The University Of Sheffield.

School of Health And Related Research.

An assessment of quality assessment and cost-effectiveness within an MCDA framework relating to Specialised Commissioning

Prepared by: Professor Simon Dixon, Dr Suzy Paisley, Professor Ron Akehurst, Dr Louise Preston, Dr Praveen Thokala, Milad Karimi, Professor Allan Wailoo and Professor John Brazier

CONTENTS

EXECUTIVE SUMMARY

The aim of this research is to support NHS England in their development of a MCDA methodology, by assessing methods to measure quality of effectiveness evidence and cost-effectiveness.

The methods of measurement are assessed using four evaluation criteria:

- Authority. The degree to which there is a body of evidence and high profile endorsement of the method.
- Consistency. The degree to which the method can be used across different topics within NHS England's portfolio and is aligned, conceptually, to methods used elsewhere in England.
- Relevance. The degree to which the methods are aligned to the type of evidence encountered and the needs of the Specialised Commissioning process.
- Workability. The degree to which the method can be used, practically, within a time-constrained process.

Four case studies were chosen by NHS England to test quality assessment and cost-effectiveness methods identified from the literature. The case studies were services that had been assessed in 2015 by NHS England's Clinical Priorities Advisory Group (CPAG) and our work was based on the same evidence base made available to the CPAG. The case study topics were:

- Robotic-assisted surgical (RAS) procedures for prostate cancer
- Treatment as prevention in HIV (HIV TasP)
- Rituximab for the treatment of steroid resistant nephrotic syndrome (SRND) in paediatrics and rituximab for the treatment of relapsing steroid sensitive nephrotic syndrome (SSNS) in paediatrics
- The use of rituximab as a second line agent for the eradication of inhibitors in patients with acquired haemophilia (R2AH)

Quality assessment (QA)

Three reviews were undertaken relating to QA of evidence on rare diseases, QA by key decision-making bodies used in the United Kingdom (UK) and latest methodological best

practice on QA. These reviews identified 21 quality assessment tools that were considered potential candidates for application to the case studies.

The most promising tools were selected for the case study work. These included a range of different types of QA tool (e.g. checklist, non-study specific tools, hierarchy of evidence), tools that were considered to score highly on the authority criteria (e.g. GRADE, Oxford Centre for Evidence Based Medicine scale (OCEBM)) and workability criteria (OCEBM, National Service Framework Long Term Conditions scale (NSF-LTC)) and a tool specific to the assessment of evidence in rare diseases (COMPASS).

The case study work showed the four tools allowed for very different assessments. No single tool was best across all evaluation criteria. For authority and consistency we considered GRADE to be the best performing tool. The NSF-LTC tool was the most relevant and workable.

The complexity of GRADE meant that it was not feasible to use the full GRADE process within the case studies given the (necessarily limited) information provided to the CPAG. It was also felt that it would not be feasible to implement the full GRADE process in the NHS England Specialised Commissioning process. However, it was felt it would be useful to incorporate the principles underpinning GRADE in the decision-making process and to combine this with the NSF-LTC which performed very well in terms of workability in all case studies.

Based on the case studies, the recommended process for the quality assessment of evidence would apply the NFS-LTC tool to summarise the quality of the body of evidence presented to the CPAG and that the CPAG use an adapted GRADE approach to support it's discussions and decision.

It was not clear how the quality of the evidence base had been assessed in the evidence reviews underpinning the commissioning documents. Without a systematic and transparent QA process at this stage, it will not be feasible to apply the recommended QA tools consistently in the commissioning process. Whilst the evidence review falls outside the scope of the current project, a further recommendation is that QA within the evidence reviews should be undertaken using a systematic and consistent approach in order that an effective summary of the quality and strength of the evidence base can be generated to inform the commissioning process.

The recommended process highlights the need for QA at three stages of the commissioning process, using separate tools that are best suited to the type of evidence available and the specific requirements of each stage of the commissioning process:

- *Stage 1: Evidence review prior to CPAG.* Individual studies included in the evidence review should be assessed using study specific quality assessment checklists.
- *Stage 2: Evidence summary documents prepared prior to the CPAG.* The NSF-LTC should be used to summarise the overall quality and strength of the evidence base presented to the CPAG. This could be done as part of the evidence review and the summary extracted for inclusion in the CPAG documentation.
- *Stage 3: Consideration of strength of evidence by the CPAG.* Onakpoya's adaptation of GRADE should be used to consider the strength of the evidence base in the context of the decision to be made.

Further adaptations of Onakpoya's GRADE method to make it more relevant to the precise decision making context of Specialised Commissioning could be considered. This may be to better reflect the nature of the evidence or the constraints of the decision making process.

Cost-effectiveness

Two searches were used to identify value criteria used in applied studies of MCDAs in health care. One search was based around references within forthcoming methodological, whilst the other examined the literature relating to value based pricing (VBP) or value based assessment (VBA). Studies known to the research team were added to these two searches.

Once all value criteria had been identified, those studies that included cost-effectiveness as a criterion were read in greater detail in order to assess the relevance of the approach used to measure performance against that criterion.

Value criteria from thirty six papers were extracted. The three most common value criteria were budgetary impact/affordability, cost-effectiveness and effectiveness of technology. In line with NHS England's requirements, studies were excluded if they used incremental cost-effectiveness ratios or quality adjusted life years (QALYs). From the 18 papers with a cost-effectiveness criterion, 16 were excluded.

The two remaining scales – one produced by the Chemotherapy Clinical Reference Group (CCRG) in relation to a previous version of the Cancer Drugs Fund and another by Diaby and colleagues – were applied to the four case studies.

The CCRG scale has limited authority as it lacks face validity. It can be consistently applied to a range of topics, but has little relevance as the vast majority of technologies are expected to fall within one of its four categories; "If no QALY available and costs more than current alternative". The scale would, however, be easy to apply to the evidence.

The Diaby scale has very limited authority as it is highly subjective and lacks face validity. The effectiveness component of the scale will be difficult to apply consistently across topics due to its subjectivity. The cost component of the scale was problematic as it was based on the distribution of costs across a set of potential programmes and not absolute costs. The overall cost-effectiveness scale is relevant, however, this only because its vagueness would mean that it can be applied to anything. The scale would, however, be easy to apply to the evidence.

Further work was undertaken to develop new scales that would differentiate services in all parts of the cost-effectiveness plane, and in particular, the north east quadrant (i.e. positive incremental costs and effects). Five measurement were considered:

1. Life-years gained.
2. Likert-type scale, for example, 'small improvements' in health gain, 'moderate improvements' and 'large improvements'.
3. Likert-type scale, as above, but with examples of health gains in order to make measurement of programmes less subjective and approximately interval. For example, 'small improvements in health gain, e.g. 3 months increase in survival, or a permanent 10% increase in quality of life'.
4. A scale based around the concept of minimum clinically important difference (MCID) whereby health gains could be measured in multiples of MCID relevant to each particular programme. This has the advantages of having interval properties and having a (weak) link through to utility. This was termed the extended north east quadrant (ENEQ) scale.
5. Use MCID (as above) but with additional descriptors that include effects not captured by the outcome measure to which MCID relates. This was termed the extended north east quadrant (ENEQ) scale with co-morbidities.

Each of the scales had considerable difficulties associated with them. Consequently, an alternative way of incorporating cost-effectiveness into an MCDA was considered. The 'extended MCDA' approach excludes cost and/or cost-effectiveness from the valuation stage of MCDA, but includes this in the subsequent funding decision.

If NHS England wishes to include a cost-effectiveness scale within the value function generated by an MCDA, then the ENEQ scale is considered the most relevant to the Specialised Commissioning process. However, its two major drawbacks should be given further consideration before being adopted. Firstly, its authority/validity should be further tested both quantitatively and qualitatively. Secondly, the availability of MCID information

and its relevance to the effectiveness evidence available to the CPAG should be assessed. If any serious failings are encountered, then an extended MCDA approach should be adopted as that removes cost and cost-effectiveness from the value function.

If a cost-effectiveness scale is used then good practice for reporting economic evaluations should be followed. In addition, work should be undertaken to specify clear and relevant reporting of economic evidence to the CPAG.

On theoretical grounds, removing cost and cost-effectiveness from the value function is the preferred way forward as it is the most appropriate way to address the issue of opportunity cost. In practical terms, this 'extended MCDA' approach also overcomes the problems encountered by the development and use of a cost-effectiveness scale for use within the estimation of the value function.

If the extended MCDA approach is used, then consideration needs to be given to benchmarking the process against the appropriate measure of opportunity cost, e.g. NHS expenditure or NHS England expenditure.

The process of undertaking an MCDA requires a number of design decisions to be made that are interdependent. The work in this report relates to just one part of this process without a full consideration of the many linkages to the rest of the process. Consequently, there needs to be further consideration of the issues raised in this report once other parts of the process are developed.

1. AIMS AND OBJECTIVES

NHS England is currently developing the methods for implementing a new prioritisation process for Specialised Commissioning, which will be based around multi-criteria decision analysis (MCDA). The value criteria for the MCDA will be based on NHS England's key principles that were consulted upon earlier this year (Appendix 1), consequently, the appropriateness of these criteria are not considered here. The work presented in this report supports this process by describing previous work that has been undertaken to measure quality of effectiveness evidence and cost-effectiveness.

The aim of this research is to support NHS England in their development of a MCDA methodology, by assessing methods to measure quality of effectiveness evidence and cost-effectiveness.

The objectives of this research are to:

- Search the literature to identify methods for Quality Assessment (QA) of effectiveness evidence.
- Identify a set of methods for QA that are appropriate for Specialised Commissioning.
- Evaluate the use of those methods in four case studies and make recommendations for use within MCDA.
- Search the literature to identify those value criteria that have been used in previous healthcare MCDAs.
- Identify a set of methods for measuring cost-effectiveness that are appropriate for Specialised Commissioning.
- Evaluate the use of those methods in four case studies and make recommendations for use within MCDA.

The report is based around two parts – the measurement of evidence quality and the measurement of cost-effectiveness. Each part is structured in the same way; reviews, case studies, development of new methods and finally, recommendations. Each part of the report can be read independent of the other. The Executive Summary, however, is a summary of both parts of the report.

2. OVERVIEW OF MCDA

2.1 Introduction

The purpose of this Section is to give an overview of MCDA so that the position of the commissioned research within the process is clarified.

2.2 MCDA approaches

It is important to understand that MCDA comprises a broad set of methodological approaches spanning a number of disciplines [1]. Whilst MCDA methods are widely used in public-sector and private-sector decisions on transport, immigration, education, investment, politics, environment, energy, defense, etc [2], the health care sector has been relatively slow to apply MCDA.[1] However, recently there has been a sharp increase in its use in health care [3].

Formal MCDA approaches can be broadly classified into value measurement (weighted-sum) models, outranking models, and reference-level models [4]. To these, we should also add 'partial MCDA methods'. In summary:

- *Value measurement models* involve constructing and comparing numerical scores representing overall value, to identify the degree to which one decision alternative is preferred over another. They most frequently involve 'weighted-sum' models; multiplying a numerical score for each criterion by the relative weight for the criterion and then summing these weighted scores to get a 'total score' for each alternative.
- *Outranking methods* involve decision-makers pairwise ranking alternatives relative to each other on each criterion in turn and then combining these pairwise rankings in order to obtain a measure of support for each alternative. Outranking algorithms include the ELECTRE family of methods [5, 6], PROMETHEE [7] and GAIA [8].
- *Reference level modelling* involves searching for the alternative that is closest to attaining pre-defined minimum levels of performance on each criterion [9]. These are broadly based on linear programming techniques and include goal, or aspiration methods [10].

[1] Although option appraisal, which could be considered a form of MCDA, has been widely used for several decades within the NHS.

- *Partial MCDA methods* do not require modelling or aggregation, although some applications have included aggregation. At the most rudimentary level, the alternatives' performance on criteria can simply be reported in a table, known as a 'performance matrix'. This matrix is then used, as a point of reference for decision-makers' deliberations.

Value measurement methods are the most widely used in healthcare [11] and is anticipated to be the approach used by NHS England. As such, the following text will focus attention on this approach.

2.3 An overview of steps in MCDA models

Value measurement modelling entails following a process commonly described in terms of eight steps (Table 1). This report is focused on Step 3, the measurement of performance against criteria, and in particular, the measurement of quality of evidence on effectiveness and cost-effectiveness.

Table 1: Steps in a value measurement MCDA process

Step	**Description**
1. Defining the decision problem	Identify objectives, type of decision, alternatives, decision-makers, other stakeholders and output required.
2. Selecting and structuring the criteria	Specify appropriate criteria for the decision problem that are relevant to decision-makers and other stakeholders.
3. Measuring alternatives' performance	Gather data about the alternatives' performance on the criteria and summarize this in a 'performance matrix'.
4. Scoring the alternatives	Eliciting stakeholders' priorities or preferences for changes within criteria (scores)
5. Weighting the criteria	Eliciting stakeholders' priorities or preferences between criteria (i.e. the weights placed on the criteria).

Step	Description
6. Calculating aggregate scores	Multiply the alternatives' scores on the criteria by the weights for the criteria and sum to get 'total scores' – by which the alternatives are ranked.
7. Dealing with uncertainty	Perform uncertainty analysis to understand the robustness of the MCDA results.
8. Interpretation and reporting	Interpret the MCDA outputs, including sensitivity analysis, to support decision-making.

As can be seen, the measurement of performance against criteria is typically undertaken after two preceding stages; defining the decision problem and selecting the criteria.

Defining the decision problem and the objective

The starting point for any MCDA involves understanding and defining the decision problem and corresponding decision goal. However, the full details of this are not yet known. Key issues that have yet to be fully resolved (to our knowledge) are:

- The MCDA approach. Whether the method to be used is a value measurement (weighted-sum) model, outranking model, reference-level model, or informal methods.
- The decision outcome. This can be the value of the alternatives (e.g., understanding the value of treatment for a subsequent decision), ranking a finite set of alternatives (e.g., prioritising investments), or a binary outcome (e.g., 'approve' or 'deny' recommendations for new technologies).
- The position of cost-effectiveness within the MCDA process. Whilst it is possible to include 'cost' or 'cost-effectiveness' as a criteria within an MCDA objective function, this generally misrepresents opportunity cost in the presence of a fixed budget. Consequently, if opportunity cost is an important consideration, MCDA is best used to generate aggregate value, which is then combined with cost information in an 'extended MCDA'. This point is considered in greater detail later in the report. Interestingly, some researchers consider the generation of an incremental cost per QALY gained to be a form of (extended) MCDA; quality of life criteria are weighted

and summed to produce utilities, then combined with time in a value function to produce QALYs and finally combined with cost information.

Selecting and structuring the criteria

The criteria for the NHS England MCDA process will be based on their key principles. However, the precise definitions of these are not yet known. Without robustly developed and precisely defined criteria, our ability to assess measurement of performance against those criteria can be limited. So, for example, it is not clear whether uncertainty around the economic evidence should be considered by the cost-effectiveness criterion, or whether it will be picked by a criterion looking at 'robustness of evidence'.

It should also be noted that the criteria used in a weighted sum value measurement model should meet certain requirements such as completeness, non-redundancy, non-overlap and preferential independence.[2] Previous studies that include cost-effectiveness as a criterion have been criticised as there can be overlap and preferential interdependence between cost-effectiveness and the other criteria (for example, effectiveness). However, an evaluation of the full MCDA being developed by NHS England is beyond the scope of this work.

Measuring the performance of the alternatives

Once the criteria are agreed upon, the performance of the alternatives on each of the criteria is determined. Measurement is generally continuous (e.g., survival or cost) or ordinal (e.g., hierarchy of evidence), although count or categorical measurement is also possible. The choice of measure also has an impact on subsequent steps in the MCDA. Despite this inter-relatedness, this report will not consider the subsequent impact of the recommended measures on the other stages of the MCDA.

2.4 Summary

The process of undertaking an MCDA requires a number of design decisions to be made. The work in this report relates to decisions on the measurement of performance against two

[2] Completeness relates to the degree to which the criteria cover the relevant valuation space. Non-redundancy relates to the degree that criteria are not necessary. Non-overlap relates to the degree that the criteria do not measure the same part of the valuation space. Preferential independence relates to the degree that stakeholder preferences for one criterion are not influenced by their preferences toward other criteria. In theory, interdependence can be accommodated by interaction terms within the estimated value function.

criteria. This work will influence, and be influenced by, other ongoing work being undertaken by NHS England. Consequently, further iterations are likely to take place when other parts of the process are developed, however, these additional changes are beyond the scope of this work.

