Making best use of evidence for explicit decisions in health care

Volume 1 of 2

Marta Soares

PhD

University of York

Health Sciences

April 2017

Abstract

This thesis explores how decision makers can make best use of evidence to inform unavoidable decisions about the appropriate use of resources in health care. A central aspect of the work is that in addition to decisions directly about (the funding of) health care technologies it also considers that further research may be needed to substantiate these decisions. Joint decisions about funding technologies and further research require an appropriate evaluative basis. This thesis reports such a framework, which also identifies the critical assessments required and, importantly, demonstrates their feasibility and applicability to analysts and policymakers. The relevance of this work has been recognised by key policymakers and has informed the most recent update of the methodological guidance of the National Institute for Health and Care Excellence (NICE) in the UK. However, the development of a framework is necessary but not sufficient for ensuring that best use is made of evidence to inform these joint decisions. These decisions also require that the evidence base is appropriately described with uncertainties characterised explicitly. This thesis compiles a number of examples where standard methods would have fallen short and hence more novel approaches were required. The approaches used generally aim to: make best use of available *data* (even if related but not directly relevant to the decision problem) or generate *evidence* by appropriately eliciting the judgements of relevant experts. The examples provide important methodological and applied contributions that demonstrate both the feasibility and value of ensuring that unavoidable decisions are informed by the best use of evidence.

List of Contents

Acknowledgements

I would like to thank my advisors (Catherine Hewitt, Stephen Palmer and Karl Claxton), all the contributors to the research included in the collection, and my family and friends.

Author's declaration

The six papers [Papers 1–6] that form the core of this thesis are listed below. Whilst I am not the first author on all of the papers below I have significantly contributed to a substantial part of the work. My contribution to each paper is described after the citation of the respective paper. The integrative chapter that binds these papers into the thesis is solely my own work. I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Paper 1. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, McKenna C, Soares M, Spackman E, Youn J. Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development. Health Technology Assessment 2012;16(46):1-323. doi: 10.3310/hta16460.

[Contribution of the candidate: I have contributed to the development of the framework as part of the wider research team. I have applied the methods of assessment to the clopidogrel case study and was responsible for preparing and revising Appendix 8 of the report. A summary of this component of the work is included in Chapter 5 of the report.]

Paper 2. Soares MO, Welton NJ, Harrison DA, Peura P, Shankar-Hari M, Harvey SE, Madan J, Ades AE, Rowan KM, Palmer SJ. Intravenous immunoglobulin for severe sepsis and septic shock: clinical effectiveness, cost-effectiveness and value of a further randomised controlled trial. Critical Care. 2014 Dec 1;18(6):649.

[Contribution of the candidate: I designed, developed and conducted the cost-effectiveness and value of information analyses. I prepared the manuscript and dealt with subsequent revisions.]

Paper 3. Soares MO, Dumville JC, Welton N, Ades T. Treatment comparisons for decision making: facing the problems of sparse and few data. Journal of the Royal Statistical Society: Series A (Statistics in Society), 177(1), 2014, 259-279

[Contribution of the candidate: I designed, developed and conducted the analyses. I prepared the manuscript and dealt with subsequent revisions]

Paper 4.   Corbett M, Soares M, Jhuti G, Rice S, Spackman E, Sideris E, Moe-Byrne T, Fox D, Marzo-Ortega H, Kay L, Woolacott N, Palmer S. Tumour necrosis factor-α inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis: a systematic review and economic evaluation. Health Technology Assessment 2016 Feb;20(9):1-334.

[Contribution of the candidate: I designed, developed and conducted the extended synthesis. I prepared and revised Chapter 5 of the report]

Paper 5.   Soares MO, Bojke L, Dumville J, Iglesias C, Cullum N, Claxton K. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. Statistics in Medicine 2011; 30(19): 2363-*2380.*

[Contribution of the candidate: I designed, developed and conducted the elicitation exercise, and analysed the results. I prepared the manuscript and subsequent revisions]

Paper 6.   Soares MO; Dumville J; Ashby R; Iglesias C; Bojke L; Adderley U; McGinnis E; Stubbs N; Torgerson D; Claxton K; Cullum N. Methods to assess cost effectiveness and value of further research when data are sparse: negative pressure wound therapy for severe pressure ulcers. Medical Decision Making. 33(3), 2013, 415-36.

[Contribution of the candidate: I designed, developed and conducted the analyses. I prepared the manuscript and subsequent revisions]

Integrative chapter

This thesis explores how decision-makers can make best use of evidence to inform unavoidable decisions about the appropriate use of resources in health care. A central aspect of the work is that in addition to decisions directly about (the funding of) health care technologies it also considers that further research may be needed to substantiate these decisions. Joint decisions about funding technologies and further research require an appropriate evaluative basis. This thesis reports such a framework which also identifies the critical assessments required and, importantly, demonstrates to analysts and policymakers its feasibility and applicability. The relevance of this work has been recognised by key policymakers and has informed the most recent update to the methodological guidance of the National Institute for Health and Care Excellence (NICE) in the UK. However, the development of a framework is necessary, but not sufficient, for ensuring that best use of evidence is made to inform these joint decisions. These decisions also require that the evidence base is appropriately described, with all uncertainties characterised explicitly. This thesis compiles a number of examples where standard methods would have fallen short and hence more novel approaches were required. The approaches used generally aim to: make best use of available *data* (even if relevant but not directly related to the decision problem) or generate *evidence* by appropriately eliciting the judgements of relevant experts. The examples provide important methodological and applied contributions that demonstrate both the feasibility and value of ensuring that unavoidable decisions are informed by the best use of evidence.

Whilst the integrative chapter aims to lay down these principles in a simple and accessible way, some of the methodological specificities are given in footnotes. Note, however, that the papers and monographs in the collection give a detailed account of the complexity of the methodological developments underlying this thesis.

Context

Decisions about health technologies lead to (funding) recommendations based on the intervention which is expected to provide the most value to the population of potential users. Assessing value requires that the direct health benefits of competing interventions are considered. In many jurisdictions (such as in the UK), value also incorporates the implications to health of the investment interventions require. Health economists assess the health that could be achieved if the same level of resource being used to fund an intervention were used to benefit other patients: this is the *health opportunity cost* of the investment required[1,2]. The technology that is expected to offer the most health, discounted of its health opportunity cost[3], constitutes the best use of resources (is cost-effective) and should be recommended for wider use (i.e. 'accepted').[4] The less valuable alternatives should thus be 'rejected'. Such principles have been used in policy for many years, namely by NICE in the UK. NICE issued its first guidance document in the year 2000[5] and since then has issued 663 recommendations (419 guidance documents) within the technology appraisal (TA) process. NICE has also expanded its remit to cover wider decision-making problems, for example around diagnostic technologies and public health interventions. Internationally, cost-effectiveness is also commonly used (e.g. CADTH in Canada[6], PBAC in Australia[7] and INFARMED in Portugal[8]).

The role of evidence in decision-making

To establish cost-effectiveness, a single piece of evidence is unlikely to be sufficiently informative.[9] For example, a single randomised controlled trial (RCT) often will not capture the costs and health benefits of all competing interventions over a sufficient time frame. Instead, evidence on a range of relevant factors (or parameters) is often gathered using a formal process called Health Technology Assessment (HTA).[10] Examples of the type of information collected within HTA include[11]:

- clinical effectiveness,
- baseline risks (natural history) of disease/condition,
- potential harm from the technologies,

- long-term disease progression across a set of relevant health states, and

- costs and quality of life associated with those health states.

Where multiple studies inform the same factor, or parameter, there is the need to *synthesise* (or pool)[12-16] this subset of the evidence base. *Models* are then used to combine all synthesised data for relevant parameters; here models define the relationships between different parameters to establish the cost and health consequences of alternative competing interventions – a decision analytic modelling framework[17-19]. Specifically, the often complex models that are developed allow the relevant but varied evidence (on the parameters listed above) to be drawn together to establish the speed of progression through disease with the alternative treatment options, and determine time spent in relevant health states. This occurs in tandem with the application of health-related quality-of-life weights and costs to time spent in the different health states, allowing cost-effectiveness to be ascertained.