3. LITERATURE SEARCHES RELATING TO QUALITY ASSESSMENT

3.1 Introduction

In order to inform methods for assessing the quality of research evidence in the NHS England Specialised Commissioning process three focussed reviews were undertaken. The specific purpose of the reviews was to identify four candidate assessment tools that could be applied to three case study Clinical Commissioning Policy (CCP) documents.

The aims of the individual reviews were as follows:

- Review 1: to identify tools specific to the quality assessment (QA) of evidence on rare diseases.
- Review 2: to identify tools used by key decision-making bodies used in the United Kingdom (UK).
- Review 3: to identify tools that represented the latest methodological best practice on the quality assessment of evidence.

Several key factors guided the review process:

- The scope of the reviews should be wide in order to identify a broad cross-section of QA tools.
- The purpose of the reviews was to identify QA tools that had the potential to meet the evaluative criteria agreed with NHS England:
    - Authority. The degree to which there is a body of evidence and high profile endorsement of the method.
    - Consistency. The degree to which the method can be used across different topics within NHS England's portfolio and is aligned, conceptually, to methods used elsewhere in England.
    - Relevance. The degree to which the methods are aligned to the type of evidence encountered and the needs of the Specialised Commissioning process.
    - Workability. The degree to which the method can be used, practically, within a time-constrained process.
- NHS England requires QA tools that assess the strength of a body of evidence and that can be used in a decision-making process. Tools that included some form of judgment about the strength of evidence in informing a decision or in making recommendations were likely to be more useful than tools that just assessed study quality *per se*.

- QA tools for evidence on rare diseases would be highly relevant to the Specialised Commissioning process and general issues relating to QA in the context of rare diseases would be of interest to the Specialised Commissioning process.

3.2 Methods

Review 1: QA of evidence in rare diseases

Literature searches of Medline and the Cochrane Library were undertaken. (Strategies are reported in Appendix 2). Search results were imported into Reference Manager (RM) software. Titles and abstracts were sifted to identify:

- Specific tools used to assess the quality of evidence for rare diseases
- Papers that described and commented on the nature and quality of evidence used to assess interventions for rare diseases
- Organisations involved in the commissioning of interventions for rare diseases or specialised services.

Leads to other sources of interest (not specifically relating to rare diseases / specialised commissioning) were followed up and added to the RM databases if considered to be relevant to the review.

A data extraction table was designed to summarise information relating to specific tools, taking into account the key factors that guided the review process as described above. The table was populated with data extracted from papers that provided information on specific QA tools used to assess evidence on rare diseases.

Within the timeframes of the project it was not possible to summarise all papers that described the nature and quality of evidence used to assess interventions for rare diseases. A review, undertaken by the Institute for Quality and Efficiency in Healthcare (IQWiG) in Germany [12] described some of the issues associated with QA evidence assessment in rare diseases. These have been summarised in the discussion section (4.4) in the QA case studies chapter.

Review 2: QA tools used in key decision-making bodies in the UK

The websites of the following health-related decision-making bodies were surveyed:

- NICE Guidelines
- NICE Appraisals
- NICE Interventional Procedures
- NICE Diagnostics
- NICE Highly Specialised Technologies (HST)
- NICE Medical Technologies Evaluation Programme (MTEP)
- Scottish Intercollegiate Guidelines Network (SIGN)
- Scottish Medicines Consortium (SMC)
- Scottish Health Technologies Group
- All Wales Medicine Strategy Group
- UK National Screening Committee (NSC)
- Joint Committee on Vaccination and Immunisation (JCVI)
- Northern Ireland Guidelines and Implementation Network (GAIN)
- NHS England Cancer Drug Fund (CDF)

Leads to other sources of interest (not restricted to the UK) were followed up and added to the review if considered relevant.

All relevant documentation which could be identified via the websites, including methods and process manuals and submission templates were scanned for methodological information relating to the quality assessment of evidence. Details of specific QA tools were extracted and added to the data extraction table used in Review 1.

Review 3: Methodological best practice in the quality assessment of evidence

Literature searches of Medline, the Cochrane Library and four evidence synthesis methodological journals were undertaken. (Strategies are reported in Appendix 2). Search results were imported into Reference Manager software. Titles and abstracts were sifted to identify:

- Specific QA tools considered relevant to the Specialised Commissioning process and not identified by Reviews 1 and 2

- Comment on recent methodological developments and 'best practice' in the quality assessment of evidence

Details of specific QA tools were extracted and added to the data extraction table.

It was not possible, within the timeframe of the project, to summarise all papers on recent methodological developments and 'best practice'. These papers were useful however in the selection of specific tools for inclusion in the reviews and for application to the case studies.

Selection of QA tools for application to case studies

Details of the QA tools identified in the three reviews were entered into the single data extraction table. The selection of tools for application to the case studies was based on an initial consideration of the four evaluative criteria to be used in the case studies and on the extent to which it was felt the tools could be applied to a NHS England's decision-making process. Preliminary selections were discussed with NHS England at two teleconferences and the final selection was agreed at a third teleconference.

3.3 Results

Review 1 searches retrieved 1147 references. Review 3 searches retrieved 1051 references. Following the title and abstract sift, 120 papers, together with the information from the UK decision-making body websites in Review 2, were used to draw up a list of 21 quality assessment tools that were considered potential candidates for application to the case studies. The 21 QA tools included in the review are listed below. The key characteristics of each tool, on which the final selection of tools for application to the case studies were based, are summarised in Table 1. It should be noted, that there is considerable overlap between many of the tools.

Critical Appraisal Skills Programme (CASP) Checklists [13]
Consolidated Health Economic Evaluation Reporting Standards (CHEERS) [14]
Cochrane Effective Practice and Organisation of Care (EPOC) suggested risk of bias criteria[15]
Cochrane Risk of Bias (RoB) Tool [16]
Clinical Evidence of Orphan Medicinal Products – An Assessment Tool (COMPASS) [17]
Centre for Reviews and Dissemination (CRD) Checklists [18]

Drummond BMJ Checklist [19]
FORM Grading System [20]
Grading of Recommendations, Assessment, Development and Evaluation (GRADE) [21]
Hierarchy of Evidence and Appraisal of Limitations (HEAL) [22]
NHS England Cancer Development Fund (CDF) [23]
National Service Framework for Long Term Conditions (NSF-LTC) Typology [24]
Oxford Centre for Evidence Based Medicine (OCEBM) Levels of Evidence [25]
Onakpoya, 2015 (Study to assess orphan drugs using OCEBM and GRADE in combination)[26]
Ottawa Panel Evidence Based Clinical Practice Guidelines [27]
Quality of Health Economic Studies (QHES) [28]
Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) [29]
Scottish Intercollegiate Guidelines Network (SIGN) checklists [30]
Scottish Medicines Consortium (SMC) checklists [31]
Strength of Recommendation Taxonomy (SORT) [32]
US Preventive Services Taskforce (USPST) [33]

Assessment of tools against evaluative criteria

Eleven[3] QA tools were considered to meet the authority criteria as they were used or recommended for use to support high profile decision-making processes in the UK (e.g. NICE), were used by Cochrane or were prominent in the methodological literature relating to quality assessment of evidence.

Seven[4] QA tools were considered to meet the consistency criteria as they could, potentially be used across the range of topics considered by NHS England and aligned, conceptually, with methods used elsewhere in England.

Nine[5] QA tools were considered to meet the relevance criteria as they could be applied to a range of study designs, were linked to making judgments about the strength of evidence to support decisions and make recommendations and/or were used in the assessment of evidence in the context of rare diseases.

[3] CASP, CHEERS, Cochrane EPOC, Cochrane RoB, CRD, Drummond BMJ, GRADE, QUADAS, OCEBM, SIGN, SMC

[4] CASP, Cochrane RoB, CRD, Drummond BMJ, GRADE, NHS England CDF, NSF-LTC

[5] COMPASS, FORM, GRADE, NSF-LTC, Onakpoya, Ottawa Panel, SIGN, SORT, USPST

Four[6] QA tools were considered to meet the workability criteria as they had the potential to work within a time-constrained process.

[6] NHS England CDF, NSF-LTC, OCEBM, Onakpoya

Table 1: Quality assessment tools identified by reviews

Tools	Non-study specific	Study specific	Linked to Recommendation / judgment	Rare Diseases / Spec Services	Authority	Consistency	Relevance	Workability	Notes
CASP		✓ Suite of checklists			✓	✓			NICE HST (CASP cohort)
CHEERS		✓ Economic evaluation			✓				NICE Guidelines
Cochrane EPOC	✓				✓				
Cochrane RoB	✓				✓	✓			Cochrane, NICE Guidelines
COMPASS	✓			✓			✓		RD specific
CRD		✓ Suite of checklists			✓	✓			NICE Appraisals
Drummond BMJ		✓ Economic evaluation			✓	✓			NICE HST
FORM	✓		✓				✓		Developed by Australian National Health and Medical Research Council (NHMRC)
GRADE	✓		✓	✓	✓	✓	✓		SIGN, NICE Guidelines, US AHRQ, WHO
HEAL	✓		✓						Published 2015
NHS England CDF		✓ Modified OCEBM	✓			✓		✓	
NSF-LTC Typology	✓		✓			✓	✓	✓	NSF Long term conditions

Non-study specific=generic checklist, not intended for a specific study design. Study specific=checklists comprises questions to assess specific study design (e.g. RCT) Linked to recommendation / judgment=checklist used in conjunction with some form of recommendation process, with strength of evidence based on strength of evidence, RD / Spec Services=use or potential use for rare disease / spec services highlighted in literature. Notes=highlights where checklist is used in / associated with specific process

Table 1: Quality assessment tools identified by reviews (cont)

Tools	Non-study specific	Study specific	Linked to Recommendation / judgment	Rare Diseases / Spec Services	Authority	Consistency	Relevance	Workability	Notes
OCEBM		✓ Evidence hierarchy			✓			✓	
Onakpoya, 2015		✓ (OCEBM)	✓ (Adapted from GRADE)	✓			✓	✓	Combined use of OCEBM and GRADE to assess orphan drugs
Ottawa panel		✓ Evidence hierarchy	✓				✓		Ottawa Panel Evidence-Based Clinical Practice Guidelines
QHES		✓ Health economics							
QUADAS		✓ Diagnostic studies			✓				NICE Diagnostics, NSC
SIGN		✓ Suite of checklists	✓		✓		✓		SIGN
SMC		✓ Suite of checklists			✓				SMC
SORT		✓ Evidence hierarchy					✓		American Family Physician Published 2004
USPST		✓ Suite of checklists	✓				✓		

Non-study specific=generic checklist, not intended for a specific study design. Study specific=checklists comprises questions to assess specific study design (e.g. RCT) Linked to recommendation / judgment=checklist used in conjunction with some form of recommendation process, with strength of evidence based on strength of evidence, RD / Spec Services=use or potential use for rare disease / spec services highlighted in literature. Notes=highlights where checklist is used in / associated with specific process

3.4 Discussion and conclusions

Based on the qualitative evaluation process, a set of tools that are potentially applicable to the Specialised Commissioning process were identified. The selection process ruled out a number of QA tools (e.g. CRD checklists) on the grounds that they could be used to assess individual studies but not to assess the strength of a body of evidence to inform a decision or make a recommendation. Other QA tools (e.g. USPST, FORM) were ruled out if an equivalent type of tool (e.g. hierarchy of evidence), that had evaluated well against the criteria, had been developed and was used in the UK.

The final selection included a range of different types of QA tool (e.g. checklist, non-study specific tools, hierarchy of evidence), tools that were considered to score highly on the authority criteria (e.g. GRADE, OCEBM) and workability criteria (OCEBM, NSF) and a tool specific to the assessment of evidence in rare diseases (COMPASS). A brief description of the tools is given below.

GRADE

GRADE[21] is used in systematic reviews and guideline development to make transparent judgments about the quality of a body of evidence and to indicate the strength of recommendations based on that evidence. The GRADE tool assesses the evidence according to outcomes across studies rather than assessing the quality of individual studies. The process involves a number of stages as follows:

1. Formulate question
2. Specify outcomes
3. Rate importance of each outcome
4. Generate evidence profiles rating quality of evidence for each outcome
5. Upgrade or downgrade quality rating according to consideration of factors relating to internal and external validity
6. Based on evidence profiles, make recommendations for or against the use of a technology, based on evidence classed as strong or weak/conditional.

GRADE was selected for use in the case studies because it can be used to assess bodies of evidence across study designs. In addition, one study identified by the reviews used adapted GRADE criteria to assess the quality of evidence used to support orphan drug approvals.[26] The selection of GRADE was based mainly however on the extent to which it met the authority criteria. It is used by a broad range of organisations both within the UK and internationally. It accounted for a large proportion of papers retrieved by Review 3 on current best practice in the quality assessment of evidence. The ability to upgrade or

downgrade evidence according to factors such as inconsistency, indirectness and imprecision and to use this to assess confidence in estimates of effect are seen as particular strengths of GRADE. Drawbacks are that the process of applying GRADE is perceived as being complex.

With regard to the latter, it quickly became clear that it would not be possible to apply the complete GRADE process to the information available in the case study Clinical Commissioning Policy documents and it was felt that the process would not be workable in the overall Specialised Commissioning process. Given the extent to which GRADE met the authority criteria it was felt that it could not be excluded from the case study evaluation. As such, it was decided that some form of assessment, based on the GRADE principles of assessing the quality of evidence according to categories, in particular, of inconsistency, indirectness and imprecision, should be included. A GRADE approach had been taken in the study by Onakpoya, using GRADE in combination with the OCEBM hierarchy.[26] Whilst the GRADE ratings (high, moderate, low, very low) used in the Onakpoya study looked useful, it was not clear how the GRADE assessment had been applied to the evidence and therefore it was felt it would not be possible to replicate the approach in the case studies. From the large body of GRADE-related literature identified by the review, we selected a checklist developed by CRD to aid the assessment of evidence using GRADE criteria.[26] The GRADE CRD checklist specifies a series of questions requiring largely yes / no / unclear responses for each of the GRADE upgrading / downgrading categories. This constitutes the GRADE approach used in the case studies.

OCEBM

OCEBM[25] is a well-established, study design-based hierarchy of evidence developed by early proponents of the evidence-based medicine movement and revised in 2011. The hierarchy specifies a number of different types of question (including diagnosis, prognosis, treatment benefits, treatment harms) then for each question type, specifies different study designs to represent evidence quality levels (level 1 being considered high quality evidence, level 5 being considered the lowest quality).

OCEBM was selected due to its authority and chiefly because its simple format was judged to be amenable to a time-constrained process. An example of the use of OCEBM in the assessment of evidence on orphan drugs was also identified.[26] The potential for practical workability was judged to be the main strength of the OCEBM hierarchy. Potential drawbacks were that OCEBM might be too blunt an instrument, with quality assessment

being restricted to judgments largely according to study type rather than the overall methodological quality of study design.

NSF-LTC

NSF-LTC[24] was developed to assess the quality of evidence and support recommendations as part of the National Service Framework on Long Term Conditions. The NSF-LTC typology rates individual sources of evidence according to three categories; design, quality and applicability. The design category accommodates primary and secondary evidence and quantitative, qualitative and mixed methods. The quality category scores study quality up to 10 based on 5 quality assessment questions. The applicability category assesses the directness of the evidence to the research questions or context of the decision or recommendation. The NSF-LTC then assesses the rating of the individual studies to generate a grade for the overall body of evidence.

NSF-LTC was selected because it is can be used to assess a body of evidence and can be used to assess any type of study design. It was selected chiefly because it was judged to be amenable to a time-constrained process. This was judged to be the main strength of the NSF-LTC typology. Its potential drawbacks were perceived as being that it might be too brief and generic to assess evidence in sufficient detail and that it did not have the same level of authority or widespread use as other tools such as GRADE. An evaluation of NSF-LTC against GRADE and SIGN levels of recommendation was identified by the literature search.[34] This found that "all three systems had strengths and weaknesses depending on the type of evidence being graded. GRADE was seen as the most complex but rigorous system, while SIGN and NSF were seen as easier and more flexible to use, but less methodologically rigorous. It is recommended that specialist societies consider the type of evidence they will be evaluating and the research experience of the appraisers before selecting a grading system. Additionally, appraisers should have training in appraising and grading evidence using the system to be employed."

COMPASS

COMPASS[17] is a three-part quality assessment checklist designed specifically for the assessment of evidence on orphan drugs for rare diseases. COMPASS was developed based on evidence submitted for licensing by the European Medicines Agency (EMA). The questions which comprise parts1 and 2 of the checklist focus specifically on information

relating to the licensing process and as such were considered not relevant to the case study evaluation. Part 2 of the COMPASS checklist focusses on study quality and comprises a series of questions, with pre-specified tick box response options, covering nine aspects of study design

COMPASS was selected because it was designed specifically for evidence relating to rare diseases. This was considered its main strength. Potential drawbacks were that COMPASS is a quality checklist for the assessment of individual studies and cannot be used to assess the strength of a body of evidence. Also, COMPASS was developed based on evidence used in EMA licensing decisions. Its use is not well established and its generalisability beyond the EMA process is not clear.

4. APPLICATION OF QA TOOLS TO THE CASE STUDIES

4.1 Introduction

In order to assess the quality and strength of evidence used in the decision making process for NHS Specialised Commissioning, we used a sample of four Clinical Commissioning Policy (CCP) reports. We applied the four identified tools to the available clinical-effectiveness information in the CCP documents and assessed the extent to which the tools met four predetermined criteria (authority, consistency, relevance and workability).

Three case studies were selected by NHS England to represent a range of topics and types of evidence base. The case studies were:

- Robotic-assisted surgical (RAS) procedures for prostate cancer
- Treatment as prevention in HIV (HIV TasP)
- Rituximab for the treatment of steroid resistant nephrotic syndrome (SRND) in paediatrics and rituximab for the treatment of relapsing steroid sensitive nephrotic syndrome (SSNS) in paediatrics

An additional case study was subsequently added as the nature of the SRND/SSNS case study did not allow the four scales to be successfully applied (although our attempt is reported). The additional case study was:

- The use of rituximab as a second line agent for the eradication of inhibitors in patients with acquired haemophilia (R2AH)

4.2 Methods

The four evidence assessment tools identified in the review were applied to each of the four case studies by one reviewer (LP). An additional reviewer (SP) applied the tools to two of the case studies. The information available to the research team for each of the case studies represented the information that was presented to the CPAG in the CCP document, specifically Section 5 (Evidence Base) and Section 6 (Rationale behind the policy statement). For the GRADE tool, the validated checklist developed by the Centre for Reviews and Dissemination (CRD) (Meader et al 2014) was used as using the full GRADE process was not feasible (due to both the evidence and the resources available). Then a qualitative assessment of each tool was made, based on the criteria of authority, consistency, relevance and workability. No formal inter-rater agreement was undertaken; however a qualitative

discussion took place regarding the *performance* against the evaluation criteria of each of the tools (as opposed to comparing the *QA results* from each tool used).

4.3 Results

The results for each case study are presented in a similar format. The results gained from using the tool and a brief assessment of its performance, given the information provided are presented in Appendix 3. Below there is a narrative assessment of the information provided in the CCP and a narrative assessment of the performance of the tools for each case study.

Case Study 1: Robotic-assisted surgical (RAS) procedures for prostate cancer

Narrative assessment of the information provided in the CCP

The information provided in the CCP is a summary of the evidence review undertaken by Solutions for Public Health. The summary considers robotic versus laparoscopic and robotic versus open surgery. The outcomes are not consistently reported. The only detail given is the study type and the intervention. The NICE guidance on robotic assisted prostatectomy is summarised, emphasising the information from a paper based on a HTA. Referencing is a little inconsistent. Some statements derived from evidence e.g. "Estimated blood loss is less with robot-assisted prostatectomy than with either alternative procedure" are not supported by either a source for this statement or numerical evidence. Some assumptions have to be made, for example, the summary refers to Gandaglia et al (2014), then to a "similar study", which, it is assumed, is also a controlled, non-randomised study. Upon further examination of the evidence in the summary, two of the randomised controlled trials that are reported prior to the reporting of a systematic review are themselves included in the systematic review (through reporting of a meta-analysis in which they are included), therefore increasing the weight of importance of this evidence in the decision making process through, effectively, double counting the RCT studies.[35]

Narrative assessment of the performance of the tools for the case study.

The NSF-LTC was the most straightforward tool to apply, with the information available in the CCP. There was insufficient evidence to apply the CRD GRADE checklist. It was possible to implement the OCEBM hierarchy with the evidence scoring relatively highly on the OCEBM criteria. Only a few of the questions on the COMPASS checklist were answerable with the information provided.

Case Study 2: Treatment as prevention in HIV

Narrative assessment of the information provided in the CCP

The information provided is a summary of the evidence review produced by 'Public Health'and the HIV Clinical Reference Group. The 'evidence base' section of the CCP covers a number of different topics, including: evidence about the effectiveness of the intervention (one RCT and three cohort studies), effectiveness of a comparator intervention (condom use) (one modelling study, one meta-analysis of non-randomised studies and two studies where study type is not given), effectiveness of the intervention in different situations (non-sexual transmission) (three studies where study type is not given), safety of the intervention (one study where study type is not given) and cost effectiveness of the intervention (six studies, study type given as cost effectiveness analyses). The evidence is described in terms of its quality intermittently throughout the 'evidence base' section. Some numerical evidence is provided in terms of study sizes, reduction in viral load etc. There appears to be a structure to the section (which may based on the evidence review) of efficacy, safety, impact on quality of life and cost-effectiveness). The 'rationale behind the policy statement' describes the evidence as "high quality' but no reason is given as for this rating.

Narrative assessment of the performance of the tools for each case study.

This case study was differentiated by the variety of different types of evidence considered – using randomised controlled trials to demonstrate efficacy but other study types as well. The evidence summary also included information on intervention acceptability. Again, the NSF LTC allowed for the most complete assessment of the available evidence. The OCEBM did not capture the complexity of the issue and the fact that different types of evidence need to be combined to address the commissioning question.

Case Study 3: Rituximab for the treatment of steroid resistant nephrotic syndrome (SRND) in paediatrics and rituximab for the treatment of relapsing steroid sensitive nephrotic syndrome (SSNS) in paediatrics

No evidence was presented in the Consultation Report, although mention was made of the NHS England Specialised Commissioning Clinical Effectiveness lead evidence review.

Case Study 4: The use of Rituximab as a second line agent for the eradication of inhibitors in patients with Acquired Haemophilia

Narrative assessment of the information provided in the CCP

The CCP refers to two systematic reviews of case reports and other non-randomised studies addressing the topic; however it does not give any bibliographic detail for one of the two reviews. The policy summarises the studies, outcome measures and critiques the review, in terms of confounding variables and length of follow up.

Narrative assessment of the performance of the tools for the case study

COMPASS, GRADE and OCEBM prioritise evidence from randomised controlled trials, so, even if sufficient information were available in the CCP, they would have assessed the evidence supporting the decision as being limited.

Assessment of the tools against the evaluation criteria

Table 2 brings together the performance of the different tools with the predetermined criteria to evaluate to what extent the different checklists meet the needs of the NHS England Specialised Commissioning Process. The boxes in that are highlighted in grey are the "best performing" by our qualitative assessment.

Table 2: Summary of tool ratings

	COMPASS	GRADE	OCEBM	NSF-LTC
Authority[7]	COMPASS is a relatively new tool and is not used frequently (due to the nature of the checklist), therefore does not have the high profile of GRADE and OCEBM.	GRADE is the most authoritative tool available for assessing the strength of bodies of evidence. Based on the volume of papers retrieved on GRADE it is perceived as being highly important in informing 'best practice' on the quality assessment of evidence.	This is a long established tool. It is well recognised.	This tool is relatively new and there is not much evidence for its use within the academic literature.
Consistency[8]	It would be challenging to use this tool across different topics as it is very focussed on medicinal products.	It would be possible to use this across different topics and the tool is the most closely aligned to current NHS practice.	It would be possible to use this across different topics. The challenge is in differentiating between topics – for example, in some topics, case reports may constitute the best available evidence.	This tool could be used across the portfolio of topics. The limitation of the tool would be to what extent it allows comparability across topics, where the quality and (study) type of evidence differs.
Relevance[9]	The evidence encountered with the CCP is not sufficiently detailed to apply the COMPASS checklist, which requires evidence at individual study level. It doesn't assess the strength of a body of evidence. COMPASS is more focussed on medicinal products. The NHS Specialised Commissioning process has a much wider portfolio than medicinal products alone.	The evidence in the CCP is not sufficiently detailed for the GRADE process. The recommendations that the GRADE process generates may be of relevance to the Specialised Commissioning process as they base the recommendation on the strength and the quality of the evidence.	This checklist allows for an assessment of all different types of evidence, so in that sense, would be a good fit for the Specialised Commissioning process. However study quality is defined strictly according to study type and it does not allow for the fact that decision making processes often have to consider lower level evidence as this is sometimes the best available evidence.	The NSF LTC does allow for a variety of study types to count as relevant evidence and the research design of a study does not determine its quality. The ability to assess the quality of evidence separately from study design or type is highly relevant within this process.
Workability[10]	The checklist is not onerous to use, but does rely on having evidence at an individual study level.	The GRADE process is necessarily time consuming, based on the information that is required to produce the eventual recommendations. It is not practical to use within a time-constrained process.	The tool can be used within a time constrained process to evaluate the evidence that is included in the CCP.	This tool is very workable. Allows rapid assessment.

[7] The degree to which there is a body of evidence and high profile endorsement of the method.
[8] The degree to which the method can be used across different topics within NHS England's portfolio and is aligned, conceptually, to methods used elsewhere in England.
[9] The degree to which the methods are aligned to the type of evidence encountered and the needs of the Specialised Commissioning process.
[10] The degree to which the method can be used, practically, within a time-constrained process.

4.4 Discussion

For authority and consistency we considered GRADE to be the best performing tool. The NSF-LTC tool was the most relevant and workable. The complexity of GRADE meant that it was not feasible to use within the (necessarily limited) information provided in the CCP.

The four tools allowed for very different assessments. COMPASS is very focused on medicinal products and orphan drugs/rare diseases. However it is a checklist of study quality and best applied at individual study level. It cannot be used to assess the strength of a body of evidence. The OCEBM tool allows a rapid assessment of the certainty from which conclusions can be drawn from a study, based according to study type. However it does not account for the variety of study types that constitute evidence in the CCP case studies. The GRADE process is well established and clear but not considered workable for the Specialised Commissioning process. However it does offer principles which may be of use. The NSF-LTC is a flexible and workable tool for examining individual studies and for generating a summary assessment of the strength and quality of a body of evidence. It is straightforward to apply, but lacks the authority of some of the more long standing and frequently applied tools.

A standard procedure for consolidating the evidence included in the Evidence Review into the Summary of Evidence for the CCP would make any assessment of the quality of the evidence more straightforward. Tabulating this evidence in a standard format would allow for comparisons between different Clinical Commissioning Policies (and therefore decisions).

Choice of tools

Based on the reviews and the case studies it is not possible to recommend one single QA tool which would meet all the requirements of the Specialised Commissioning process. Working with the CPAG documentation in the case studies it is also considered that recommendations on the overall process of QA, in addition to recommendations relating to specific QA tools might be helpful.

The overall recommendation is that some form of QA process is required at three stages of the Specialised Commissioning process. This would ensure that the quality of evidence was systematically assessed at the evidence review stage, was presented as a body of evidence in a meaningful summary format to the CPAG (and in the final CCP document) and was

systematically discussed in the context of the prioritisation decision being made by the CPAG. Each of these stages involves slightly different QA activities.

The recommended QA process includes the specification of different tools considered relevant to each of these activities. For the evidence review, quality assessment is recommended using study specific quality assessment checklists. For generating the summary of the quality of the body of evidence the NSF-LTC typology is recommended. For the consideration of evidence quality in the context of the prioritisation decision, a discussion by the CPAG, based on the GRADE criteria, particularly relating to inconsistency, indirectness and imprecision, is recommended. This discussion could be summarised to record the judgment about the evidence base by the CPAG in the context of the prioritisation decision. It is felt that the GRADE CRD checklist, used in the case studies, is overly detailed for supporting a discussion and therefore the GRADE CRD checklist is not recommended for this QA activity. The discussion could, however, be summarised using the categories reported in the Onakpoya study included in the review.[26] The rationale underpinning the approach is discussed below. The recommended approach is presented in more detail in section 5.

The requirements of the Specialised Commissioning prioritisation process were the key considerations in the choice of recommended QA tools and processes. Specific considerations, relating particularly to the workability criterion, include the need to support batch prioritisation processes within relatively short timescales, the need to accommodate a broad range of evidence types and the need to implement workable, consistent practices across the prioritisation process, including the different Clinical Reference Groups. Whilst GRADE was considered the most authoritative quality assessment tool, representing state of the art 'best practice' in QA methods, it was judged that the GRADE process was overly complex and did not fit well within the Specialised Commissioning process. A key strength of GRADE, in terms of its ability to incorporate judgments about confidence in the estimate of effect, including the directness of the evidence to the context of the decision is important however, and has been incorporated in the recommended approaches to QA.

Tools designed specifically for the assessment of evidence on rare diseases were considered, potentially, to be highly relevant. Such tools were perceived as being capable of accommodating the broad range of evidence types used to inform the commissioning process and of taking account of methodological weaknesses in study design, or difficulties of achieving adequate study design, perceived as being typical of evidence on specialised interventions and interventions for rare diseases. The COMPASS tool was selected for application to the case studies because it was designed specifically for orphan drugs. It was

not considered to perform particularly well in the case studies and did not seem to offer much in terms of requirements specific to orphan drugs and rare diseases. That is, it did not accommodate a broad range of study designs and it did not address sufficiently issues relating to weaknesses in study design. Whilst the COMPASS tool does not form part of the recommended QA process, the recommended approaches do, nonetheless, include QA tools that take account of a broad range of evidence types and do not assess the quality of evidence solely according to a strict study design hierarchy.

However, a discussion remains on whether the QA process should allow for methodological weaknesses considered typical in evidence on specialised interventions and rare diseases. A review of evidence characteristics for studies on rare diseases, undertaken by IQWiG and identified in Review 1 is useful in informing this discussion.[12] The IQWiG review provides expert opinion on methodological issues relating to the design of studies on interventions for rare diseases and reviews the characteristics of studies used in the approval of orphan drugs in Europe. The review concludes that there is no scientific rationale for using different approaches in the assessment of interventions for rare versus non-rare diseases. There are no specific designs for rare diseases that could not also be used for non-rare diseases. Moreover, it reports that approvals for orphan drugs are largely based on randomised studies, concluding that the feasibility of conducting RCTs is not in question. It acknowledges however, that in practical terms, in making decisions about interventions for rare diseases it may be necessary to *'make compromises with regard to the reliability of the conclusions'* of studies. These are described, in descending order of acceptability, as:

- compromises with regard to the level of precision of study conclusions,
- compromises with regard to external validity including the acceptance of evidence from similar therapeutic indications or using established surrogate outcomes
- compromises with regard to internal validity by considering non-randomised data from disease registries *'with clear quality criteria'* or where the effect size is so large that it cannot be explained by bias alone. Conclusions based on these type of data are typically considered more robust when combined with clinical plausibility of a clear effect, for example, as seen with enzyme replacement therapies.

In making recommendations on QA for the Specialised Commissioning process it was important to accommodate the wide range in the quality of evidence available to the prioritisation process and to consider how this should be accounted for in the quality assessment process. That is, whether 'allowances' should be made for the varying quality of evidence in this context. In line with the IQWiG conclusion that there is no scientific rationale for using different approaches in the context of rare diseases it was decided that, in

the first instance the scientific quality of evidence should be assessed according to usual scientific criteria (i.e. focussing on internal validity). This approach is recommended for the evidence reviews underpinning the commissioning process and for the presentation of the evidence base in the Clinical Commissioning Policy document. In line with the IQWiG conclusion that it may be necessary to 'make compromises' with regard to the reliability of results when making decision about interventions for rare diseases, it was decided that the quality of the evidence, defined according usual criteria, should be discussed in the context of the decision to be made. That is, the impact of the quality of the evidence on the decision-making process should form part of the CPAG discussion.