Evidence, at the heart of cost-effectiveness analyses, is, however, often uncertain. Uncertainty stems from numerous sources, one being that studies in the evidence base recruit only a sample of patients (and not the whole of the population)—this source of uncertainty is here referred to as *inferential* uncertainty.[20] Inferential uncertainties are conventionally described using distributions and propagated in the model through to cost-effectiveness using Monte Carlo simulation methods, in a process called probabilistic sensitivity analysis (PSA)[21], reflecting an inherently Bayesian view of uncertainty[22]. There are, however, other causes of uncertainty such as *structural* uncertainties underlying methodological choices and judgements made when gathering and modelling the evidence, and *temporal* uncertainty, when the timeframe to which the evidence relates is limited. Recent research has outlined how these other sources of uncertainty can be appropriately considered in cost-effectiveness analyses.[23,24]

Uncertainty and the value of further research

Where complex models are defined, unbiased assessments of the expected costs and effects require that uncertainty is considered.[25] But inferential,

structural and temporal uncertainties also mean that the value of the technologies is itself uncertain. Hence, there may be uncertainty over whether the particular technology identified for approval is, in fact, best. And, if the value of an approved technology is not realised in practice, resources are wasted.[26,27] The magnitude of potential losses (or the cost of uncertainty) can be quantified using well-known extensions of the analyses of uncertainty typically required for cost-effectiveness, called expected value of information (EVI) analyses.[26,28-31] Further research can be considered to offer value if its costs are less than the costs of the uncertainty it aims to reduce. In terms of policy, alongside adoption or rejection, decision-makers thus ought to consider whether appropriately designed research is worthwhile conducting, with the purpose of reviewing the original decision once its results become available.[32,33]

Considering research alongside technology adoption is becoming even more relevant in the current policy context, where the provision of 'earlier' access to new (and promising) medicines is being incentivised (e.g. adaptive pathways process by the European Medicines Agency[34] and accelerated access review in the UK[35]). Earlier access means a less developed evidence base at the point of launch, which determines the need for post-marketing clinical research.[36] Moreover, as in the UK, many jurisdictions widened the scope of decision processes to include technologies such as diagnostics, medical devices and public health or social care interventions (co-dependent technology process is Australia[37], and the Diagnostic Assessment Review process in the UK[38]). These interventions are not required to demonstrate effectiveness for regulatory purposes, and are thus likely to have a less developed evidence base than medicines.[39] Finally, recent policy actions are increasingly considering the use of real world data to support decision-making – for example, in 2016 the new Cancer Drugs Fund in the UK[40] established that cancer drugs for which significant clinical uncertainties remain can still be funded, but only under an agreed concomitant data collection programme which may involve observational data collected through the Systemic Anti-Cancer Therapy Dataset (SACT).[41]

Joint research and adoption decisions

In the UK, NICE's TA process has offered, since inception, a third policy option alongside 'accept' and 'reject': recommending the use of a particular technology only within a specific programme of research ('Only in research', OIR). The first of such recommendations was issued in the year 2000, and up to 2004 24% of all recommendations issued were OIR (13/54). Since, the use of this policy option has diminished: between 2005 and 2009 only 4.6% were OIR (8/171), between 2010 and 2014 there were 1.7% (4/234) and, finally, in 2015 and up to November 2016 no OIR guidance was issued despite 115 recommendations being made.[42]

Given that drugs are often priced close to their value-based price[43] where uncertainty is most significant, OIR recommendations are likely being underused. A potential reason for this is that decision-making committees are not being presented with adequate information on the need for further research. Besides cost-effectiveness, NICE requires that probabilistic (PSA) and deterministic sensitivity analyses (DSA) are conducted, but not EVI analyses.[44] PSA establishes the level of support of the evidence (on parameterised, inferential uncertainties) for each of the technologies (i.e. probability of each being cost-effective).[17,45] DSA establishes the influence of uncertainties that have not been parameterised (and have thus been excluded from the PSA) on expected cost-effectiveness. Whilst PSA and DSA can usefully inform 'adoption' and 'rejection' decisions[46] they are insufficient to determine whether further research is worthwhile for which assessments of the consequences of uncertainty (EVI analysis) are required.

Perhaps realising that the assessments presented to decision-makers are, by themselves, insufficient to inform joint research and adoption decisions[33], in 2010 NICE liaised with the Medical Research Council to commission methodological research on this topic. The research was published in 2012 and is included in the collection of papers for this thesis (Paper 1). It established the principles of joint research and adoption decisions which have since been incorporated in NICE's methods guide. But the research also determined the specific assessments needed, and how they should be conducted, and unequivocally demonstrated their applicability in practice. However, the NICEs

methods guide has not incorporated this specific infromation[44], and thus the set of analyses presented to committees were (and still are) inadequate to inform joint decisions about adoption and research. The results of the research are summarised next.

A framework for informing policy decisions under uncertainty

Paper 1 in the collection establishes the principles and assessments required to inform recommendations on the use of health care technologies considering the possibility of also recommending appropriately designed programmes of accompanying research. The framework proposed allows for two policy options, in addition to 'accept' and 'reject': the abovementioned OIR option, that restricts the use of new technology to those patients who are involved in the research, and an 'Approval with Research' (AWR) option that approves the technology for widespread use on the condition that additional evidence to support its use is collected.[47] The framework is based on an explicit quantification of the value of the alternative policy options, coherent with the principles of decision-making in health care, particularly those set out by NICE. Key elements include assessments of the value of treatments—using cost-effectiveness analyses that already form the basis of the evidence presented to NICE appraisals—and of the value of future research using EVI analyses, which is currently not specified by NICE but for which there is a vast body of methods and applied literature.[28-31]

To establish the value of the alternative policy options additional considerations are, however, required. Firstly, it is important to recognise that research takes time to complete. Before the required further research data is available, individuals subjected to the initial, uncertain, treatment decision face the consequences of uncertainty. It is only after new research data become available, and a review of the initial decision is made, that future cohorts are able to benefit from the research. Hence, to establish the value of OIR/AWR policies, there needs to be some judgement made on the expected time for research to report, and some consideration of the possibility it does not report [48]. Additionally, the framework recognises that the value of the

treatments and/or of the research may be affected by other changes occurring over time, for example: a comparator becoming off-patent and its price reduced; another piece of related research reporting or the emergence of new technologies that might make existing ones obsolete. Also, if the further research needed is comparative, it is important to consider those patients that will lose out by receiving a technology expected to be less valuable. Finally, the costs of conducting the research should also be considered.

Based on the above, assessments of the value of the four alternative policy options— accept, reject, OIR and AWR—can be established. For a technology expected to be cost-effective, an AWR decision could be anticipated to be more valuable than an OIR decision. But the work identified two circumstances in which OIR could become more valuable. The first is when research is not possible with approval[49], which may depend, in part, on whether the type of evidence required has an experimental design; for example, more precise estimates of relative treatment effect are likely to require an RCT to avoid selection bias, but this is unlikely to be possible once a technology is approved for widespread use. In this case, it may be worth delaying widespread approval under an OIR policy, even if the technology is expected to be cost-effective. The second circumstance in which OIR, or even rejection, should be considered for a cost-effective technology is where implementing approval commits resources that cannot be recovered if guidance changes at a later date.[33,50,51] These 'irrecoverable costs' could be: capital expenditure on equipment or facilities; additional resources required to implement guidance or to train staff to use a new health technology; or a period of 'learning'. Or irrecoverable costs may be related to the fact that most new technologies impose initial per-patient treatment costs that exceed the immediate health benefits and are offset only in the longer run, and thus may be seen as irrecoverable if guidance changes before the initial losses are compensated by later gains.[52]

To facilitate the use of the framework, Paper 1 develops a list of assessments that guide analyses and target decision-makers' considerations. These have been structured into a checklist (reproduced in Box 1). Answers to each item ('Yes' or 'No') create a profile that, via an algorithm, identifies the appropriate

policy option for a particular decision problem (i.e. OIR, AWR, Approve, or Reject).

Box 1: Assessment Checklist for coverage decisions conditional on evidence collection (OIR, AWR, Approve, or Reject) – Paper 1 in the collection

1 Is it cost-effective?
2 Are there significant irrecoverable costs?
3 Does more research seem worthwhile?
4 Is the research possible with approval?
5 Will other sources of uncertainty resolve over time?
6 Are the benefits of research greater than the costs?
7 Are the benefits of approval greater than the costs?

The applicability of these critical assessments is demonstrated in Paper 1 using a series of case studies, one of which is presented below in more detail.

Demonstration of the framework: clopidogrel case study

This case study concerns the management of patients with non-ST-segment elevation acute coronary syndromes presenting with a moderate to high risk of ischemic events. The use of clopidogrel in this indication was appraised by NICE (TA80 in 2004[5]), resulting in positive guidance on its use as an adjunct to standard therapy (aspirin alone). Further research was also recommended to inform the optimal treatment duration.

For the purpose of this case study, we take the position of the reimbursement agency at the time of this initial appraisal, but consider different durations of clopidogrel use, ranging from 1 month to 12 months. To establish cost-effectiveness, a decision analytic model considered effects on mortality and on the occurrence of non-fatal myocardial infarctions (MIs). Results showed that twelve months' treatment with clopidogrel was the regimen expected to be cost-effective (the incremental cost-effectiveness ratio was below £20,000 per Quality Adjusted Life Year, QALY). Considering the potential population of users (60,000 patients per year[53]), it was estimated that the National Health Service (NHS) in the UK could gain at least 14,786 QALYs (or approximately

£300 million)[i] by changing its guidance to consider 12 months clopidogrel instead of NHS standard care over the technology time horizon[54] (assumed to be 10 years). Clopidogrel was thus deemed cost-effective, which completes the first assessment of the framework ('Point 1: Is it cost-effective?', Box 2).