It was agreed that the GRADE criteria for upgrading or downgrading evidence according to inconsistency, imprecision and indirectness reflected, in part, the compromises described by IQWiG. Consequently, the recommended QA process here is that a discussion should take place according the GRADE criteria (based on Onakpoya [26]) in order to make explicit the bearing of the quality of evidence on the commissioning decision. The overall QA process, therefore, is a hybrid process that incorporates established quality criteria for the assessment of evidence *per se* and an assessment of the quality of evidence in the context of Specialised Commissioning decisions.

It might be, however, that NHS England may wish to amend the GRADE criteria further to enhance its relevance and ease of use to the CPAG. For example, in the case of single arm studies, the confidence about the prognosis of untreated patients will influence an assessment of effect size (e.g. with enzyme deficiencies, we are very confident about prognosis even in the absence of RCTs). Whilst this may be captured within a full GRADE process undertaken by experts, it may be not be captured in a time limited process.

Limitations of the case studies

A clear limitation of the assessment of QA tools in the case studies lay in the inconsistencies and the lack of detail in the presentation of clinical effectiveness evidence in the Clinical Commissioning Policy documents. It is recommended that an important part of a QA process should be an effective summary of the quality and strength of the body of evidence available to the commissioning process in order to facilitate a transparent discussion on the impact of the quality of evidence on the decision making process. This is particularly important in avoiding an 'evidence drift' in the interpretation of the effectiveness evidence and in specifying the rationale behind the policy statement. An effective and transparent QA approach would make clear the extent to which the policy rationale is informed by the evidence and the extent to which the statements which form the rationale can be linked directly to the summary of the clinical effectiveness evidence.

A possible limitation of the recommendations is that they are based on the application of selected QA tools to a small number of case studies. In addition, the case study work was a desk-based exercise rather than being embedded within Specialised Commissioning process with its own particular contextual factors. It is possible that the recommended QA tools may need further explanation or adjustment to the requirements of the commissioning process. In particular, it is felt that questions in the NSF-LTC, relating to study design and methods may need further specification in order to capture information in sufficient depth. The means by which a CPAG discussion, based on the GRADE criteria, can best be captured and represented using the approach used by Onakpoya may need also agreement.

5. BUILDING ON THE QA CASE STUDY WORK

5.1 Introduction

A single scale that is appropriate to the Specialised Commissioning process was not identified. However, the relative advantages of the most appropriate scales were identified in the case study work. From this, and with the needs of NHS England in mind, a hybrid approach was developed that can be applied within the Specialised Commissioning process.

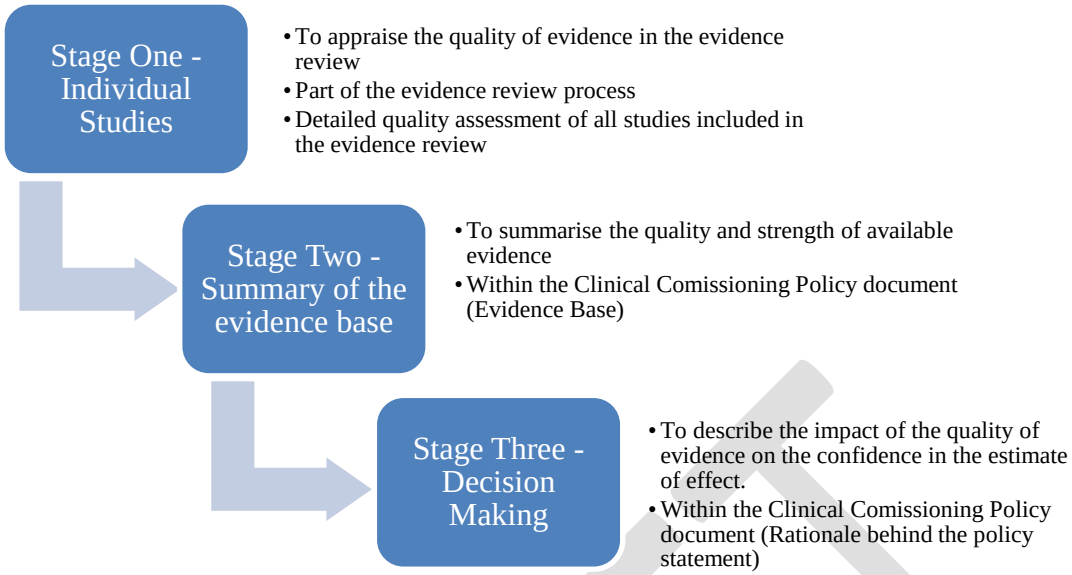
Consequently, recommendations for the assessment of the quality of evidence used to support NHS England's Specialised Commissioning decisions have been made. The process includes recommendations for the use of specific QA tools and highlights the need for QA at three stages of the commissioning process, taking into account the need to:

- Assess evidence according to established scientific quality criteria,
- To present a summary of this in a format that can support the commissioning decision-making process, and
- To assess the quality of the evidence in the context of the decision being made.

The process and assessment methods are detailed below, which also highlights who and when the different assessments are made.

5.2 Development of a hybrid approach

The quality assessment of evidence should be considered systematically, using consistent methods, at three points of the commissioning process. Each of the three stages would serve a specific purpose in the overall process and would bring together the use of different types of quality assessment tool.



Specific recommendations for each stage of quality assessment are given below.

5.3 Recommendations

Stage 1: Evidence review prior to the CPAG

Individual studies included in the evidence review should be assessed using study specific quality assessment checklists.

Recommended QA tool: This stage of quality assessment falls outside the scope of the current project and therefore, no specific recommendations are made. The review team has knowledge of the types of QA tool that would be appropriate and can discuss this with NHS England if the recommendation to include this stage of quality assessment were to be adopted.

Stage 2: Evidence summary documents prior to the CPAG

A clear summary of the strength of the body of evidence should be presented in the 'Evidence Base' section of the Commissioning policy document. The summary should be based on the

quality assessment of individual studies in the evidence review (as recommended above). The summary could be developed as part of the evidence review process and extracted to form part of the Clinical Commissioning Policy document.

Recommended QA tool: NSF-LTC. Consideration should be given to providing more guidance on how to score questions relating to study design and methods.

Stage 3: Consideration of strength of evidence by the CPAG

A discussion on the strength of the body of evidence and its impact on the commissioning decision should be transparent and systematic. This should be summarised in the 'Rationale behind the policy statement' section of the Commissioning policy document.

Recommended QA tool: Onakpoya's adaptation of GRADE. A discussion incorporating the factors identified by GRADE for the upgrading or downgrading of evidence is recommended. These include risk of bias, inconsistency, indirectness, imprecision, publication bias and effect size. It might be helpful to summarise this discussion using the categories used by GRADE and adapted by Onakpoya (high, moderate, low, very low) to indicate impact of the quality of evidence on the confidence in the estimate of effect. As highlighted earlier, further adaptations of this to make it more relevant to the precise decision making context of Specialised Commissioning could be considered. Further engagement with NHS England, and potentially further research, would be required to develop any adaptations.

6. LITERATURE SEARCHES RELATING TO VALUE ATTRIBUTES

6.1 Introduction

Three approaches were used in order to produce a rapid review of relevant attributes used in previous MCDA studies in health care. The methods used different search strategies and each targeted at different parts of the knowledge base that were known to represent well developed MCDA research with a focus on HTA. The methods were:

1. Identification of papers cited in the forthcoming ISPOR Good Practice Guidelines on MCDA as reporting value attributes. The Guidelines have been drawn together by a prestigious international group of researchers and includes a substantial set of cited studies. Whilst the citations may not have been systematically identified, they are expected to represent the most important papers relating to the application of MCDA to HTA.
2. Identification of journal articles and grey literature relating to value based pricing (VBP) or value based assessment (VBA). It is known that the discussions relating to VBP and VBA generated a lot of interest in MCDA methods and so a search on these topics was considered to be useful to identify contemporary, UK policy-relevant papers.
3. Identification of MCDA studies not captured by the previous two searches. The very nature of rapid reviews means that there is a danger that all relevant evidence sources are not captured. Consequently, we reviewed the full set of papers identified by the first two approaches and added missing evidence that were known to us.

Once all value criteria had been identified, those studies that included cost-effectiveness as a criterion were read in greater detail in order to assess the relevance of the approach used to measure performance against that criterion.

6.2 Methods

Review 1: Identification of papers cited in the forthcoming ISPOR GPG on MCDA

The ISPOR Guidelines included two systematic reviews of MCDA relating to health technologies [36, 37]. Both reviews used a previous classification of healthcare decision criteria developed by Guindo et al [38]. The authors of Cromwell et al [36] were contacted and their 'criteria matrix', which listed the Guindo criteria and recorded their use in the individual studies, was adapted for the purposes of this study. The principal adaptation was the inclusion of additional studies from the Wahlster review [37], plus the inclusion of

studies from the two other reviews undertaken for this project (described below). This is referred to as the consolidated criteria matrix henceforth. The Cromwell review also included programme budgeting and marginal analysis studies. Whilst these studies use evaluation criteria to identify services for investment and disinvestment, their decision making contexts are quite different from that of Specialised Commissioning. Consequently, these studies are excluded from our review.

Wahlster included 22 MCDAs whilst Cromwell included 19 MCDAs. With eight studies included in both, this left 33 unique studies.

Identification of papers relating to value based pricing or value based assessment

A focused search was conducted on Medline (via Ovid) using terms relating to value based pricing and value based assessment. The search retrieved 392 unique references. The complete search strategy can be found in Appendix 4.

In addition, grey literature was retrieved via a number of online sources:

- International Society for Pharmacoeconomics and Outcomes Research (ISPOR)
- World Health Organization (WHO)
- Health Technology Assessment International (HTAi)
- Organisation for Economic Co-Operation and Development (OECD)
- European Federation of Pharmaceutical Industries and Associations (EFPIA)
- European Public Health Alliance (EPHA)
- Directorate General for Health & Food Safety
- European Commission
- EUnetHTA
- European Medicines Agency (EMA)
- Association of the British Pharmaceutical Industry (ABPI)

The websites of various European bodies (including NICE, HAS, iQWIG, AIFA, AETS, LBI-HTA, SBU, NHIF) were also searched for any relevant grey literature. Ninety-two references were retrieved from the grey literature search.

The aim of the data extraction for these sources was to identity work that included the identification of value criteria in the context of Value-based assessment or Value-based pricing in the UK. From the 392 Medline papers, 370 were removed after reading the title and 13 were removed after reading the abstract. The majority of papers were removed because their subject was value-based purchasing in the United States' Medicare program. This left nine papers which were read in full. After this, eight were removed after reading the

full paper and one paper included, which was added to the consolidated criteria matrix. The eight papers were removed because although they discussed VBA or VBP, they did not suggest ways of measuring performance against these criteria. It is also worth noting that these papers did not propose value criteria that went beyond those identified by the Department of Health (i.e. burden of disease, therapeutic improvement, innovation and wider societal benefits).

From the grey literature 92 references were reviewed. The references were posters, reports, or working papers. All were excluded after review.

Identification of MCDA studies not captured by the previous two searches

The ScHARR project team were aware of two other initiatives that used MCDA techniques to assess the value of health technologies; the NHS Cancer Drugs Fund Prioritisation Tool [23] and the American Society of Clinical Oncologists Value Framework [39]. In addition, one other paper was known to the team [40] which was added to the consolidated criteria matrix.

Review of cost-effectiveness criteria

Each paper that included cost-effectiveness as a criterion was read in detail and the method for measuring performance against the criterion described. Those methods that were considered relevant to the NHS England decision making context were then identified, with the intention that they be applied to the three case studies. Studies were excluded if they:

- used incremental cost-effectiveness ratios
- used Quality Adjusted Life Years (QALYs)

These exclusion criteria were specified by NHS England as the level of information and analysis required to generate these statistics for all topics is unlikely to be available.

From the 18 papers read 16 were excluded (Appendix 5). Two of these papers were removed because they did not provide a scale for cost-effectiveness. All others were removed because they required a calculation of (incremental) cost-per-QALY or cost-per-DALY.

6.3 Results

Identification and classification of value criteria

The consolidated criteria matrix is shown in Appendix 6, and shows which criteria were used in which paper. Note that some studies may have included more than one criterion for any particular category; as such, the totals represent the number of studies that included a

category (and not the total number of criteria). The criteria used in the most number of studies are given in Table 3.

Table 3: Most frequent value criteria used

Criterion	**Frequency***
Budgetary impact/affordability	17
Cost-effectiveness	17
Effectiveness of technology	15
Disease characteristics/severity	14
Quality of evidence	12
Safety	11
Equity/reducing inequalities	10
Current burden (morbidity and mortality)	10
Magnitude of benefit (number of patients)	10
Age/risk of target group	9**

* There were 36 papers in total.

** 'Miscellaneous' criteria were also used in 9 papers.

Assessment of performance against cost-effectiveness criteria

The CDF draft consultation document (NHS, 2014), written by the Chemotherapy Clinical Reference Group (CCRG), contained a cost attribute that included the joint consideration of effectiveness and cost (Table 4). However, in subsequent CDF documents, this scale is replaced by a consideration of "the median cost of the drug under evaluation" [23] . The degree to which the 2014 document and its associated cost scale was implemented is unclear as all published applications of the CDF process (which include the points assigned to each criterion), do not include points relating to cost. Consequently, to avoid confusion with the

CDF process that has generally been applied, the cost-effectiveness scale in Table 4 is referred to as the CCRG scale.

Table 4: The Chemotherapy Clinical Reference Group scale

Criteria	Score
Superior efficacy and cost saving compared to currently used alternative	3
Cost saving and non-inferior to current equivalent	2
Cost neutral and non-inferior to current equivalent but has other advantages less toxic, oral administration	1
If no QALY available and costs more than current alternative	0

Diaby & Lachaine created a four level cost-effectiveness classification (very cost-effective, cost-effective, low cost-effectiveness, not cost-effective)[41]. This classification was created in three steps. First interventions were classified into four cost categories (very expensive, expensive, relatively expensive and not expensive). All interventions were assigned into each category so that each category contained the same number of interventions. Second, medical benefits were classified in four categories (highly significant medical benefit, significant medical benefit, relatively significant medical benefit, and non-significant medical benefit). Third, each combination of a cost and benefit category was assigned to one of the four cost-effectiveness classifications (Table 5). No information was provided on how cost-effectiveness levels were allocated to cost-effect pairings. Each category of cost-effectiveness was rated by experts with scores of 1 (very cost-effective), 0.75 (cost-effective), 0.5 (low cost-effectiveness), and 0.25 (not cost-effective).

Table 5: The Diaby & Lachaine scale

	Highly significant medical benefit	Significant medical benefit	Relatively significant medical benefit	Non-significant medical benefit
Not expensive	+++	++	+	-
Relatively expensive	+++	++	+	-
Expensive	++	+	+	-
Very expensive	+	+	-	-

+++ indicates very cost-effective, ++ indicates cost-effective, + indicates low cost-effectiveness, - indicates not-effective

6.4 Discussion and conclusions

Value criteria

The identification and classification of value criteria made use of two pre-existing systematic reviews that used a classification system. It is possible that there is some inconsistency in the classification of criteria between Cromwell [36], Wahlster[37] and the additional studies that we identified. However, we are not able to identify and resolve any such issues within this project. In particular, the level of reporting of the criteria within the supplementary material relating to the Wahlster paper, is much less than that for Cromwell. Indeed, two identified papers were missing any classification of criteria [42, 43].

However, this is not expected to impact on our work substantially as we were able to identify a wide range of methods that were used to measure cost-effectiveness that could inform our recommendations to NHS England. Also, qualitatively, it appears that the reporting of any cost-effectiveness criteria were more comprehensive within the Wahlster paper than for the other criteria.

Cost-effectiveness criterion

Eighteen papers were identified as having a cost-effectiveness criterion, however, only 2 papers measured cost-effectiveness using methods that did not require the calculation of an incremental cost-effectiveness ratio. Two papers did not describe their methods in sufficient detail to pass comment on.

Measurement of cost-effectiveness within the CCRG framework is based on a simple four category ordinal scale (Table 4). The scale has limited face validity, as it is quite possible that if the net monetary benefits were calculated, the ordering of the three highest scoring categories could change. For example, 'cost-neutral with reduced toxicities' may be more cost-effective than 'cost-saving and clinically superior' if the toxicities are severe, the cost-savings are marginal and the clinical superiority is marginal or translates into negligible QALY gains. However, it could be argued that these circumstances would be few and far between, and as such, the scale is reasonable in the main.

Measurement of cost-effectiveness within the Diaby paper requires separate assessments of cost and effectiveness. The assessment of cost is based on the distribution of costs across all programmes within the prioritisation process; schemes are ranked and divided into quartiles, then allocated to the four cost categories, ranging from 'not expensive' to 'very expensive'. The assessment of effectiveness is based on four categories, however, no detail is given about how programmes are allocated to these. The cost-effectiveness scale has limited face validity, as again, the order of actual cost-effectiveness may differ from the ordinal rankings dictated by the MCDA. For example, it is quite easy to see a programme classified as 'not expensive' and 'relatively significant medical benefit' being more cost-effective than 'relatively expensive' and 'significant medical benefit'. The classification of cost by quartiles makes the measurement properties of this scale even less predictable.

The practical aspects of applying these two scales will now be assessed in the case study work.

7. APPLICATION OF COST-EFFECTIVENESS SCALES TO THE CASE STUDIES

7.1 Introduction

Three case studies were selected by NHS England to represent a range of topics and types of evidence base. The case studies were:

- Robotic-assisted surgical (RAS) procedures for prostate cancer
- Treatment as prevention in HIV (HIV TasP)
- Rituximab for the treatment of steroid resistant nephrotic syndrome (SRND) in paediatrics and rituximab for the treatment of relapsing steroid sensitive nephrotic syndrome (SSNS) in paediatrics

An additional study was subsequently added, as the nature of the SRND/SSNS case study did not allow the two scales to be successfully applied (although our attempt is reported). The additional case study was:

- The use of rituximab as a second line agent for the eradication of inhibitors in patients with acquired haemophilia (R2AH)

7.2 Methods

The two methods identified in the literature review – CCRG and Diaby - were applied to each of the four case studies independently by two reviewers (SD and MK). The information available for the case studies represented the information that was presented to CPAG. The basis of each reviewer's classification was noted. In addition, other issues that were flagged up by the case study work are highlighted.

With only four case studies, levels of agreement cannot be assessed quantitatively in any meaningful way. As such, this is a qualitative assessment using the criteria used for the quality assessment work reported earlier; authority, consistency, relevance and workability.[11]

[11] Authority is the degree to which there is a body of evidence and high profile endorsement of the method. In addition to how this criterion was used for the QA work, we also consider theoretical validity under this. Consistency is the degree to which the method can be used across different topics within NHS England's portfolio and is aligned, conceptually, to methods used elsewhere in England. Relevance is the degree to which the methods are aligned to the type of evidence encountered and the needs of the Specialised Commissioning process. Workability is the degree to which the method can be used, practically, within a time-constrained process.

7.3 Results

None of the case studies could be assigned to the cost scale within Diaby's cost-effectiveness criterion as it requires a distribution of costs across programmes, then the identification of quartiles. With three case studies, this is impossible.[12] The reviewers classifications for the Diaby and CCRG scales are given in Tables 6 and 7, respectively.

Table 6: Summary of ratings for cost-effectiveness using the Diaby scale (effectiveness only)

	RAS		R2AH		HIV TasP	
	Rev 1	Rev 2	Rev 1	Rev 2	Rev 1	Rev 2
Highly significant benefit						✓
Significant medical benefit			✓		✓	
Relatively significant medical benefit	✓	✓		✓		
Non-significant medical benefit						

Rev= Reviewer

Table 7: Summary of ratings for cost-effectiveness using the CCRG scale

	RAS		R2AH		HIV TasP	
	Rev 1	Rev 2	Rev1	Rev 2	Rev 1	Rev 2
Superior efficacy and cost saving compared to currently used alternative			✓		✓	✓
Cost saving and non-inferior to current equivalent						
Cost neutral and non-inferior to current equivalent but has other advantages less toxic, oral administration						
If no QALY available and costs more than current alternative	✓	✓		✓		

[12] Although it is possible that this distribution could be applied to an annual prioritisation process where tens of topics are considered.

	RAS		R2AH		HIV TasP	
	Rev 1	Rev 2	Rev1	Rev 2	Rev 1	Rev 2

Rev= Reviewer

Classification for SRNS/SSNS was not possible. The proposed policy involved the transfer of an existing service funded by one part of NHS England to the NHS Specialised Commissioning budget. Consequently, the comparator for the service is the same service; cost neutrality and equivalent outcomes are virtually guaranteed. If the cost-effectiveness of rituximab for SRNS/SSNS is required, then an alternative treatment needs to be specified (which could be 'no rituximab).

Classification of RAS, R2AH and HIV TasP topics was possible, however it was rather subjective. The sources of information that were considered most pertinent by the reviewers are shown in Appendix 7. These and other issues that need consideration in any future assessment of cost-effectiveness within an MCDA are listed below:

- RAS has two comparators – laparoscopic radical prostatectomy and open radical prostatectomy – therefore any classification becomes a qualitatively weighted assessment of the two evidence bases.
- SRNS/SSNS relates to two distinct indications, therefore any classification becomes a qualitatively weighted assessment of the two indications.
- Within the evidence presented to CPAG, there is a mix of findings (positive and negative) and types of evidence, with no central estimate of effect.
- Within the evidence presented to CPAG, reporting of effect size and statistical significance is highly variable; sometimes given, sometimes not.
- Within the evidence presented to CPAG, clinical significance or standardised effect sizes are not reported, so the importance of effects is qualitatively assessed by each reviewer.
- Within the evidence presented to CPAG, it is not clear whether costs are absolute or incremental to current treatments.
- Within the evidence presented to CPAG, the comparators in the trials are not reported clearly.
- Within the evidence presented to CPAG, quality of the economic evidence is taken on trust. For example, TasP cost-effectiveness is largely predicated on the very high costs of lifetime HIV treatment (c£280K-£360K), yet another study is cited that uses much lower estimates, but which was largely ignored.

- A policy decision may be made based on issues outside the normal process (e.g. SRND/SSNS), therefore the available information does not always match the required information.
- For SRND/SSNS and R2AH, the economic evidence related to budget impact and not cost. For SRND/SSNS, rituximab treatment was considered cost neutral as treatment with rituximab was already paid from another part of the NHS budget. For R2AH, part of the additional cost was already paid from another part of the NHS budget. These data are irrelevant for an assessment of cost-effectiveness. The costs should be incremental to the alternative treatments (e.g. not getting rituximab) and not the overall impact on the budget.

Assessment of the scales against the evaluation criteria

The CCRG scale has limited authority as it lacks face validity (as highlighted earlier). It can be consistently applied to a range of topics, but has little relevance as the vast majority of technologies are expected to fall within the category of "If no QALY available and costs more than current alternative". The scale would, however, be easy to apply within the prioritisation process.

The Diaby scale has very limited authority as it is highly subjective and lacks some face validity. The effectiveness scale will be difficult to apply consistency across topics due to its subjectivity. The cost scale would be applied consistently within any set of programmes, but would potentially be inconsistent between sets of programmes assessed at different times. The scale is relevant, essentially because its vagueness would mean that it can be applied to anything and this means that it would be easy to apply within the prioritisation process.

7.4 Discussion

Several other issues were highlighted by the case study work:

- Whilst it is accepted that the evidence base for services related to Specialised Commissioning may be weaker, the reporting of the evidence to the CPAG of the evidence that was available was flawed. Examples highlighted above include the reporting of absolute levels of cost and effectiveness in the new treatment group, without reference to the levels observed in the control group. Whatever scale is used in future, the reporting of evidence must be improved in order to facilitate the measurement of performance on that scale.
- Differences in interpretation of the evidence between the reviewers was apparent. For example, one reviewer had a tendency to dismiss effects that were not statistically

significant, whilst the other was more forgiving. More generally, the approach to 'marking down' poor quality evidence differed between reviewers. This was, in part, due to the nature of the exercise, which did not include the development of guidance for the reviewers and training. Whatever scale is used in future, clear guidance on measurement, and training for CPAG members will be required.

- Neither of the scales linked well the common approach of assessing cost-effectiveness in terms of incremental costs and incremental effects. Whilst the CCRG scale can be interpreted as such, it was unable to differentiate between services that had both positive incremental costs and incremental benefits. As highlighted earlier, this situation is expected to be quite common among the services to be evaluated by the CPAG. This 'north-east quadrant problem' is discussed in greater detail in the next section where we attempt to create a scale that is able to differentiate between services of this kind.
- The extent to which assessment against this criterion should take into account the quality of the economic evidence, and hence its inherent uncertainty, needs to be considered once the wider MCDA process is known. In particular, will ther bean 'uncertainty' criterion and/or will there be uncertainty/sensitivity analysis?

8. BUILDING ON THE CASE STUDY WORK

8.1 Introduction

The review and case study work suggest that the Chemotherapy Clinical Reference Group (CCRG) cost-effectiveness scale is the best available scale that has been used in previous healthcare MCDAs and which meets the requirements of the Specialised Commissioning process. However, it is quite crude and does not address the 'north-east quadrant problem' (valuing better outcomes achieved at a higher cost). In this section we examine the measurement of cost-effectiveness from first principles and develop other scales that are capable of differentiating between programmes in the north-east quadrant of the cost-effectiveness plane. These are then assessed qualitatively using the same criteria as before.

8.2 Measuring cost-effectiveness

The cost-effectiveness plane is the most frequently used approach to illustrating the measurement of cost effectiveness. Shown in Figure 1, the plane has incremental costs on the y-axis and incremental effectiveness on the x-axis, generating a graph with four quadrants. The quadrants – labelled by their analogous geographical positions – represent programmes with distinct combinations of incremental costs and benefits:

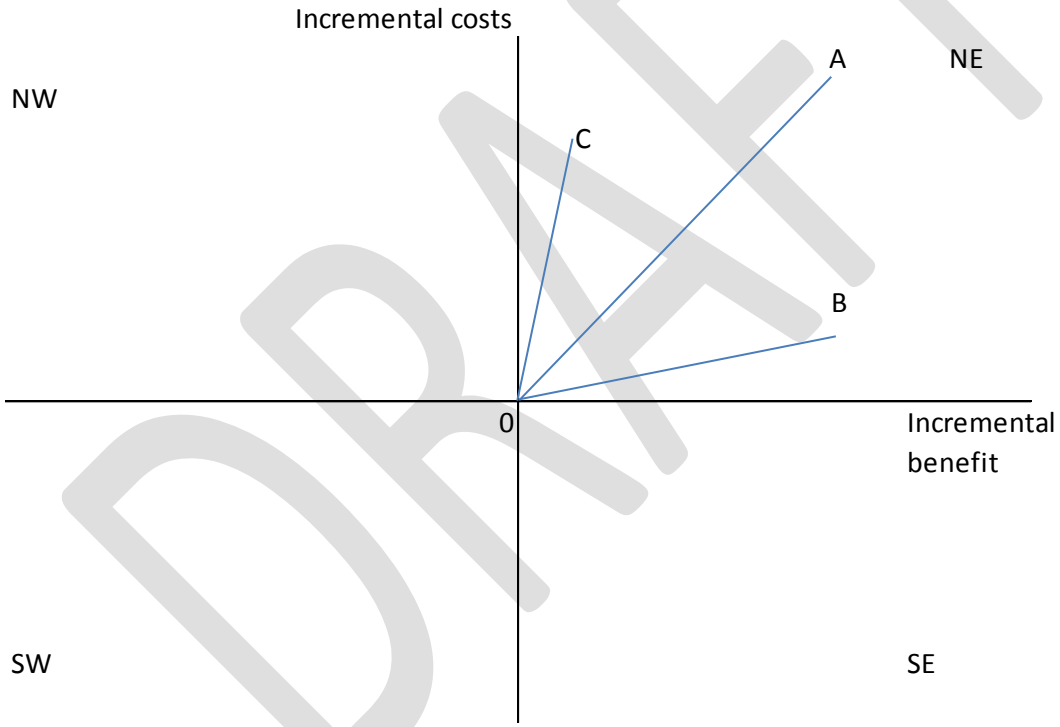
- North-west (NW) quadrant, represents programmes with additional costs and reduced benefits. Such programmes are unequivocally not cost-effective.
- South-east (SE) quadrant, represents programmes with reduced costs and additional benefits. Such programmes are unequivocally cost-effective.
- South-west (SW) quadrant, represents programmes with reduced costs and reduced benefits. The new programmes may be cost-effective, as the released costs could be re-allocated to other programmes in a way that would more than offset the reduced benefits generated by the new programmes.
- North-east (NE) quadrant, represents programmes with additional costs and additional benefits. The new programmes may be cost-effective, as the increased benefits could more than offset the additional costs and necessary disinvestments elsewhere.

It should be noted that this description is based on deterministic assessments of cost-effectiveness. Within a probabilistic framework, and in economic evaluations that undertake sensitivity analyses more generally, cost-effectiveness is represented by an area that may span more than one quadrant [44]. In these circumstances, the conclusions are less clear

than those described above. For the rest of this report, however, we will assume that a deterministic framework without sensitivity analysis is being adopted for the assessment of cost-effectiveness by NHS England. This is not to say that uncertainty should not be considered at all; it's just that we recognise that uncertainty about the results is likely to be considered elsewhere in the decision making process.

Furthermore, it is likely that the context within which the Specialised Commissioning Programme currently operates will make the relevance of the SW quadrant questionable. Consequently, despite the theoretical validity of choosing programmes that are located in the SW quadrant [45], we will ignore this possibility for the rest of this report.

Figure 1: The cost-effectiveness plane



The cost-effectiveness plane can be used to identify two points on a cost-effectiveness scale that have unequivocal rankings. SE is best and NW is worst. The ranking in respect of the NE quadrant is dependent on the relative magnitudes of costs and effects. For example, a programme in the north-east quadrant may produce a greater net monetary benefit (i.e. be more cost-effective) than another programme in the south east quadrant. Whilst an unequivocal ranking is possible that would produce a measurement scale that could be used within a MCDA, its two categories do not address the NE quadrant problem.

Assessing cost-effectiveness in the NE quadrant is simple when costs and effectiveness are measured on an interval scale. Firstly, a ratio of incremental costs divided by incremental effectiveness is calculated – represented by the line passing through the origin and the point on the plane depicting the relevant programme – for example, A, B and C in Figure 1. Secondly, a judgement about relative cost-effectiveness can be made based on the slopes of two or more lines representing competing programmes – for example, we can say that B is cost-effective relative to A, which is more cost-effective than C. Finally, a judgement about absolute cost-effectiveness can be made based on the magnitude of the ICERs relative of an incremental cost/value threshold.

However, an appropriate effectiveness metric to allow this conventional approach to assessing cost-effectiveness is not available within the context of the Specialised Commissioning decision making process; QALYs and ICERs have been explicitly ruled out. This forces us to consider an alternative scale of cost-effectiveness. The key issue with this, is that magnitudes of incremental costs and effectiveness need to be measured in order for the cost-effectiveness of different programmes to be differentiated from one another.

In the context of the cost-effectiveness plane, we want to be able to differentiate between A, B and C. Given the lack of detailed evidence for many of the programmes that are assessed by NHS England, we need to be able to do this based on simple measures of incremental cost and benefit. Several possibilities were considered for the measurement of benefits and the development of ordinal categories.

1. Use life-years gained, on the basis that this is a fundamentally important outcome measure, for which data is almost always available. The main problem with this is that it produces a bias against programmes that produce mainly or only quality of life benefits.
2. Use a Likert-type scale, for example, small improvements in health gain, moderate improvements and large improvements. This mirrors the approach adopted by Diaby. The main problems with this are that it is highly subjective and purely ordinal. Consequently, when combined with the incremental cost categories, a clear ranking of programmes is not possible.
3. Use a Likert-type scale, as above, but with examples of health gains in order to make measurement of programmes less subjective and approximately interval. For example, 'small improvements in health gain, e.g. 3 months increase in survival, or a permanent 10% increase in quality of life'. Whilst this approach will make measurement less subjective, it will be difficult to generate a set of valid descriptors, for example, what is a 10% increase in quality of life, or what quality of life will

someone experience in their 3 additional months of life? Given development time, a more valid set of descriptors could be produced. However, it is felt that the level of complexity required, and the inherent limitations of this approach, would leave this approach lacking any authority.