Box 2: Assessment Checklist for coverage decisions conditional on evidence collection [Claxton 2012]	
1 Is it cost-effective?	Yes
2 Are there significant irrecoverable costs?	No
3 Does more research seem worthwhile?	Yes
4 Is the research possible with approval?	No
5 Will other sources of uncertainty resolve over time?	Yes
6 Are the benefits of research greater than the costs?	Yes
7 Are the benefits of approval greater than the costs?	Yes
Policy recommendation	**Approve**

Analyses also determine that, although treatment with clopidogrel initially has negative net health effects that are irrecoverable, these should not be regarded as significant, as the treatment decision for a presenting patient is irreversible in relevant time frames (results not presented here). This settles the second assessment ('Point2: Are there significant irrecoverable costs?'), with a judgement that irrecoverable costs are insignificant in this case study.

The third assessment ('Point 3: Does more research seem worthwhile?') requires a more in-depth consideration of the uncertainty and its sources. The implications of parameterised uncertainties for decision uncertainty suggest that approving clopidogrel for 12 months' use has a 0.5 chance of being incorrect[ii]. The consequences of this level of decision uncertainty (associated with parameterised uncertainties) were valued at 5,194 QALY (or £103.9

---

[i] This estimate is based on the cost-effectiveness decision analytic model. It is expressed in net health[3], and hence considers expected health opportunity costs (here valued at 1 QALY per £20,000).

[ii] This was informed by PSA, where parameterised uncertainties (commonly inferential in nature) were made explicit using statistical distributions, and Monte Carlo sampling undertaken to evaluate uncertainty over net health effects (NHEs).[20] This allows recording the proportion of simulations in which each intervention offers greater NHEs – which is interpreted as the probability of a particular intervention being cost-effective.

million)[iii], which exceeds the likely costs of further research. The parameter whose uncertainty is associated with the most consequences was relative effectiveness of clopidogrel on overall mortality.[iv]

Structural (not parameterised) sources of uncertainty were also considered. A source of structural uncertainty was informed by a *post-hoc* analysis of the RCT[53], which suggested that the relative benefits of treatment with clopidogrel may be highest within the first 3 months of treatment. A scenario was thus analysed considering an interaction between the relative treatment effect of clopidogrel and the duration of use.[v]   Although the main analysis was regarded as more credible by the Appraisal Committee at NICE, the scenario nevertheless carried some weight. Thus, the consequences of this additional source of uncertainty were quantified.[vi] When both scenarios are assumed equally credible, the between-scenario uncertainty is valued at 85 QALYs, which questions the need for further research to resolve this uncertainty.

Most of what might be gained from further evidence within this case study was associated with the parameters (namely relative effectiveness on overall mortality), with the value of research exceeding the likely costs of conducting

---

[iii] To ascertain the consequences of decision uncertainty, PSA was extended into EVI analysis.[17] These analyses simply use the difference between the NHEs of 12 months' clopidogrel and the option associated with greatest NHE for that particular simulation. For the clopidogrel case study, most commonly (52.4%) there were no consequences because 12 months' treatment with clopidogrel is the correct decision. When it is not, there is a greater chance of relatively small consequences of error (30% are below 10,000 QALYs), which occur predominantly when 6 months' treatment duration offers the highest NHE (18% chance). There is a small chance of much larger consequences (< 5% chance that they are above 30,000 QALYs). These occur only when standard NHS treatment offers the highest NHE. The expected consequence of uncertainty (5194 QALYs) is simply the average over this distribution. It can be interpreted as an estimate of the population NHE that could be gained over the time horizon of this technology if the uncertainty about treatment and its duration could be immediately resolved.

[iv] Analyses supporting this conclusion are a form of EVI that considers the impacts of resolving uncertainty over particular parameters of the decision model—parameter EVI[17] (further detail in pages 69-74 of Paper 1).

[v] Where more than one scenario might be regarded as credible there will be uncertainty *between* as well as *within* each of the scenarios. For the appraisal of clopidogrel, the alternative assumption made longer durations of clopidogrel less cost-effective and reduced the expected consequences of uncertainty from 5,195 QALY to 3,969 QALYs.

[vi] That is, what might be gained if evidence could immediately distinguish which scenario was 'true'. Methods for quantifying the consequences of between-scenario uncertainty were developed within this work (reported in Appendix 11 of Paper 1).

the research. Therefore a judgement at this point that more research might be worthwhile seems reasonable ('Point 3: Does more research seem worthwhile?'). Research was, however, not considered possible with approval (point 4: 'Is the research possible with approval?') because the type of experimental design needed to provide more precise estimates of the relative effects of clopidogrel is unlikely to be possible if treatment is already approved for widespread NHS use.

An assessment of point 5 in the checklist ('Will other sources of uncertainty resolve over time?') identified that clopidogrel was to become off-patent 7 years on from the appraisal and likely to see its price reduced. A price reduction means clopidogrel becomes more cost-effective, the consequences of uncertainty less pronounced and research less valuable.

At this point in the checklist, potential policy responses are 'Approve' or 'OIR' (not shown here). Analyses indicate that, if the research reports before year 3, the initial losses caused by restricting access to clopidogrel are offset by the potential gains from the research findings. But, if it reports later than 3 years, an unconditional approval would be more appropriate. When considering the possibility of research not reporting, a lower time-to-reporting period is required to make an OIR policy valuable. For example, when there is a 75% chance of research reporting, OIR should only be considered if research reports within 1 year, otherwise unrestricted approval is best (for a graphical display of this information, consult Figure 21 in Paper 1). In the case of clopidogrel, it seems unlikely that the type of research required could report quickly enough, therefore the benefits of approval are judged to exceed the opportunity costs (Point 7 in the checklist), and 'Approve' is more appropriate (Box 2).

This case study demonstrates the analyses needed to inform decision-making, and Paper 1 suggests numerous new graphical displays that help communicate findings and facilitate the judgements needed. It demonstrates that the checklist developed is an extremely useful tool to guide assessments and inform policy decisions.

Appropriate description of inferential uncertainties

The previous work demonstrates that policy responses on approval and research can be adequately informed when existing uncertainties are identified and appropriately characterised. But the evidence base can be limited in ways that are not easily described. For example, where evidence is absent on one particular comparator of interest, quantitative descriptions of its effectiveness cannot be produced. Or evidence may exist but show significant heterogeneity (estimated effects of the intervention differing across studies beyond what would be expected by chance[55]). Studies may also be flawed in their design and there may be the potential for estimates drawn to be biased, but the extent of bias unknown. Finally, existing evidence may even not be entirely relevant for the population or setting of interest.

There is a lack of specific guidance for these situations within current processes, and standard methods appear to fall short. When evidence is absent on a particular comparator, for example, this comparator is often excluded, leaving decision-makers to ascertain any implications qualitatively. When the evidence base is heterogeneous, the extent to which heterogeneity affects adoption and research decisions is often unexplored.

Thoughtful and novel approaches to describing uncertainties are thus required, particularly where the evidence base presents limitations. This thesis presents the development of strategies to better use available data in two contexts (detailed further in the next subsections): (1) where there is significant heterogeneity in the evidence base, and (2) where the sharing of information across relevant data can better inform decision-making. It uses the following as case studies:

- Paper 2 in the collection evaluated the clinical and cost-effectiveness of IntraVenous ImmunoGlobulin (IVIG) in severe sepsis, to establish whether or not an RCT is necessary and feasible, and whether or not the costs of carrying out an RCT are outweighed by the potential benefit of the resulting information. The contribution of this paper relates to in-depth analyses of the implications of heterogeneity to

decisions on the use of IVIG and on the need for (and value of) further research.

- Paper 3 relates to the clinical effectiveness of a medical device, Negative Pressure Wound Therapy (NPWT), in treating severe pressure ulcers. NPWT is a relatively expensive treatment used widely and claimed to speed wound healing and reduce infection rates. However, the evidence base supporting these claims is sparse. The contribution of this paper concerns the development of methods, particularly challenging under sparseness, to explore <u>heterogeneity</u> and to <u>share information</u> with related evidence (draw on similarities) when data on a particular comparator is uninformative.

- Paper 4 reports a technology appraisal conducted for NICE. It focused on appraising the effectiveness and cost-effectiveness of Tumour Necrosis Factor α inhibitors (anti-TNFs) for axial spondyloarthritis, an inflammatory rheumatologic disease. This work developed methods for the <u>sharing of information</u> to best describe the policy problem where complex relationships exist between evidence sources**.**

These examples provide important methodological and applied contributions to demonstrating both the feasibility and value of ensuring that unavoidable decisions are informed by the best use of evidence.