4. Use the concept of minimum clinically important difference (MCID) as a 'unit of health gain', such that gains could be measured in multiples of MCID relevant to each particular programme. This has the advantages of having interval properties and having a (weak) link through to utility in that two studies have estimated MCID to range between 0.03 and 0.07 on the health-related utility scale [46, 47]. The link through to utility would facilitate the estimation of QALYs and the potential to derive an approximate incremental cost per QALY gained. This approach was taken forward as the most promising approach.
5. Use of MCID, as above, but with additional descriptors to capture other health effects that fall outside the outcome measure to which the MCID refers. These health effects may be additional benefits or adverse events, and consequently, a net estimate would need to be approximated. To keep this manageable, the scale assumes that positive net effects improve cost-effectiveness by one category, whilst negative net effects reduce cost-effectiveness by one category. This is clearly a simplification, but without careful measurement of all of these additional health effects on a scale that is compatible with the MCID of the primary outcome measure, few other approaches are available.

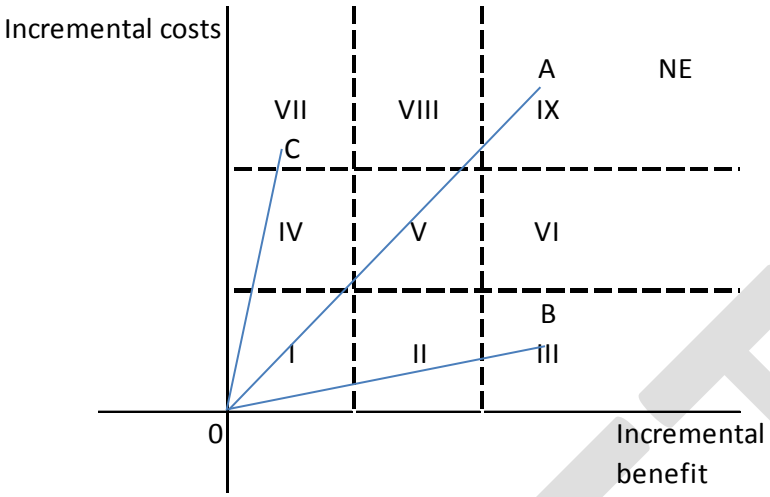
In addition, two possible approaches to measuring incremental cost and the development of ordinal categories were considered. Firstly, incremental cost could be categorised in terms of absolute cost, e.g. <£1000 per patient more, or ≥£1000 per patient more. Secondly, incremental cost could be categorised in terms of proportional cost, e.g. <5% increase in costs, or ≥5% increase in costs. The advantage of the first approach is that it is clearly linked to the cost-effectiveness plane and the calculation of ICERs. However, it could be argued that it may disadvantage specialist, high-cost programmes. The first approach was taken forward based on its link to accepted cost-effectiveness methods and the expectation that the value of specialist, high-cost programmes will be accounted for elsewhere in NHS England's MCDA.

Developing a NE quadrant scale

The effect and cost dimensions of the proposed scale are MCID and total NHS costs. These are then categorised to produce a finite set of cost-effectiveness categories that are ranked in terms of the relative cost-effectiveness of the central point of the category. As shown in

Figure 2, the ranking of the 9 categories from most cost-effective to least cost-effective is III, II, VI, followed by I, V and IX tied, then VIII, IV and VII.

Figure 2: Cost and effect categories within the NE quadrant



Clearly, whilst this ranking is correct for the centre points of the nine areas within the NE quadrant, they are not correct for all points within the areas. A low cost point in area II will be more cost-effective than a high cost point in area III. However, such problems are inevitable if we are to move away from the simple two category ranking based on the quadrants.

To complete the scale, the boundaries on the categories need to be identified. It is proposed that the incremental effectiveness categories should be based on MCID:

- < 1 x MCID
- between or equal to 2 x MCID and 3 x MCID
- > 3 x MCID.

It is proposed that the incremental cost categories should be based on £900 per patient per annum bands. Based on an approximate MCID of 0.03, the mid-point of area I will produce an ICER of £30,000 per QALY gained (£450/0.015 QALYs, assuming costs are annual and a permanent utility gain).[13]

[13] Note that other utility values for MCID were produced, and therefore, other cost categories would be needed to match the £30,000 threshold.

Developing an extended NE quadrant scale

The preceding NE quadrant scale implicitly assumes that the only effect of any value is that to which the MCID refers. So, if the MCID relates to a reduction in pain, then any gain in physical functioning or anxiety/depression will be not included in the measurement of cost-effectiveness.[14] Two approaches were considered with respect to this. Firstly, these additional effects are ignored, but that when allocating a programme to a category, a judgement is made about whether to amend the measurement based on MCID alone. Secondly, these additional effects are added to the descriptive system of the measurement scale. For example, Area I would refer to "less than one MCID health gain and less than an additional £900 per patient per annum, or less than one MCID health gain, £900-£1800 additional costs per annum, plus other positive net health effects (i.e. gains minus adverse events).

8.3 Comparison of possible scales

Four scales were compared qualitatively in terms of authority, consistency, relevance and workability. They were the simple three quadrant scale, the CCRG scale, the NE quadrant scale with additions from the CCRG scale to recognise combinations of costs and effects outside the NE quadrant (which we refer to as the 'expanded north east quadrant scale'), and an expanded north east quadrant scale that allows for multiple health effects. The four scales are shown in Table 8, and aligned in rows where they match (although sometimes this is only an approximate match).

[14] This point is debatable. Some of the estimated utility values for MCID were based on generic scales, which include multiple health domains. The degree to which these are picked up this methodology is unclear.

Table 8: Four possible cost-effectiveness scales

Cost-effectiveness plane quadrants	CCRG scale	Expanded north east quadrant scale	Expanded north east quadrant scale with multiple health effects	Cost-effectiveness plane code#
1. Less effective and costs more		1*.Less effective and costs more	1.Less effective and costs more	NW
2. More effective and costs more	1. If no QALY available and costs more than current alternative	2*. Less than 1 times MCID better and more than£1800 pppa more expensive	2. As per 3*, but with other negative net health effects.	NE VII
		3*. Less than 1 times MCID better and £900-£1800 pppa more expensive	3. As per 4*, but with other negative net health effects. OR, as per 2* with other positive net health effects.	NE IV
		4*. Between 1 and 2 times MCID better and more than £1800 pppa more expensive	4. As per 5*, but with other negative net health effects. OR, as per 3* with other positive net health effects.	NE VIII
	2. Cost neutral and non-inferior to current equivalent but has other advantages less toxic, oral administration	5*. Less than 1 times MCID better and less than £900 pppa more expensive, OR between 1-2 times MCID better and £900-£1800 pppa more expensive, OR more than 2 times MCID better and more than £1800 pppa more expensive	5. As per 6*, but with other negative net health effects. OR, as per 4* with other positive net health effects.	NE I, V or IX

 # Letters refer to the quadrants in Figure 1. Roman numerals refer to the areas in Figure 2.
pppa = per patient per annum.

Table 8: Four possible cost-effectiveness scales (cont…)

Cost-effectiveness plane quadrants	CCRG scale	Expanded north east quadrant scale	Expanded north east quadrant scale with multiple health effects	Cost-effectiveness plane code#
		6*. More than 2 times MCID better and £900-£1800 pppa more expensive	6. As per 7*, but with other negative net health effects. OR, as per 5* with other positive net health effects.	NE VI
		7*. Between 1 and 2 times MCID better and less than £900 pppa more expensive	7. As per 8*, but with other negative net health effects. OR, as per 6* with other positive net health effects.	NEII
		8*. More than 2 times MCID better and less than £900 pppa more expensive	8. As per 9*, but with other negative net health effects. OR, as per 7* with other positive net health effects.	NE III
			9. As per 8* with other positive net health effects.	-
	3. Cost saving and non-inferior to current equivalent	9*. Non-inferior and costs less	10. Non-inferior and costs less	-
3. More effective and costs less	4. Superior efficacy and cost saving compared to currently used alternative	10*. More effective and costs less	11. More effective and costs less	SE

 # Letters refer to the quadrants in Figure 1. Roman numerals refer to the areas in Figure 2.
pppa = per patient per annum.

The scale based on the cost-effectiveness plane quadrants is highly authoritative, as the ranking is universally agreed. It can be consistently applied to a range of topics, but has little relevance as the vast majority of technologies are expected to fall outside the scale. As such, it will be of little use in differentiating between the cost-effectiveness of different programmes. The scale would, however, be easy to apply within the prioritisation process.

The CCRG scale was assessed in the case study section of this report.

The ENEQ scale has limited authority, for although it is tied to the cost-effectiveness plane and (very weakly) to QALYs and the NICE threshold, there are likely to be many situations where the categorisation will lack face validity. It is also completely untested, either empirically or qualitatively (outside the research team). Additionally, estimates of MCID vary between studies and so any chosen value will be open to dispute. The use of a universal measure of effect – MCID – means that it can be applied to a wide range of topics. However, the MCID will not be known for many services, consequently, the approach lacks relevance to the prioritisation process. Additionally, the focus on the primary outcome measure of interest limits its relevance to the evaluation of complex conditions. The relevance of the scale, however, is enhanced by its number of categories compared to the CCRG scale. The scale could be difficult to apply within the prioritisation process as additional searches and evidence synthesis would be required to identify the best estimate of the MCID, then applying this to each of the effectiveness studies.

The ENEQ with multiple health effects scale is essentially the same as the ENEQ, but will command less authority due to the addition of a subjective and unvalidated dimension (i.e. the other health effects). Its relevance will be increased due to its ability to capture multiple health effects. In practical terms, the benefits of the extended scale may be very slight as its application to evidence will still need a substantial amount of judgement in order to assess whether the net effects are positive or negative, and whether the re-categorisation relative to the basic ENEQ is correct.

On balance, the ENEQ appears to strike a balance between the need for a granular scale that addresses the NE quadrant problem, and the practicalities of the Specialised Commissioning process. However, it needs empirical and stakeholder testing, to assess its authority, relevance and workability. We were unable to apply the ENEQ to the four case studies as MCID estimates were not available in the CPAG documentation.

8.4 An alternative approach to incorporating cost-effectiveness in an MCDA

Whilst a large proportion of MCDAs in healthcare include cost as criterion, it has been argued that these approaches do not adequately capture the opportunity cost of alternatives [48]. This is because opportunity cost is represented by the shadow price of disinvestment within the relevant budget, and as such, is dictated by the cost-effectiveness of existing treatments at the margin. So whilst the inclusion of cost (or cost-effectiveness) does recognise the existence of a trade-off, it is very unlikely that the combination of scoring,

weighting and aggregation developed within an MCDA, will reflect the actual opportunity cost of displaced activities in the NHS.

More formally, it is known that to maximise the value of outputs within a fixed budget, investments should only be made when:

(Value x Lambda) – Cost >0, where Lambda =Cost-effectiveness of displaced treatments (1)

A special case of Equation (1) is where value is represented by QALYs and Lambda is the incremental cost per QALY gained of the displaced treatments. A broader representation of value is possible (e.g. generated by MCDA), which in turn, would require a new value of Lambda. However, the investment rule remains, if value maximisation is desired.

Now consider a value measurement MCDA. Value is estimated in the following way:

$Value = \sum W_i \ P_{ij}$, *where W is weight for criterion i, and P is performance of option j (2)*

Clearly, adding cost into Equation 2 as a criterion, then making investment decisions based on an overall value (or relative performance in the case of a prioritisation process) is not equivalent to Equation 1.

Consequently, in order to avoid a potential mismatch between an MCDA and value maximisation within a fixed budget it has been argued that cost should not be included as criterion. Instead, it is argued that MCDA should be used to generate an estimate of aggregate value, which is then combined with cost information in an additional step to the MCDA.

The identification or estimation of a threshold is not part of our work. However, one method that would appear appropriate to this decision making context would be to identify a threshold through revealed preference, i.e. base it on an examination of past decisions using the chosen MCDA method. The range within which the threshold falls would be informed by the 'worst' approved technology (as defined by Equation 1) and the 'best' rejected technology (as defined by Equation 1). Additionally, this exercise could be tied to the NICE appraisal process by including TAs that are relevant to Specialised Commissioning, for example, hepatitis C treatments and rituximab for vasculitis, which are commissioned by NHS England. The merit of this approach, and indeed others that are possible, will depend on what NHS England consider to be the opportunity cost of investments to be, i.e. do the investments displace activity across the NHS or just those within the NHS England budget.

8.5 Recommendations

1. If NHS England wishes to include a cost-effectiveness scale within the value function generated by an MCDA, then the ENEQ scale is considered the most relevant to the Specialised Commissioning process. However, its two major drawbacks should be given further consideration before being adopted. Firstly, its authority/validity should be further tested both quantitatively and qualitatively. Secondly, the availability of MCID information and its relevance to the effectiveness evidence available to the CPAG should be assessed. One part of this should be to investigate the feasibility of MCID being reported routinely together with the other provided information. For example, if it requires additional searches its feasibility would be questioned, however, registration processes generally require an estimate of MCID, in which case, reporting would be straightforward.If any serious failings are encountered, then an extended MCDA approach should be adopted as that removes cost and cost-effectiveness from the value function.

2. If a cost-effectiveness scale is used, this needs to be undertaken in tandem with clear guidance on the use of the measurement scale and training of CPAG members. It is not possible to give complete guidance on the use of the ENEQ scale at present, as it will be partly dependent on the other criteria within the MCDA. For example, if quality of evidence and uncertainty are included as criteria, then assessment of cost-effectiveness may not need to take into account statistical significance of outcomes or details of research design.

3. In order for the CPAG to measure performance against any cost-effectiveness scale, then good practice for reporting economic evaluations should be followed. Whilst full reporting is probably not necessary, particular principles need to be clear:
    a. The perspective of the analysis needs to be stated
    b. Discounting of costs and outcomes has been undertaken appropriately
    c. Incremental analysis is undertaken based on the best available estimate, which requires,
        i. Total per patient NHS costs for the new service/therapy.[15]
        ii. Total per patient NHS costs for the 'old' therapy (or therapies).[15]
        iii. Per patient outcomes for the new service/therapy.
        iv. Per patient outcomes for the 'old' therapy.

[15] Ideally, this should relate to the time period over which there may be significant cost and health consequence. However, in light of the limited data seen in the case studies, this may not be possible and so some guidance on how this can be approximated.

v. Additionally, for the ENEQ scale, the patient outcomes need to be expressed in terms of multiples of MCID or capable of being so expressed.

d. Sensitivity analysis has been undertaken appropriately. At the very least this should be a deterministic analysis using alternative, plausible estimates.

e. Budget impact should not be considered (for this criterion).

4. On theoretical grounds, removing cost and cost-effectiveness from the value function is the preferred way forward as it is the most appropriate way to address the issue of opportunity cost. In practical terms, it also overcomes the problems encountered by the development and use of a cost-effectiveness scale for use within the estimation of the value function, e.g. overlap and preference interdependence of criteria, identifying a MCID that can be applied to all available data, etc.

5. If the extended MCDA approach is used, then consideration needs to be given to benchmarking the process against the appropriate measure of opportunity cost, e.g. NHS expenditure or NHS England expenditure.

6. Once the MCDA process has been formulated, the approach to assessing cost-effectiveness requires further review as some important issues could not be fully explored in this report due to the ongoing development process. For example, there are potential problems with overlap and preference interdependence relating to the cost-effectiveness criteria and there is uncertainty around how the quality of economic evidence will be factored into the MCDA and what sensitivity analyses will be undertaken.

Appendix 1: Key principles

Clinical effectiveness principles:

a) There must be adequate and clinically reliable evidence to demonstrate clinical effectiveness.
b) There must be a measurable benefit to patients.
c) The intervention should offer equal or greater benefit than other forms of care routinely commissioned by the NHS.
d) While considering the benefit of simulating innovation, NHS England will not confer higher priority to a treatment or intervention solely on the basis it is the only one available.

Fairness and equity principles:

a) NHS England may agree to fund interventions for rare conditions where there is limited published evidence on clinical effectiveness.
b) The intervention must be available to all patients within the same patient group (other than for clinical contra-indication).
c) The intervention should be likely to reduce health inequalities, and NHS England will have regard to any relevant broader equality issues.
d) The intervention should benefit the wider health and care system.
e) The intervention should advance parity between mental and physical health.

Financial principles:

a) The intervention should demonstrate value for money.
b) We will then apply the principle of affordability and only commission for those treatments and interventions that are affordable within the annual allocation to specialised commissioning, and those that enable resources to be released for reinvestment.

Appendix 2: QA search strategies

Review 1: QA of evidence in rare diseases

Medline

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) <1946 to Present>
Search Strategy:

---

1 orphan drug$.tw. (839)
2 rare disease$.tw. (14771)
3 special$ service$.tw. (2599)
4 1 or 2 or 3 (18003)
5 evidence.tw. (1273978)
6 reimburs$.tw. (19421)
7 commission$.tw. (29965)
8 5 or 6 or 7 (1318757)
9 4 and 8 (1563)
10 limit 9 to yr="2005 -Current" (1122)

Cochrane Library

Search Name: SpecComm
Last Saved: 29/10/2015 13:45:17.908
Description: Orphan drug etc. search

ID	Search
#1	"rare disease":ti,ab,kw (Word variations have been searched)
#2	"orphan drug"
#3	"special service"
#4	#1 or #2 or #3

Review 3: QA Methodological best practice

Medline

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) <1946 to Present>
Search Strategy:

---

1 quality assessment.tw. (10296)
2 critical appraisal.tw. (5083)
3 risk of bias.tw. (6646)
4 (grade$ or grading).tw. (320838)
5 (score$ or scoring).tw. (618528)
6 1 or 2 or 3 or 4 or 5 (917765)
7 systematic review$.tw. (73416)
8 meta-analy$.tw. (84686)
9 evidence.tw. (1281005)
10 recommend$.tw. (455390)
11 7 or 8 or 9 or 10 (1748042)
12 ((quality assessment or critical appraisal or risk of bias or (grade$ or grading) or (score$ or scoring)) and (systematic review$ or meta-analy$ or evidence or recommend$)).ti. (1379)
13 (method$ or tool$).tw. (4701999)
14 ((quality assessment or critical appraisal or risk of bias or (grade$ or grading) or (score$ or scoring)) adj3 (systematic review$ or meta-analy$ or evidence or recommend$)).tw. (7186)
15 ((quality assessment or critical appraisal or risk of bias or (grade$ or grading) or (score$ or scoring)) adj3 (systematic review$ or meta-analy$ or evidence or recommend$) adj3 (method$ or tool$)).tw. (322)
16 12 or 15 (1674)
17 limit 16 to yr="2010 -Current" (1034)

Journal searching

Journal titles

Systematic Reviews
Research Synthesis Methods
BMC Health Services Research
BMC Medical Research Methodology

Appendix 3: Detailed QA case study results

Case Study 1: Robotic-assisted surgical (RAS) procedures for prostate cancer

COMPASS	
Results from the tool	Assessment of the tool
The majority of the checklist was completed by ticking the 'not known' or 'not applicable' boxes Information on study design and whether trials were randomised allowed these sections of the checklist to be completed.	Individual study level information provided was not sufficient to complete the checklist.
OCEBM	
Results from the tool	Assessment of the tool
Seven different sources of evidence were referenced in the CCP. Of these, two were level two evidence and five were level three evidence.	Easy to apply as evidence types given in the CCP.
GRADE	
Results from the tool	Assessment of the tool
It was not possible to assess the evidence provided against the GRADE criteria for assessing the quality of evidence due to insufficient information included in the CCP.	Evidence cannot be categorised. In addition, evidence summaries, as recommended by GRADE are unfeasible. Therefore it is impossible to both make recommendations and determine the strength of these recommendations.
NSF-LTC	
Results from the tool	Assessment of the tool
Of the seven different sources of evidence referenced in the CCP, all could be assessed according to the NSF-LTC based on the information provided. They were a combination of P1 and R1 studies (design) scored between 0/10 and 3/10 for (quality) and in terms of applicability, were all direct evidence.	The CCP supplied sufficient information to apply this tool.

Case Study 2: Treatment as prevention in HIV

COMPASS	
Results from the tool	Assessment of the tool
It was not possible to complete the tool to give a meaningful result.	The majority of the tool was completed by ticking the 'not known' or 'not applicable' boxes as insufficient individual study level information was provided. Information on study design and whether trials were randomised allowed these sections of the checklist to be completed. Information was available on the study population for some of the included studies.
OECBM	
Results from the tool	Assessment of the tool
The evidence summary is limited in its description of the levels of evidence. There is only one RCT, a number of cohort studies and a meta-analysis of cohort studies.	As the evidence summary itself outlines, RCTs may not be the most appropriate form of evidence for a topic such as this, especially when comparing the intervention with no intervention.
GRADE	
Results from the tool	Assessment of the tool
The tool was not completed.	The use of GRADE in assessing the evidence provided for the use of TaP in HIV is challenging as limited RCT evidence exists, due to the known efficacy of the intervention and lack of a comparable intervention. There was insufficient information in the evidence summary to populate GRADE and in addition, the production of an evidence table would have been challenging as the evidence used in support of the policy varied from RCTs to prevalence and qualitative data about acceptability of the intervention.
NSF-LTC	
Results from the tool	Assessment of the tool
The body of evidence is Research grade C. The body of evidence is rated as lower quality, but this is in part due to the lack of information provided in the Clinical Commissioning Policy. The evidence is direct with a combination of P1, P2 and S1 evidence.	This performed well. All of the included evidence was direct and it was possible to assess study design for most of the studies. However it was not possible to assess quality for some studies, although a better assessment of the key RCT in the area was made due to having more available information.

Case Study 4: The use of Rituximab as a second line agent for the eradication of inhibitors in patients with Acquired Haemophilia

COMPASS	
Results from the tool	Assessment of the tool
None of the questions in COMPASS could be answered.	The information in the policy comes from case reports which have few features similar to trials, for which this checklist was designed.
OCEBM	
Results from the tool	Assessment of the tool
The evidence included in the Clinical Commissioning Policy is Level 4 evidence, consisting of a systematic review of case reports, case series and a non-randomised trial.	The tool allowed for the evidence to be rated.
GRADE	
Results from the tool	Assessment of the tool
There was insufficient evidence in the CCP to populate the GRADE evidence table and therefore make recommendations.	The information was from a systematic review of non-randomised studies, so many of the questions in GRADE could not be answered.
NSF-LTC	
Results from the tool	Assessment of the tool
The study is of Research grade C as the study is of low quality, but this is in part due to the lack of information provided in the CCP. The evidence is direct, review based evidence but scores low quality due to the paucity of information in the policy.	The tool was straightforward to apply, and allowed for the fact the evidence in the review was non-standard RCT evidence.

Appendix 4 – Medline Search Strategy

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) <1946 to Present>

Search Strategy:

---

1 (value based adj7 assess*).ti,ab.

2 value based pricing.ti,ab.

3 value based decision*.ti,ab.

4 value based care.ti,ab.

5 value based purchas*.ti.

6 1 or 2 or 3 or 4 or 5

****

Appendix 5 – Summary of cost-effectiveness criteria

Study	Method of measuring performance against cost-effectiveness criterion
Marsh et al. [49]	Cost per QALY gained
Goetghebeur et al. [50]	Incremental cost per natural unit
Poulin et al. [51]	No mention but says Economic Analysis (needs a cost–benefit analysis)
Poulin et al. [52]	No explicit mention but refers to HTA reports - 12.1 Is there evidence to support the costeffectiveness of the technology?
Defechereux et al. [53]	Cost/DALY > GDP/capita
Makundi et al. [54]	Total costs per life year saved (USD) and Total cost per DALY saved (USD)
Youngkong et al. [55]	Not described
Miot et al. [56]	Cost per life-year gained and Cost per QALY gained
Baltussen et al. [57]	Cost per DALY compared to GDP/capita
Jehu-Appiah et al. [58]	Cost per DALY compared to GDP/capita
Youngkong et al. [59]	Incremental cost per QALY compared to per-capita GDP
NHS England [60]	Ranking scores of clinical benefit in combination with assessment of positive, negative or neutral costs
Baeten. [61]	Cost per qaly
Baltussen. [62]	Cost per DALY compared to GDP/capita
Diaby. [41]	Creation of 4 categoris (very cost effective, cost effective, low cost effectiveness and not cost effective)
Goetghebeur et al. [63]	Not described
Tony et al. [64]	Committee members were instructed to score individually (on a scale of 0 to 3) each criterion of the MCDA Core Model, using evidence synthesized for each of them (by-criterion HTA report).
Venhorst et al. [65]	Costs per gained healthy life year compared to Gross Domestic Product (GDP) per capita

Appendix 6: Consolidated criteria matrix

Criteria	Marsh [49]	Goetgheheur [50]	Johnson-Masotti [66]	Poulin [51]	Nuti [67]	Wilson [68]	Honoré [69]	Wilson [70]	Poulin [52]	Defechereux [53]	Kreng [71]	Makundi [54]	Youngkong [59]	Miot [56]	Baltussen [57]	Jehu-Appiah [58]	Youngkong [55]	Lee [72]	Baeten [61]	Baltussen [62]	Diaby [41]	Goetgheheur [63]	Tony [64]	Venhorst [65]	Bots [73]	Cho [74]	Golan [75]	Hilgerink [76]	Hummel [77]	Le Gales [78]	Shin [79]	Sloane [80]	Linley [81]	Sussex [40]	ASCO [39]	CCRG [60]	Sum
Program outputs																																				
Effectiveness of program/technology		1		1		1	1	1	1				1	1			1							1				1		1			1	1	1	15
Quality of life impact		1				1																1											1		1	5
Incremental health gain										1		1			1				1	1		1		1												7
Community/social cohesion																																1				1
Contribution to research/Innovation									1																							1	1			3
Engagement								1																												1

Appendix 6: Consolidated criteria matrix (cont…)

Criteria	Marsh [49]	Goetghebeur [50]	Johnson-Masotti	Poulin [51]	Nuti [67]	Wilson [68]	Honoré [69]	Wilson [70]	Poulin [52]	Defechereux [53]	Kreng [71]	Makundi [54]	Youngkong [59]	Miot [56]	Baltussen [57]	Jehu-Appiah [58]	Youngkong [55]	Lee [72]	Baeten [61]	Baltussen [62]	Diaby [41]	Goetghebeur [63]	Tony [64]	Venhorst [65]	Bots [73]	Cho [74]	Golan [75]	Hilgerink [76]	Hummel [77]	Le Gales [78]	Shin [79]	Sloane [80]	Linley [81]	Sussex [40]	ASCO [39]	CCRG [60]	Sum
Organization inputs																																					
Adherence to policy		1				1		1	1													1	1													6
Alignment with current priorities			1																																	1
Feasibility		1	1		1	1		1															1						1							7
Following professional standards									1																												1
Meeting recognized professional standards		1		1										1								1															4
Presence/absence of alternatives													1																				1	1		1	4
Focus on prevention						1	1						1																								3
Sustainability				1					1																												2
Raise profile of low profile condition						1																															1
Comparative intervention limitations		1																				1															2
Meeting health needs of population						1								1																							2
Meeting identified health need			1			1																															2
Meeting public expectation						1								1																							2
Public health interest		1																				1															2

Appendix 6: Consolidated criteria matrix (cont…)

Criteria	Marsh [49]	Goetghebeur [50]	Johnson-Masotti	Poulin [51]	Nuti [67]	Wilson [68]	Honoré [69]	Wilson [70]	Poulin [52]	Defechereux [53]	Kreng [71]	Makundi [54]	Youngkong [59]	Miot [56]	Baltussen [57]	Jchu-Appiah [58]	Youngkong [55]	Lee [72]	Baeten [61]	Baltussen [62]	Diaby [41]	Goetghebeur [63]	Tony [64]	Venhorst [65]	Bots [73]	Cho [74]	Golan [75]	Hilgerink [76]	Hummel [77]	Le Gales [78]	Shin [79]	Sloane [80]	Linley [81]	Sussex [40]	ASCO [39]	CCRG [60]	Sum
Stakeholder involvement																																					
Political/historical context		1																					1													2
User feedback possible						1																															1
Stakeholder pressures		1																					1									1					3
User involvement in decision-making						1																															1
Willingness to subsidize										1																											1
Community involvement						1																															1
User-view experience/satisfaction						1																															1
Program delivery																																					
Type of service		1												1								1															3
Sufficient knowledge/staff base				1	1				1																												3
Service impact on other agencies						1			1																											1	3
Impact on future decisions														1																							1
Impact on future practice														1																							1
Partnerships/Integration with other programs				1		1								1																							3
Maintenance of quality						1																															1
Safety		1	1	1					1															1				1	1			1		1	1	1	11

Appendix 6: Consolidated criteria matrix (cont…)

Criteria	Marsh [49]	Goetghebeur [50]	Johnson-Masotti	Poulin [51]	Nuti [67]	Wilson [68]	Honoré [69]	Wilson [70]	Poulin [52]	Defechereux [53]	Kreng [71]	Makundi [54]	Youngkong [59]	Miot [56]	Baltussen [57]	Jehu-Appiah [58]	Youngkong [55]	Lee [72]	Baeten [61]	Baltussen [62]	Diaby [41]	Goetghebeur [63]	Tony [64]	Venhorst [65]	Bots [73]	Cho [74]	Golan [75]	Hilgerink [76]	Hummel [77]	Le Gales [78]	Shin [79]	Sloane [80]	Linley [81]	Sussex [40]	ASCO [39]	CCRG [60]	Sum
Ethics																																					
Acceptability						1		1					1											1						1							5
Accessibility			1	1		1			1															1													5
Poverty reduction															1	1			1	1																	4
Variation in practice																	1																				1
Age/risk of target group										1					1	1			1	1	1				1				1			1					9
Equity/Reducing inequalities	1	1						1				1					1				1		1	1			1					1					10
Gender of target group													1																								1

Appendix 6: Consolidated criteria matrix (cont…)

Criteria	Marsh [49]	Goetghebeur [50]	Johnson-Masotti	Poulin [51]	Nuti [67]	Wilson [68]	Honoré [69]	Wilson [70]	Poulin [52]	Defechereux [53]	Kreng [71]	Makundi [54]	Youngkong [59]	Miot [56]	Baltussen [57]	Jehu-Appiah [58]	Youngkong [55]	Lee [72]	Baeten [61]	Baltussen [62]	Diaby [41]	Goetghebeur [63]	Tony [64]	Venhorst [65]	Bots [73]	Cho [74]	Golan [75]	Hilgerink [76]	Hummel [77]	Le Gales [78]	Shin [79]	Sloane [80]	Linley [81]	Sussex [40]	ASCO [39]	CCRG [60]	Sum
Health inputs																																					
Current burden (morb/mort)		1				1	1	1	1			1		1	1										1							1				10
Risk/benefit of treatment						1								1																						2
Disease characteristics/severity		1				1				1		1			1	1	1		1	1	1	1										1	1		1	14
Magnitude of benefit (number of people)				1						1		1	1	1		1	1		1			1			1											10
Proportion elligible	1																																			1
Quality of Evidence	1	1	1			1		1				1	1	1								1		1			1								1	12
Completeness and consistency of reporting evidence														1								1														2
Health Economics (CE)	1	1		1					1	1		1	1	1	1	1	1		1	1	1	1		1											1	17
Economics																																					
Appropriateness (make best use of resources)		1				1							1																							3
Efficiency			1																				1													2
Budgetary impact/Affordability	1	1	1	1	1		1		1					1			1			1		1		1			1	1	1	1		1				17
Impact on other spending														1								1														2
Cost to user																	1																	1	1	3
Miscellaneous		1		1													1								1	1	1		1	1	1					9

Appendix 7: Rationale for ratings by the two reviewers for RAS, R2AH andHIV TasP

Table 1: Diaby classification

	RAS		R2AH		HIV TasP	
	Reviewer 1	**Reviewer 2**	**Reviewer 1**	**Reviewer 2**	**Reviewer 1**	**Reviewer 2**
Highly significant benefit						RCT and cohort data indicate an overall risk reduction of transmission in discordant couples as 96%. However, rates not reported, so numbers needed to treat could be very high. ART is well tolerated.
Significant medical benefit			"The Franchini review included 27 studies (19 case reports, 7 case series and 1 open label trial). Complete remission was reported in 57 of the 65 participants, partial response in 2, minor response in 1 and no response in 5 participants. This suggests that rituximab appears to be an effective option for patients with AHA for who established therapies have failed" No alternatives to this medication are reported but "Patients with AHA are at high risk of severe and fatal haemorrhage until the inhibitor has been eradicated"		Public Health England estimate that 1,800 new HIV infections will be prevented as a result of this policy, impacting significantly on reducing the HIV epidemic. Quality of life has not been specifically studied when treatment is used as prevention.	