Heterogeneity

Where multiple studies investigate the same parameter their findings will inevitably differ. These differences may mean that the effects of the intervention will differ across studies beyond what would be expected by chance alone – this is often termed 'heterogeneity'[55]. Heterogeneity may result from clinical variation in important treatment effect modifiers, variation in the way effects have been measured, or bias.[55] Where heterogeneity is detected, its level is only of concern if decisions are sensitive to it. In this case, more stringent criteria can be used to select studies (in the review of the literature) but this may lead to relevant evidence being excluded. For example, in the context of a network meta-analysis (NMA) where evidence on multiple

treatments is synthesised together[15], stricter criteria may mean no evidence remains for one (or more) of the treatments of interest.

To avoid excluding relevant evidence it is often necessary to work with a heterogeneous evidence base. If there is a sufficient number of studies, heterogeneity can be described statistically. A particular methodology is often considered – the random-effects model[56-58] – which extricates the additional variation due to heterogeneity from the expected variation due to chance. Such a model, however, subsumes heterogeneity as additional uncertainty (through the predictive distribution[59]), but leaves it unexplained. This makes it difficult for policymakers to interpret decision uncertainty and consider the implications of heterogeneity for decisions about adoption and research.[60]

Issues regarding the use of random effects models are illustrated in Paper 2 of the collection. In this example, the evidence base on the effectiveness of IVIG on all-cause mortality was characterised by significant heterogeneity, with previous meta-analyses concluding that more evidence was required in the form of a large, well-conducted, RCT. The work in Paper 2 started by examining the potential causes of the considerable heterogeneity in the RCT evidence[vii], and identified a number of possible scenarios that considered not only measures of study quality, but also dose and duration of IVIG therapy. There was no clear clinical rationale for the impact of dose and duration on the effectiveness of IVIG, exposing a lack of evidence for understanding its mechanism of action in severe sepsis.   The implications of each scenario on cost-effectiveness were further explored[viii]. The results show that any conclusions regarding the cost-effectiveness of IVIG were highly sensitive to the choice of scenario. The most favourable estimate of the incremental cost-effectiveness ratio (£16,177) was obtained using a random-effects model. In

---

[vii] A Bayesian network meta-analysis was used to synthesise the effectiveness data. A formal selection process was designed and implemented to identify the key covariates (and combinations of covariates) explaining heterogeneity and to adjust for potential confounding. For further details on methods see Paper 2, page 17 of the full report[61], or the accompanying publication[62].

[viii] A de novo cost-effectiveness model was developed, and the literature carefully examined to inform the structure of the model and the values of parameters[61]. Analyses of two sources of primary data were required, which identified further sources of heterogeneity in short- and long-term outcomes (see paper 2 or full report[61] for further details).

the least favourable scenario IVIG appeared dominated when a random-effects model was used with an adjustment for publication bias. EVI analysis was conducted[ix], and across the majority of scenarios a study collecting data on the relative effectiveness of IVIG appeared the most efficient research design to invest in. However, results on the value of conducting such research were sensitive to the clinical effectiveness model used. Because there was no clear clinical rationale for the different scenarios, and despite the need for a further RCT, designing this study was deemed complex when uncertainties at this level existed. The research team hence recommended that further research focuses on filling the mechanistic knowledge gaps prior to a multicentre RCT.

This example illustrates that merely describing heterogeneity can misguide decisions: in the case study, had a random-effects model been adopted without further consideration for the sources of uncertainty, a further RCT would have been recommended without acknowledging important design issues, and thus valuable resources would have been potentially wasted in an inappropriately designed study. Instead, it is important to make the best use of the available data by attempting to ascertain the sources of heterogeneity and considering these in determining appropriate policy responses.

In the example described above, heterogeneity could be examined analytically. There are other cases, however, where there may be evidence of sources of clinical or methodological heterogeneity (e.g. differences in the characteristics of the patients included across trials, or in the design of the studies), but its implications are difficult to explore statistically because the evidence base is sparse. A sparse evidence base refers to a small number of events being reported for one or more comparisons because, for example, few individuals are recruited or the follow-up time is short. In general, small RCTs have the potential to be more heterogeneous in their design, as well as less detailed in their reporting. Additionally, small sample populations run a higher risk of

---

[ix] EVI analyses were undertaken (see Paper 2), including EVPPI and expected value of sampling information analyses, EVSI. The EVSI typically requires two nested simulation procedures. The decision model implemented here, however, applied relative treatment effects to short term benefits only (structured as a decision tree) and hence long-term outcomes did not depend on the treatment received. Because of this, we were able to avoid simulation for one of the two nested expectations. Further details are reported in Appendix 5 (page 140) of the full report[61] that led to Paper 2.

being unrepresentative — though poor reporting can make this difficult to assess. Although it is important to evaluate the extent of heterogeneity[63], sparse data may limit investigations. For example, the number of studies in each particular comparison may be insufficient to allow for statistical heterogeneity to be described and explored, in which case heterogeneity cannot be easily distinguished from the effects of the treatments themselves.

Where data are sparse, strong assumptions are normally required to enable modelling and estimation, but at the same time these assumptions are difficult to test. Firstly, it may not be possible to evaluate alternative model specifications, and secondly, under sparseness, uncertainty will be large and statistically significant results are unlikely.[64] In this context, NMA may offer some advantages compared with standard meta-analytic approaches as strength may be borrowed across the entire network through indirect evidence and the sharing of parameters across trials and/or treatments[65].

A relevant example of a sparse evidence base supporting an NMA is presented in Paper 3. In this application, different studies recruited individuals with different grades of ulcers, a measure of severity known to affect healing. Some studies recruited only less severe cases (with grade 3 ulcers), others only more severe cases (grade 4 ulcers), and other studies recruited a mix of patients, with some of these reporting subgroup outcome estimates (see Figure 1 in Paper 3 for the network diagrams). This raised the need to explore whether grade could act as a treatment effects modifier, i.e. be a source of heterogeneity. In our analysis, an initial modelling framework was developed in which both baseline hazards and treatment effects were modified by grade.[x]

---

[x] A Bayesian NMA model was developed, specifying a binomial likelihood for the number of healed patients in each (sub)group, and a cloglog link for the probability parameter, allowing for the observed proportions of participants healed to be a function of duration of follow-up. For RCTs reporting outcomes on grade-specific (sub)sample(s) of participants, the log-hazards of healing were regressed linearly on indicator variables for ulcer grade, treatments and on an interaction term between these. The baseline log-hazard for grade 3 ulcers was left unconstrained. The outcomes observed in studies reporting outcomes on a mixed-grade sample of patients (five RCTs) were expressed as a function of the proportion recruited from each grade (observed or imputed using a hierarchical beta–binomial model) and of the grade-specific log-hazard parameters, although the latter quantities were not observed within this subset of trials. This means the information was shared across the two sets of trials.

This approach allowed the available data from RCTs to be fully utilised. Results demonstrated that grade determines the absolute hazard of healing but it was less clear whether it modified the effectiveness of treatment. Given the sparse data, not obtaining a statistically significant result for the effect of grade on treatment effect was not surprising; however, it made the relevance of the findings unclear. As an alternative, a simpler modelling approach was implemented that did not include grade of ulcer as a treatment effect modifier, but where the effect of grade of ulcer on baseline hazard of healing was embedded in the study-level baseline hazards.[xi]   Results showed that the data fitted better in this simpler model.[xii]

This example illustrates that exploring heterogeneity using a sparse evidence base can be difficult as, on the one hand, results must reasonably reflect the uncertainties that are shown in the data but, on the other hand, analyses are based on difficult-to-verify assumptions owing to the lack of robust data. In this situation, novel approaches for pooling the evidence need to be explored, and Paper 3 of the collection illustrates such a framework, where the potential implications of treatment effect modifiers are made explicit.

Sharing of information

Where the evidence base is limited there are ways to impose a level of sharing of information between related data which could address these limitations. For example, evidence on one or more endpoints could be used to strengthen inferences about others (through, for example, multivariate meta-analyses[66,67]), or surrogacy relationships used to strengthen inferences over final endpoints[68]. Actually, the abovementioned example in Paper 3 already imposed a level of information-sharing, with the effect of ulcer grade on

---

[xi] The model is analogous to the previous one. However, given grade was here not explored as a treatment effect modifier, subgroup-specific outcome data was not used. Instead the healing events observed within each treatment group were modelled, with baseline healing rates left unconstrained to embed heterogeneity. Additionally, alternative models were implemented that tested assumptions made - details are outlined in Table 4 and Section 4.2 of Paper 3. We also tested the inclusion of observational data to strengthen inferences on the effect of ulcer grade on the baseline, and of elicited data to strengthen inferences over treatment effects.

[xii] This suggests that either the effect of grade was misspecified in the initial framework, or that the existing heterogeneity was not fully explained by grade.

effectiveness being assumed equal across treatments. These methods have a broad application, yet they have only been seldom applied within HTA. Instead, evidence is typically either completely excluded or considered perfectly generalisable, leading to poor characterisation of the available evidence, and of the existing uncertainties.

The first situation analysed within this thesis concerns the case where treatments within a class are expected to have similar (but not equal) effectiveness, and decision-makers require a formal exploration of the implications of such similarities.[69] In Paper 4, effectiveness evidence did not demonstrate any particular anti-TNF treatment for AS to be more effective than another. Analyses firstly adopted the standard approach, where treatment effects are assumed independent (or different) across treatments. To explore alternative assumptions and best inform decision-making, synthesis models were then extended (Chapter 5 in Paper 4) by considering common (equal) treatment effects across anti-TNFs and, alternatively, a 'class effect'[69-71] representing an assumption of similarity within the class (via a random effects distribution).[xiii] These analyses show that, in line with the published evidence, the different treatments have a similar effectiveness in AS.[xiv] However, unless we believe this assumption to hold *AND* the trials to be homogeneous in design and in the populations included, adopting a model assuming equal effectiveness across treatments would misrepresent uncertainty. The class effect model fits the data equally well to the common effect model, and may lead to more appropriate descriptions of uncertainty.[xv]

[xiii] The abovementioned models are identified as modelling approach A in Paper 4.

[xiv] Results have shown that the model considering a common effect for all anti-TNFs shows a better fit than the one estimating a different effect for each anti-TNF.

[xv] There are several summaries that can be derived from a class effect model, each with different implications for decision making.[59] A single effect for the class of treatments can be derived using the mean of the random effect distribution, with uncertainty reflecting *between-treatment* differences. Alternatively, the predictive distribution can be used, with uncertainty reflecting both between- and within-treatment differences. Hence the predictive distribution is here judged a more relevant summary of the overall effectiveness of the class. Drug-specific estimates can also be retrieved from a class-effect model (shrunken estimates, Table 70 in Paper 5); these borrow strength from the common class effect and estimates are shrunken towards its mean.

Paper 3 describes another example of the application of such a 'class effect', where the effectiveness of NPWT and 11 relevant comparators (belonging to a common class of dressings) was analysed. In this example, however, the application of the class effect served a different purpose. Because most links in the network were informed by a single study and the number of healing events in some trials was small or zero, it was not possible to obtain inferences on one of the comparators, a foam dressing. This would have precluded the inclusion of this treatment as a comparator in the cost-effectiveness study. To overcome this limitation of the evidence base, a 'class effect' was again used to describe the differences between the effectiveness of the alternative dressings. By sharing information in this way an estimate of the effectiveness of foam was obtained, encompassing not only the vague information from its trial reporting zero events in each arm, but also considering that the effectiveness of foam is expected to be within the spectrum of values observed across the class.[xvi]

The second situation analysed in this thesis where information sharing is relevant is where the evidence base is complex and standard methods of synthesis inadequately represent the decision problem.[12] For example, separate analyses of multiple outcomes reported in trials may not lead to appropriate evaluations of the parameter inputs required for an economic model.[72] Three related facets of this issue were addressed in Paper 4 in the collection. The first relates to the main outcome, a score of disease activity (the Bath Ankylosing Spondylitis Disease Activity Index, or BASDAI) that has been differentially reported across trials using two statistics: the absolute change in the scores from baseline, and the proportion of individuals that present a 50% change in score in relation to baseline (BASDAI 50). While some trials reported both statistics, others reported only one. Both of these measures were, however, relevant to the decision problem as a non-response to BASDAI 50 leads to discontinuation of treatment, and absolute changes in the scores were used to determine the magnitude of initial response to treatment and to extrapolate the effects of treatment. Given that these

---

[xvi] In this case study, the primary aim of the class effect used was to obtain an estimate for a particular treatment with uninformative data. Shrunken estimates were thus further used to inform the decision model.

outcomes were both central to informing the decision model, a model for the synthesis was developed that formally expressed BASDAI 50 as a function of the absolute change in BASDAI.[xvii] This required a multiparameter specification of the synthesis model to allow evidence on different functions of parameters to be used appropriately[73-75]. In this way, inferences can be informed by evidence on both outcomes, and all studies in the evidence base could be included.[xviii]

The second facet relates to the fact that many of the trials also report treatment effects over another outcome of treatment, the BASFI (Bath Ankylosing Spondylitis Functional Index), a measure of function rather than disease activity. In the economic model, the BASFI was used together with BASDAI score to determine the long-term QALY and cost burden of the disease. Treatments improving symptoms are expected to affect both disease activity and function and, particularly given the need to model these two outcomes together in the decision model, in Paper 4 the synthesis was also extended to consider potential correlations between effects over BASDAI and BASFI scores[xix]. Results show that estimates on BASDAI treatment effects are more precise, reflecting support for inferences from the data on BASFI; the

---

[xvii] This is referred to as modelling approach B in Paper 4. Inference was Bayesian. Data on the mean change in BASDAI score from baseline was assumed to have a Normal likelihood, and treatment effects were specified as differences in change scores across treatments. The likelihood for the BASDAI 50 data was assumed Binomial. The probability parameter of this distribution was then related to the change score using the BASDAI score at baseline (assumed correlated with the change score using a bivariate normal distribution). The correlation parameter was estimated separately for placebo and anti-TNF treatment. A class effect was assumed across treatments. See Appendix 12 in Paper 4 for a fuller description of methods.

[xviii] The class effect of anti-TNFs on change score was evaluated as slightly higher after the inclusion of BASDAI50 data in comparison to modelling approach A. Modelling approach B also returns more precise estimates of the pooled change score. This modelling approach, besides synthesising absolute change scores across a wider set of trials, can also be used to evaluate BASDAI 50 response for a specific baseline BASDAI score and change score in the placebo arm.

[xix] This was referred to as modelling approach C in Paper 4. It extends modelling approach B to include data on mean change in BASFI score, described using a Normal likelihood. As with BASDAI, treatment effects over BASFI were specified as differences in change scores across treatments. Treatment effects on BASFI were modelled as correlated to those on BASDAI across trials using a multivariate NMA approach[66,67,76]. The variation in treatment effects for both BASDAI and BASFI, and the correlation parameter between these, were estimated from the data. A class effect was assumed across the different treatments.

correlation between outcomes observed in the data allows inferences in BASDAI to borrow strength[76,77] from those on BASFI.

Finally, the third facet of this work related to evaluations of the absolute change in BASDAI for responders separately to non-responders, that is, conditional scores. This was important, given that response determined discontinuation. However, the published clinical effectiveness evidence did not report the conditional scores. Paper 4 demonstrates that a synthesis model that appropriately considers the two BASDAI variables can be extended to allow an evaluation of these conditional scores.[xx]

The examples in Papers 3 and 4 of the collection illustrate that sharing information can address important limitations of the evidence base and generate appropriate descriptions of the evidence for inclusion in the decision model – in a way that better reflects the decision-making context and better supports the adoption and research decisions being made.

Using formally elicited experts' beliefs

Whilst the above subsection considered the best use of *data*, there are circumstances where judgements are still needed to reach a decision. In this situation, the views of relevant experts can be sought to complement existing evidence, to generate descriptions of uncertainty where evidence is absent, or to determine the level of support for particular assumptions. The fact that, under limited evidence, judgements may be required for a decision to be reached creates a *prima facie* case for the use of elicitation methods with experts.[78] Whilst experts are consulted at several stages of formal decision processes, such as the appraisal processes facilitated by NICE, consultation is conducted informally and qualitatively despite being generally recognised that formal and structured methods to elicit experts' beliefs reduce bias and the use of heuristics[79], and should thus be applied.[80]

[xx] From modelling approaches B and C, it is possible to derive the conditional change score in BASDAI50 responders and non-responders using a simulation procedure that considers variability (i.e. variation across individuals). This is described in detail in Appendix 5 in Paper 4.

Methodological challenges in the design and conduct of structured elicitation exercises in HTA

To date, structured expert elicitation (SEE) has only been used to a very limited extent in HTA.[81] Contributing to this may be the fact that, despite methods of SEE having received significant attention in disciplines like engineering and Bayesian statistics[80,82], there is a lack of specific guidance specific to health care decision-making. This creates a number of challenges for the use of SEE in HTA, many of which have been identified in Paper 5, which reports the design of a large exercise aimed at systematically capturing clinical experts' knowledge and uncertainty about the speed of healing of severe pressure ulcers, and related events (such as wound-related complications), including beliefs about the impact of the alternative treatments (relative effectiveness). This exercise was specifically designed to inform the decision model over the use of NPWT for pressure ulcers, and follows from the work in Paper 3 in the collection.