Relatively significant medical benefit	There was no compelling evidence that robot-assisted approaches impact on long term oncological outcomes when compared with laparoscopic and standard approaches. There is some evidence of clinical advantages from robot-assisted prostatectomy when compared with both laparoscopic and open radical procedures. These include lower risk of incontinence or sexual dysfunction, and reduced blood loss and lengths of stay, when compared to open prostatectomy.	2 RCTs vs LRP. Higher rate of urinary continence, statistically significant in 1 trial. Recovery of sexual function higher in both trials. No significant difference in margins and relapse-free survival at one year. Non-randomised studies have mixed results across all outcomes. Other evidence suggestive of reduced hospital stay and associated quality of life effects.		Case series show high rates of remission (57/65). The impact of this on events is not estimates, however, the potential consequences of failure are severe.		
Non-significant medical benefit						

Table 2: CCRG classification

	RAS		R2AH		HIV TasP	
	Reviewer 1	**Reviewer 2**	**Reviewer 1**	**Reviewer 2**	**Reviewer 1**	**Reviewer 2**
Superior efficacy and cost saving compared to currently used alternative			This is cost-saving: "The high cost of haemostatic agents to control bleeding in patients with a persistent inhibitor almost invariably exceeds the investment required for inhibitor eradication using rituximab." This medication is more efficacious than no treatment.		Lifetime costs per-case are estimated at between £280-360k, therefore resulting in an overall saving of between £500-647 million to the NHS (page 7, CC policy document) 1,800 new HIV infections will be prevented as a result of this policy (Page 9, CC policy document)	Superior efficacy from at least one RCT. Savings modelled. Although details are not available, the conclusion has face validity. An alternative analysis produces "an incremental cost effectiveness ratio of <(£4,000)" after taking into account price reductions related to generics.
Cost saving and non-inferior to current equivalent						
Cost neutral and non-inferior to current equivalent but has other advantages less toxic, oral administration						

If no QALY available and costs more than current alternative	A health economic analysis concluded that robot-assisted prostatectomy was more expensive than laparoscopic prostatectomy The policy represents a cost pressure to the NHS	No QALYs available. Cost is approx. £6500 per patient. It is unclear whether this is an incremental cost. Savings from reduced hospitalisation and blood use unquantified.		It is stated as cost neutral within the available documentation, however, the necessary data are not shown. Cost is also linked to budgets… 'it will cost less as we are already paying for some of it'. This is irrelevant to issues of cost.		

REFERENCES

1. Köksalan, M., J. Wallenius, and S. Zionts, *An Early History of Multiple Criteria Decision Making.* Journal of Multi-Criteria Decision Analysis, 2013. **20**(1-2): p. 87-94.
2. Devlin, N. and J. Sussex, *Incorporating Multiple Criteria in HTA: Methods and processes*. 2011: Office of Health Economics, Whitehall, London, UK.
3. Diaby, V., K. Campbell, and R. Goeree, *Multi-criteria decision analysis (MCDA) in health care: A bibliometric analysis.* Operations Research for Health Care, 2013. **2**(1–2): p. 20-24.
4. Thokala, P. and A. Duenas, *Multiple Criteria Decision Analysis for Health Technology Assessment.* Value in Health, 2012. **15**(8): p. 1172-1181.
5. Roy, B., *The outranking approach and the foundations of electre methods.* Theory and Decision, 1991. **31**(1): p. 49-73.
6. Roy, B. and D. Vanderpooten, *The European school of MCDA: Emergence, basic features and current works.* Journal of Multi-Criteria Decision Analysis, 1996. **5**(1): p. 22-38.
7. Brans, J.-P. and P. Vincke, *A Preference Ranking Organisation Method: (The PROMETHEE Method for Multiple Criteria Decision-Making).* Management Science, 1985. **31**(6): p. 647-56.
8. Brans, J.-P. and B. Mareschal, *The PROMCALC & GAIA decision support system for multicriteria decision aid.* Decision Support Systems, 1994. **12**(4): p. 297-310.
9. Aouni, B.d. and O. Kettani, *Goal programming model: A glorious history and a promising future.* European Journal of Operational Research, 2001. **133**(2): p. 225-231.
10. Belton, V. and T. Stewart, *Multiple Criteria Decision Analysis: An Integrated Approach* 2002: Kluwer Academic Publishers.
11. Marsh, K., et al., *Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature.* Pharmacoeconomics, 2014. **32**(4): p. 345-65.
12. IQWiG, *Assessment and analysis of studies on rare diseases*. 2014, Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).
13. CASP, *Critical Appraisals Skills Programme (CASP)*. 2013.
14. Husereau, D., et al., *Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement.* BMJ, 2013. **346**.
15. Cochrane, E., *Cochrane Effective Practice and Organisation of Care (EPOC)*. 2015.
16. Higgins, J.P.T. and S. Green, *Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]*. 2011, Cochrane Collaboration.
17. Picavet, E., et al., *Development and validation of COMPASS: clinical evidence of orphan medicinal products - an assessment tool.* Orphanet Journal Of Rare Diseases, 2013. **8**.
18. Centre for, R. and Dissemination, *Systematic reviews : CRD's guidance for undertaking reviews on healthcare*. 2009, University of York, Centre for Reviews and Dissemination: York.
19. Drummond, M.F. and T.O. Jefferson, *Guidelines for authors and peer reviewers of economic submissions to the BMJ.* BMJ, 1996. **313**(7052): p. 275-283.
20. Hillier, S., et al., *FORM: an Australian method for formulating and grading recommendations in evidence-based clinical guidelines.* BMC Medical Research Methodology, 2011. **11**: p. 23.
21. GRADE Working Group, *GRADE*. 2014, GRADE Working Group.
22. Gugiu, P.C., *Hierarchy of evidence and appraisal of limitations (HEAL) grading system.* Evaluation & Program Planning, 2015. **48**: p. 149-159.
23. NHSEngland, *Standard Operating Procedures: The Cancer Drugs Fund (CDF) Guidance to support operation of the CDF in 2015-16*. 2015, London: NHS England.
24. DH Long-term Conditions NSF Team, *National Service Framework for long-term conditions*. 2005, Department of Health.
25. Oxford Levels of Evidence Working Group, *The Oxford Levels of Evidence 2*. 2011, Oxford Centre for Evidence-Based Medicine.

26. Onakpoya, I.J., et al., *Effectiveness, safety and costs of orphan drugs: an evidence-based review.[Erratum appears in BMJ Open. 2015;5(10):e007199corr1; PMID: 26503381].* BMJ Open, 2015. **5**(6).
27. Brosseau, L., et al., *Ottawa Panel Evidence-Based Clinical Practice Guidelines for the Management of Osteoarthritis in Adults Who Are Obese or Overweight.* Physical Therapy, 2011. **91**(6): p. 843-861.
28. Ofman, J.J., et al., *Examining the value and quality of health economic analyses: implications of utilizing the QHES.* J Manag Care Pharm, 2003. **9**(1): p. 53-61.
29. Whiting, P., *QUADAS*. 2015, University of Bristol.
30. SIGN, *Scottish Intercollegiate Guidelines Network*. 2015, SIGN.
31. SMC, *Scottish Medicines Consortium*. 2015: SMC.
32. Ebell, M.H., et al., *Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature.* Am Fam Physician, 2004. **69**(3): p. 548-556.
33. USPST, *United States Preventive Services Taskforce*. 2015, USPST.
34. Baker, A., et al., *The applicability of grading systems for guidelines.* Journal of Evaluation in Clinical Practice, 2011. **17**(4): p. 758-762.
35. Health Improvement, S., *Open, laparoscopic and robot-assisted laparoscopic radical prostatectomy for localised prostate cancer (Evidence Note 49)*. 2013, Health Improvement Scotland.
36. Cromwell, I., S. Peacock, and C. Mitton, *'Real-world' health care priority setting using explicit decision criteria: a systematic review of the literature.* BMC Health Services Research, 2015. **15**(1): p. 164.
37. Wahlster, P., et al., *Balancing costs and benefits at different stages of medical innovation: a systematic review of Multi-criteria decision analysis (MCDA).* BMC Health Services Research, 2015. **15**(1): p. 262.
38. Guindo, L., et al., *From efficacy to equity: Literature review of decision criteria for resource allocation and healthcare decisionmaking.* Cost Effective Resour Alloc, 2012. **10**(1): p. 9.
39. Schnipper, L.E., et al., *American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options.* J Clin Oncol, 2015. **33**(23): p. 2563-77.
40. Sussex, J., et al., *A Pilot Study of Multicriteria Decision Analysis for Valuing Orphan Medicines.* Value in Health, 2013. **16**(8): p. 1163-1169.
41. Diaby, V. and J. Lachaine, *An application of a proposed framework for formulary listing in low-income countries: the case of Cote d'Ivoire.* Applied Health Econ Health Policy, 2011. **9**(6): p. 389 - 402.
42. Pattanaphesaj, J. and Y. Teerawattananon, *Reviewing the evidence on effectiveness and cost-effectiveness of HIV prevention strategies in Thailand.* BMC Public Health, 2010. **10**(1): p. 401.
43. Baltussen, R., et al., *Multi-criteria decision analysis to prioritize health interventions: Capitalizing on first experiences.* Health Policy (Amsterdam, Netherlands), 2010. **96**(3): p. 262 - 4.
44. Drummond, M., et al., *Methods for the economic evaluation of health care programmes*. 3rd ed. 2005, Oxford: OUP.
45. Dowie, J., *Why cost-effectiveness should trump (clinical) effectiveness: the ethical economics of the South West quadrant.* Health Econ, 2004. **13**(5): p. 453-9.
46. Kaplan, R.M., *The minimally clinically important difference in generic utility-based measures.* Copd, 2005. **2**(1): p. 91-7.
47. Walters, S.J. and J.E. Brazier, *What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D.* Health Qual Life Outcomes, 2003. **1**: p. 4.

48. Claxton, K., *Three questions to ask when examining MCDA.* Value & Outcomes Spotlight, 2015. **1**(1).
49. Marsh, K., et al., *Prioritizing investments in public health: a multi-criteria decision analysis.* J Public Health (Oxf), 2013. **35**(3): p. 460-6.
50. Goetghebeur, M., et al., *Combining multicriteria decision analysis, ethics and health technology assessment: applying the EVIDEM decision-making framework to growth hormone for Turner syndrome patients.* Cost Effective Resour Alloc, 2010. **8**: p. 4.
51. Poulin, P., et al., *New technologies and surgical innovation: five years of a local health technology assessment program in a surgical department.* Surg Innov, 2012. **19**(2): p. 187-99.
52. Poulin, P., et al., *Multi-criteria development and incorporation into decision tools for health technology adoption.* J Health Organ Manag, 2013. **27**(2): p. 246 - 65.
53. Defechereux, T., et al., *Health care priority setting in Norway a multicriteria decision analysis.* BMC Health Serv Res, 2012. **12**: p. 39.
54. Makundi, E., L. Kapiriri, and O. Norheim, *Combining evidence and values in priority setting: testing the balance sheet method in a low-income country.* BMC Health Serv Res, 2007. **7**: p. 152.
55. Youngkong, S., et al., *Multi-criteria decision analysis for setting priorities on HIV/AIDS interventions in Thailand.* Health Res Policy Syst, 2012. **10**(1): p. 1 - 8.
56. Miot, J., et al., *Field testing of a multicriteria decision analysis (MCDA) framework for coverage of a screening test for cervical cancer in South Africa.* Cost Effective Resour Alloc, 2012. **10**(1): p. 2.
57. Baltussen, R., et al., *Priority setting using multiple criteria: should a lung health programme be implemented in Nepal?* Health Policy Plan, 2007. **22**(3): p. 178 - 85.
58. Jehu-Appiah, C., et al., *Balancing equity and efficiency in health priorities in Ghana: the use of multicriteria decision analysis.* Value Health, 2008. **11**(7): p. 1081 - 7.
59. Youngkong, S., et al., *Multicriteria decision analysis for including health interventions in the universal health coverage benefit package in Thailand.* Value Health, 2012. **15**(6): p. 961 - 70.
60. NHSEngland, *Standard Operating Procedures: The Cancer DRugs Fund (CDF). Guidance to support operatoin of the CDF in 2013-14.* NHS England NHSCB/SOP04. 2014, London: NHS England.
61. Baeten, S., et al., *Incorporating equity-efficiency interactions in cost-effectiveness analysis-three approaches applied to breast cancer control.* Value Health, 2010. **13**(5): p. 573 - 9.
62. Baltussen, R., et al., *Towards a multi-criteria approach for priority setting: an application to Ghana.* Health Econ, 2006. **15**(7): p. 689 - 96.
63. Goetghebeur, M., et al., *Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): Applying the evidem framework to medicines appraisal.* Med Decis Mak, 2012. **32**(2): p. 376 - 88.
64. Tony, M., et al., *Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada.* BMC Health Serv Res, 2011. **11**: p. 329.
65. Venhorst, K., et al., *Multi-criteria decision analysis of breast cancer control in low- and middle- income countries: development of a rating tool for policy makers.* Cost Eff Resour Alloc, 2014. **12**(1): p. 13.
66. Johnson-Masotti, A. and K. Eva, *A decision-making framework for the prioritization of health technologies.* Health services restructuring in Canada: new evidence and new directions, 2006.
67. Nuti, S., M. Vainieri, and A. Bonini, *Disinvestment for re-allocation: a process to identify priorities in healthcare.* Health Policy, 2010. **95**(2-3): p. 137 - 43.
68. Wilson, E., J. Rees, and R. Fordham, *Developing a prioritisation framework in an English Primary Care Trust.* Cost Eff Resour Alloc, 2006. **4**(1): p. 3.

69. Honore, P., et al., *Decision science: a scientific approach to enhance public health budgeting.* J Public Health Manag Pract, 2010. **16**(2): p. 98 - 103.
70. Wilson, E., et al., *Prioritizing health technologies in a Primary Care Trust.* J Health Serv Res Policy, 2007. **12**(2): p. 80 - 5.
71. Kreng, V. and C. Yang, *The equality of resource allocation in health care under the National Health Insurance System in Taiwan.* Health Policy, 2011. **100**(2-3): p. 203 - 10.
72. Lee, C. and N. Kwak, *Strategic enterprise resource planning in a health-care system using a multicriteria decision-making model.* J Med Syst, 2011. **35**(2): p. 265 - 75.
73. Bots, P. and J. Hulshof, *Designing multi-criteria decision analysis processes for priority setting in health policy.* J Multi-Criteria Decis Anal, 2000. **9**(1-3): p. 56 - 75.
74. Cho, K. and S. Kim, *Selecting medical devices and materials for development in Korea: the analytic hierarchy process approach.* Int J Health Plann Manag, 2003. **18**(2): p. 161 - 74.
75. Golan, O. and P. Hansen, *Which health technologies should be funded? A prioritization framework based explicitly on value for money.* Isr J Health Policy Res, 2012. **1**(1): p. 44.
76. Hilgerink, M., et al., *Assessment of the added value of the Twente Photoacoustic Mammoscope in breast cancer diagnosis.* Med Dev (Auckland, NZ), 2011. **4**: p. 107 - 15.
77. Hummel, J., et al., *Predicting the health economic performance of new non-fusion surgery in adolescent idiopathic scoliosis.* J Orthop Res, 2012. **30**(9): p. 1453 - 8.
78. Le Gales, C. and J. Moatti, *Searching for consensus through multi-criteria decision analysis. Assessment of screening strategies for hemoglobinopathies in southeastern France.* Int J Technol Assess Health Care, 1990. **6**(3): p. 430 - 49.
79. Shin, T., et al., *The comparative evaluation of expanded national immunization policies in Korea using an analytic hierarchy process.* Vaccine, 2009. **27**(5): p. 792 - 802.
80. Sloane, E., et al., *Using the analytic hierarchy process as a clinical engineering tool to facilitate an iterative, multidisciplinary, microeconomic health technology assessment.* Comp Oper Res, 2003. **30**(10): p. 1447 - 65.
81. Linley, W.G. and D.A. Hughes, *Societal views on NICE, cancer drugs fund and value-based pricing criteria for prioritising medicines: a cross-sectional survey of 4118 adults in Great Britain.* Health Econ, 2013. **22**(8): p. 948-64.