The overarching concern expressed in Paper 5 relates to the fact that health carers, the experts identified as substantive experts, may not necessarily have advanced quantitative skills.[83,84] This conditioned many aspects of the design of the exercise. A face-to-face meeting was planned to allow for substantial training to be delivered. Homogeneity was imposed in the quantities elicited (only proportions were elicited) in order to simplify the task and abridge the training session, and the histogram method[85] was used for its intuitiveness. In specifying the quantities to elicit, the parameters of primary interest were expressed in terms of directly observable quantities, whose distribution(s) are thought to be easier to elicit[80]. For example, transition probabilities were not elicited directly and instead information was sought on the proportion of patients healed within given time periods (more than one time period was elicited to allow the possibility of time dependency). Defining observable quantities is particularly challenging in the context of decision modelling because, firstly, the model imposes relationships between the quantities that

need to be respected and, secondly, because dependencies (correlations) need to be avoided as eliciting these is particularly burdensome.[78,86] [xxi]

Other key methodological choices were explored using a pilot study, namely regarding the use of individual or consensus approaches to aggregation[87], and whether to use calibration[88] to generate weights for the different experts within an individual elicitation approach[89] (more details in Paper 5). In the pilot, the consensus method produced more accurate responses but also generated incoherent probability statements between two related quantities. These results are in line with the view that consensus promotes the use of heuristics and may produce overconfident judgments.[90] Also, the pilot showed that calibration-based weights were sensitive to the seed question, but that it was unclear how to specify relevant seed questions.

In the main exercise, experts were asked to express their beliefs individually and they were discouraged to interact. Calibration was not undertaken and elicited quantities were mathematically pooled across experts to generate a single description of uncertainty for each parameter.[91] Experts were weighted equally.

In total, twenty-three nurses attended and completed the main elicitation exercise. Each answered more than 30 questions, 18 of which were uncertain quantities. The results demonstrate that experts expressed a range of opinions and carefully considered uncertainty (Table II in Paper 5). Participants judged the exercise to be challenging but indicated their responses to be face-valid. This application demonstrates that carefully designing an elicitation exercise is crucial to its validity.

[xxi] Several decisions were made concerning how to elicit transition probabilities under time dependency, events related to transition probabilities, and relative effectiveness. Further details can be found in Paper 5, but to exemplify I will here summarise how relative effectiveness was elicited. To avoid eliciting dependency, an approach based on conditional independence was devised. Each expert was first asked to record the probability (with uncertainty) of a patient that received a reference dressing being healed. Then, the expert was asked to assume that the value she/he expressed to best represent her/his knowledge was true (reference value). We then elicited the absolute effectiveness of NPWT (with uncertainty). Because the experts conditioned their judgement about the effectiveness of NPWT on a plausible value for the comparator, when analysing this evidence this relationship was assumed to sustain for any other possible values of the absolute effectiveness of the comparator.

Elicited evidence in support of policy decisions

A review of applications reporting the use of formal methods of elicitation in HTA[81] identified 14 studies conducted in a number of policy contexts, ranging from an 'early modelling' context where experts' beliefs support decisions regarding the research and development pathway[92], to treatments such as NPWT which are extensively used in clinical practice despite a limited evidence base and where the experience of experts may best represent the current level of knowledge to support decision-making.

In all contexts, elicited judgements can support and even shape policy decisions; the NPWT example found that elicited evidence was particularly important for the comparator for which less evidence existed, foam. Recall that RCT data on the effectiveness of this parameter was uninformative; standard methods of synthesis would not have been able to describe its effectiveness, and thus the decision model would have likely excluded this comparator. In Paper 3 in the collection, uncertainty over the relative effectiveness of foam was described using the range of effects observed across treatments of the same 'class' described using a random-effect distribution. Foam's effectiveness was thus described with wide uncertainty, reflecting the differences between existing dressings; however, a higher likelihood of effectiveness was allocated to the central area of the distribution. This meant that, based on existing evidence[xxii], foam showed a considerable probability of being effective and cost-effective (Paper 6).[xxiii] However, the experts consulted as part of the formal elicitation exercise (Paper 5) consistently indicated foam to be expected to be less effective than other dressings, despite indicating substantial uncertainty over this result. Using such elicited evidence alongside

---

[xxii] Existing evidence (identified using broad literature searches) on a number of aspects of the evaluation was used, namely on effectiveness on healing (Paper 3), and on costs and utilities. We did not find evidence on the impact of complications, closure surgery, and recurrence in UK patients with pressure ulcers or on the impact of severe pressure ulceration on death.

[xxiii] A *de novo* cost-effectiveness model was developed for this application, with clinical input to
ensure that it adequately represented the clinical trajectory of patients with severe pressure ulceration.

existing evidence[xxiv], the likelihood of foam being cost-effective reduced substantially, and analyses of the implications of uncertainty did not indicate that further research on this treatment was worthwhile.[xxv]

This case study demonstrates that multiple sources of available evidence can be included in an evaluation in a transparent way, and that the presentation of cost-effectiveness estimates after the addition of data from each source into the model ca be useful to emphasise the different conclusions that would have been drawn if analyses were limited to only one source of data. The value of further research analysis was possible when it was most important, that is, when the existing evidence base was clearly insufficient and further research needs to be prioritised. Alternative study designs were evaluated, including not only a range sample sizes of an RCT but also possible comparator arms and different follow-up times. Given the sparse evidence base, a key source of information in this case study was the beliefs of clinical experts, which allowed shaping research priorities by streamlining further research to not focus on foam as a comparator in a future RCT.

Conclusions

This thesis focuses on a fundamental, but commonly overlooked, aspect of health care policy: the unification of decisions on wider access to technologies and recommendations for further research. While the risks associated with uncertainty cannot be wholly avoided, further evidence collection can allow a

---

[xxiv] The impact of each source of evidence on cost-effectiveness was evaluated by considering the accumulation of evidence using 3 scenarios: 1) existing evidence alone, 2) existing evidence combined with elicited evidence, and 3) the most comprehensive evidence base that also considered data generated as part of a pilot RCT. The collation of evidence sources, necessary to generate the second and third scenarios, required Bayesian updating. Further details can be found in Paper 6.

[xxv] EVI analyses were conducted using the most comprehensive evidence base (scenario 3). For this case study, it was particularly relevant to evaluate alternative possibilities for comparator arms in future research; however, if there is correlation between treatment effect estimates, value of information analyses cannot consider these parameters independently. An analytical approach to evaluate the plausibility of assuming independence was developed (details presented in pages 428-430 of Paper 6). Alongside, an approach to evaluate the extent and impact of nonlinearity based on the parameter EVI was also developed to allow simplifying EVSI calculations.

timely revision of the decision and avoid the consequences of uncertainty for prospective patients. It is common for access decisions to be uncertain, not just because the evidence base rarely provides definitive answers (even at launch), but also because technologies are commonly priced at the level at which they are just cost-effective (value-based price) where decision uncertainty is highest. Approving at the value-based price in the absence of further research means potential losses associated with uncertainty fall on the NHS (i.e. on population health). If a lower price is agreed, however, the consequences of uncertainty are either shared or borne entirely by the manufacturer (if the price is such that the consequences of uncertainty are no longer significant). It is thus crucial for policy to ensure that technologies are appropriately down-priced, or to ensure that important further research is conducted.

This thesis presents the principles and assessments required to support such joint adoption-research policy decisions, which have been developed for NICE (albeit these are easily generalisable). The principles underlying the work have been incorporated into NICE processes, but the detailed assessments have not. This signals that future work may be needed to more smoothly integrate these into policy. Such integration may require exploring current processes and how these can be adapted to i) allow for the additional assessments required[93], ii) align incentives for the further research required to be conducted in a timely fashion, and iii) consider the format reappraisal decisions would take (acknowledging the potential difficulties of reversing an adoption decision). Methodologically, it is also important to get stakeholders familiarised with the additional analyses required (EVI), and investigate the potential of simplified or streamlined modifications to these complex methods that widen their application without impacting on the rigour of analysis and resulting output[31,94,95]. Finally, there is also value in examining different performance-linked risk-sharing schemes in terms of their profile of consequences of uncertainty[96].

Whilst the final approaches developed for joint adoption-research are based on extensions of well-developed methodologies (cost-effectiveness, probabilistic analysis and EVI analyses), their application requires that *all* the

limitations of an evidence base are described and quantified explicitly as uncertainties. The thesis then proceeds to illustrate (using a set of case studies) how particularly difficult-to-describe limitations of the evidence base can be successfully dealt with, highlighting the potential usefulness of relevant indirect or external information (such as other data or expert elicitation).

The approaches proposed within the specific examples can, by themselves, be considered methodological developments, as these required the specification of innovative models for the synthesis of evidence that: best considered the available data; shared information appropriately; allowed for alternative specifications and the testing of assumptions; best described the complexities of the evidence base (required for modelling); and, crucially, responded to important, but difficult to describe, uncertainties and hence shaped policy decisions. But, in widening the use of such approaches to support policy, it would be important to develop a guided framework of analyses that forged strong links between methodological research and the specific policy context. A number of issues need tackling. The first is to determine when such analytical extensions offer important added value and should thus be explored within policy. Secondly, given these methods facilitate the use of a number of alternative additional sources of evidence that may not have previously been considered (including data and beliefs), it is important that a set of principles guide the choice (or prioritisation) of the alternative data sources. This may, for example, consider the potential for each source to explain (or relate to) any existing direct data. Thirdly, after choosing the key evidence source(s) for focus, there are methodological choices to be made on how to most appropriately incorporate such evidence – further research should describe/develop alternatives and examine the level of sharing of information they impose to guide practice and interpretation. Finally, it is important to devise appropriate ways to quantify the contribution of the different sources to inference and to explore and understand their bearing on access and research decisions. This is because it is unlikely that the validity of using external/indirect data can be verified, or that the unbiasedness of elicited beliefs can be established. Whilst the work presented in this thesis demonstrates that the use of this evidence is important, particularly where otherwise essentially un- or misinformed guesses are needed to reach a

decision. However, policymakers may struggle to endorse conclusions that are contradictory to those based on the direct evidence alone, in the absence of further nuanced insights into how and why additional data influence findings. Also, indirect/elicited evidence should not be seen to replace good quality experimental evidence. Instead, it should help to inform the decision to acquire further evidence.

As a direct result of the work in this thesis, and in recognising the need for further development of methods, I have successfully secured funding from the Centre for Health Economics for a PhD studentship on '*Developing methods for the sharing of information in evidence synthesis to inform Health Technology Assessment and policy decisions*' (a student has been appointed and started work in October 2016, being expected to finish in October 2019).

In what concerns elicitation, the work undertaken as part of this thesis has become an exemplar application in HTA. It underlines the importance of appropriately designed and conducted structured elicitation exercises that can usefully inform health care decision-making. Currently, there is no guidance that supports methodological choices on expert elicitation, particularly in the context of health care decision-making. This context determines that experts are likely to be health professionals, not trained to express judgments in mathematical formats, such as probabilities. Additionally, to inform health care decision-making, information on a wide range of, sometimes complex, parameters is often required, such as costs, relative effectiveness measures, and time to event outcomes. Moreover, an assessment of uncertainty is required (that is, uncertainty in knowledge over average values), it being vital that uncertainty is not misrepresented (over- or under-confidence) or tainted by the distribution of outcomes across a patient population. Finally, it is important that heterogeneity in the subpopulations that different experts observe is appropriately reflected in the judgements elicited. These challenges require further investigation to hone optimal elicitation protocols. To this effect, along with a group of collaborators I was awarded a MRC methods programme grant to establish a protocol for elicitation to be used to support health care decision-making. This work started in November 2016 and is expected to complete in November 2018.

Appendices

Paper 1.

Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, McKenna C, Soares M, Spackman E, Youn J. Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development. Health Technol Assess. 2012;16(46):1-323. doi: 10.3310/hta16460.

Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development

K Claxton, S Palmer, L Longworth, L Bojke,
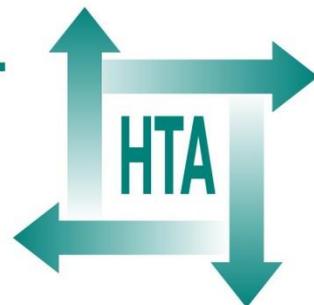S Griffin, C McKenna, M Soares, E Spackman
and J Youn

Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development

K Claxton,[1,2]* S Palmer,[1] L Longworth,[3] L Bojke,[1] S Griffin,[1] C McKenna,[1] M Soares,[1] E Spackman[1] and J Youn[3]

[1]Centre for Health Economics, University of York, York, UK
[2]Department of Economics and Related Studies, University of York, York, UK
[3]Health Economics Research Group, Brunel University, London, UK

*Corresponding author

Declared competing interests of authors: At the time that this research was conducted some of the research team at York and Brunel served on NICE Appraisal Committees (KC, SP, LL and SG), some of whom are also members of the NICE Decision Support Unit (KC, SP). York also provides analysis for the NICE appraisal process as one of a number of Evidence Review Groups.

This report should be referenced as follows:

Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, McKenna C, *et al*. Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development. *Health Technology Assessment*, 2012;**16**(46).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (*SciSearch*®) and *Current Contents*®/ Clinical Medicine.

NIHR Health Technology Assessment programme

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

> **Criteria for inclusion in the HTA journal series**
> Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme or, commissioned through the Methodology research programme (MRP), and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.
>
> Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

This issue of the Health Technology Assessment journal series contains a project commissioned by the Methodology research programme (MRP).

The Medical Research Council (MRC) is working with NIHR to deliver the single joint health strategy and the MRP was launched in 2008 as part of the delivery model. MRC is lead funding partner for MRP and part of this programme is the joint MRC–NIHR funding panel 'The Methodology Research Programme Panel'.

To strengthen the evidence base for health research, the MRP oversees and implements the evolving strategy for high quality methodological research. In addition to the MRC and NIHR funding partners, the MRP takes into account the needs of other stakeholders including the devolved administrations, industry R&D, and regulatory/advisory agencies and other public bodies. The MRP funds investigator-led and needs-led research proposals from across the UK.

In addition to the standard MRC and RCUK terms and conditions, projects commissioned by the MRP are expected to provide a detailed report on the research findings and publish the findings in the HTA monograph series, if supported by NIHR funds.

The final reports from MRP projects have been peer reviewed by a number of independent expert reviewers before publication in this journal series. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report. The views expressed in this publication are those of the authors and not necessarily those of the NIHR, MRC or the Department of Health.

Abstract

Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development

K Claxton,[1,2]* S Palmer,[1] L Longworth,[3] L Bojke,[1] S Griffin,[1] C McKenna,[1] M Soares,[1] E Spackman[1] and J Youn[3]

[1]Centre for Health Economics, University of York, York, UK
[2]Department of Economics and Related Studies, University of York, York, UK
[3]Health Economics Research Group, Brunel University, London, UK

*Corresponding author

Background: The general issue of balancing the value of evidence about the performance of a technology and the value of access to a technology can be seen as central to a number of policy questions. Establishing the key principles of what assessments are needed, as well as how they should be made, will enable them to be addressed in an explicit and transparent manner.

Objectives: The aims of this research are to (1) establish the key principles of what assessments are needed to inform an 'only in research' (OIR) or 'approval with research' (AWR) recommendation, (2) evaluate previous National Institute for Health and Clinical Evidence (NICE) guidance in which OIR or AWR recommendations were made or considered and (3) evaluate a range of alternative options to establish criteria, additional information and/or analysis that could be made available to inform the assessments needed.

Data sources: All NICE draft and final guidance up to January 2010 was considered in the review of NICE technology appraisal guidance. Four case studies were used to evaluate the range of options of what information and analysis could be made available to inform the assessment required. These were based on a reanalysis of existing health technology appraisals for NICE or the Health Technology Assessment programme.

Review methods: A critical review of policies, practice and literature was undertaken using traditional systematic searching based on initial search terms informed by key publications. An iterative approach was adopted using 'pearl growing' evaluated through capture–recapture methods. In addition, grey literature, policy documents and other sources, such as special interest groups and the expertise of the Advisory Group for the project, were used to contribute to this process.

Results: A series of recommendations, or options, for NICE to consider were developed with the involvement of key stakeholders. These establish the key principles and associated criteria that might guide OIR and AWR recommendations and identify what, if any, additional information or analysis might be included in the technology appraisal process, including how such recommendations might be more likely to be implemented through publically funded and sponsored research. To meet these aims the research is broadly structured as follows. A critical review of policy, practice and literature in this area

informs the development of a coherent conceptual framework to establish the key principles and the sequence of assessment and judgements required. This sequence of assessment and judgement is represented as an algorithm, which can also be summarised as a simple set of explicit criteria or a 7-point checklist of assessments. A review of previous NICE guidance in which OIR or AWR recommendations were either made or considered was undertaken to examine the extent to which the key principles are evident. The application of the checklist of assessment to a series of four case studies informs considerations of whether or not such assessments can be made based on existing information and analysis in current NICE appraisal and in what circumstances could additional information and/or analysis be useful. Finally, some of the implications that this more explicit assessment of OIR and AWR might have for policy (e.g. NICE guidance and drug pricing), the process of appraisal (e.g. greater involvement of research commissioners) and methods of appraisal (e.g. should additional information, evidence and analysis be required) are drawn together. At each stage this research has been informed by a diverse and international Advisory Group and the feedback from participants at two workshops involving a wide range of key stakeholders, which included members of NICE and its Advisory Committees (including lay members and other NICE programmes), patient advocates, manufacturers, and research and NHS commissioners, as well as relevant academics.

Limitations: Further research is required to establish how these considerations could be integrated within a practical value-based pricing scheme. In addition, irrecoverable opportunity costs are commonly associated with many health technologies that offer future benefits following treatment. The significance of these types of irrecoverable costs is not widely recognised and further research to demonstrate their potential impact more generally is needed.

Conclusions: The categories of guidance available to NICE have a wider application than is reflected in the review of previous guidance. Importantly, determining which category of guidance will be appropriate depends only partly on an assessment of expected cost-effectiveness. As well as AWR for technologies expected to be cost-effective and OIR for those not expected to be cost-effective, there are other important circumstances when OIR should be considered. In particular, for technologies expected to be cost-effective, OIR rather than approve may be appropriate when research is not possible with approval and OIR or even reject, rather than AWR or approve, may be appropriate even if research is possible with approval when there are significant irrecoverable costs.

Funding: The National Institute for Health Research Health Technology Assessment programme.

Contents

List of abbreviations

ACD	Appraisal Consultation Document
AG	Assessment Group
AHRQ	Agency for Healthcare Research and Quality
AWR	approval with research
BNF	*British National Formulary*
BSR	British Society for Rheumatology
CAD	coverage with appropriate determination
CCBT	computerised cognitive behavioural therapy
CED	coverage with evidence development
CFFE	conditionally funded field evaluation
CI	confidence interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CLOP	clopidogrel
CMS	Centers for Medicare & Medicaid Services
CSP	coverage with study participation
CURE	Clopidogrel in Unstable Angina to Prevent Recurrent Events
DMARD	disease-modifying antirheumatic drug
EECP	enhanced external counterpulsation
ERG	Evidence Review Group
EVPI	expected value of perfect information
FAD	Final Appraisal Determination
HAQ	Health Assessment Questionnaire
HRQoL	health-related quality of life
HTA	Health Technology Assessment
HTAi	Health Technology Assessment international
ICER	incremental cost-effectiveness ratio
ICS	inhaled corticosteroids
ICTRP	International Clinical Trials Registry Platform
IgE	immunoglobulin E
MeSH	medical subject heading
MI	myocardial infarction
MRC	Medical Research Council
MTA	multiple technology appraisal
MTAC	Medical Technologies Advisory Committee
MUST-EECP	Multicenter Study of Enhanced External Counterpulsation
NCD	national coverage determination
NETT	National Emphysema Treatment Trial
NFMI	non-fatal myocardial infarction
NHEs	net health effects
NICE	National Institute for Health and Clinical Excellence
NIHR	National Institute for Health Research
NRR	National Research Register
NSTE-ACS	non-ST segment elevation acute coronary syndrome
OCS	oral corticosteroids
OFT	Office of Fair Trading
OIR	only in research
OMAL	omalizumab
OWR	only with research

PASI	Psoriasis Area Severity Index
PATH	Programs for Assessment of Technology in Health
PPRS	Pharmaceutical Price Regulation Scheme
PSA	probabilistic sensitivity analysis
PsA	psoriatic arthritis
PsARC	Psoriatic Arthritis Response Criteria
PSS	Personal Social Services
QALY	quality-adjusted life-year
RCT	randomised controlled trial
REC	Research Ethics Committee
SIGN	Scottish Intercollegiate Guidelines Network
STA	single technology appraisal
TA	Technology Appraisal
TAR	Technology Appraisal Review
UKCRN	UK Clinical Research Network
WHO	World Health Organization

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or in the notes at the end of the table.

Executive summary

The general issue of balancing the value of evidence about the performance of a technology and the value of access to a technology can be seen as central to a number of policy questions. Establishing the key principles of what assessments are needed, as well as how they should be made, will enable them to be addressed in an explicit and transparent manner. The aims of this research are to (1) establish the key principles of what assessments are needed to inform an 'only in research' (OIR) or 'approval with research' (AWR) recommendation, (2) evaluate previous National Institute for Health and Clinical Evidence (NICE) guidance in which OIR or AWR recommendations were made or considered and (3) evaluate a range of alternative options to establish criteria, additional information and/or analysis that could be made available to inform the assessments needed. A series of recommendations, or options for NICE to consider, were developed with the involvement of key stakeholders. These establish the key principles and associated criteria that might guide OIR and AWR recommendations and identify what, if any, additional information or analysis might be included in the technology appraisal process, including how such recommendations might be more likely to be implemented through publically funded and sponsored research.

The relevance of this work, primarily to the NICE Technology Appraisal Programme, has been evaluated through two workshops involving key stakeholders, including members of NICE and its Advisory Committees (including lay members and other NICE programmes), patient representatives, manufacturers, and research and NHS commissioners, as well as relevant academics. Summaries of the research findings and key issues were provided in the form of briefing documents. These documents, which formed the basis of the workshop presentations and related group discussions, as well as a summary of feedback and list of participants, are available at www.york.ac.uk/che/research/teehta/workshops/only-in-research-workshop/.

The following elements of research form the basis of this report: (1) a critical review of policy, practice and literature in this area, (2) the key principles and the sequence of assessments needed, (3) a review of NICE technology appraisal guidance and (4) a checklist of assessments needed and its application to the four case studies using a range of additional information. Some of the possible implications for policy, process and methods of appraisal, distinguishing those issues directly relevant to the NICE remit and those that might be most relevant to other public bodies and stakeholders, were drawn together based on feedback provided at both workshops.

Key principles and assessments needed

The National Institute for Health and Clinical Evidence is increasingly making decisions about health technologies close to licence through the single technology assessment process. Inevitably these decisions are being made when the evidence base to support these technologies is least mature and when there may be substantial uncertainty surrounding their cost-effectiveness, as well as their effectiveness and potential for harms. In these circumstances further evidence may be particularly valuable as it would lead to better decisions that improve patient outcome and/or reduce resource costs. However, a decision to approve a technology will often have an impact on the prospects of acquiring further evidence to support its use. This is because, once positive guidance has been issued, the incentives for manufacturers to conduct research are limited. Also, the clinical community is unlikely to regard further randomised controlled trials to be ethical once positive guidance provides access with a funding mandate. Therefore, the decision to approve a technology should account for both the potential benefits of access to a cost-effective

technology and the potential costs to future NHS patients in terms of the value of evidence that may be forgone by early adoption.

The key principles and assessments needed fall into four broad areas: (1) expected cost-effectiveness and population net health effects (NHEs) (including benefits, harms and NHS and Personal Social Services costs), (2) the need for evidence and whether or not the type of research required can be conducted once a technology is approved for widespread use, (3) whether or not there are sources of uncertainty that cannot be resolved by research but only over time and (4) whether or not there are significant (opportunity) costs, which will be committed and cannot be recovered once the technology is approved.

Guidance will depend on the combined effect of all of these assessments because they influence whether or not the benefits of research are likely to exceed the costs and whether any benefits of early approval are greater than withholding approval until additional research is conducted or other sources of uncertainty are resolved. The key principles, represented by a sequence of assessment and judgements, can be summarised as a simple 7-point checklist that could be considered by Assessment Groups (AGs), Appraisal Committees and manufacturers:

- Is it expected to be cost-effective?
- Are there significant irrecoverable costs?
- Does more research seem worthwhile?
- Is the research possible with approval?
- Will other sources of uncertainty resolve over time?
- Are the benefits of research greater than the costs?
- Are the benefits of approval greater than the costs?

These principles suggest that the categories of guidance available to NICE have wider application than is reflected in our review of previous guidance. Importantly, which category of guidance will be appropriate depends only partly on an assessment of expected cost-effectiveness and hence this assessment should be regarded only as an initial step in formulating guidance. In general, as well as AWR for technologies expected to be cost-effective and OIR for those not, there are other important circumstances when OIR should be considered. In particular, for technologies expected to be cost-effective, OIR rather than approve may be appropriate when research *is not possible* with approval and OIR or even reject rather than AWR or approve may be appropriate even if research *is possible* with approval when there are significant irrecoverable costs.

Implications for value-based pricing

Any change in the effective price of the technology, either through patient access schemes or through direct price changes (possibly negotiated though a value-based pricing scheme), will affect the key assessments, leading to different categories of guidance. The price at which the technology would just be expected to be cost-effective is commonly regarded as the value-based price for the technology. This describes the threshold price below which approve rather than reject would be appropriate if OIR or AWR are not available as policy options. However, if they are available, there are often a number of relevant price thresholds. Once uncertainty and the need for evidence, as well as the impact of irrecoverable costs, are recognised, the threshold price that would lead to approve rather than OIR will always be lower than a single value-based price based on expected cost-effectiveness alone.

Even if price negotiation becomes possible alongside NICE appraisal, it will be important to retain OIR and AWR as available categories of guidance for two reasons: (1) there is no guarantee

that manufacturers will always agree to the lower price below which approve rather than OIR or AWR would be appropriate and (2) there may be many circumstances when no effective price reduction would make approve appropriate, for example reject or OIR guidance may be appropriate even if the effective price of a technology is zero if there is substantial uncertainty about its effectiveness and/or potential for harms.

Incentives for evaluative research

It is important that policy provides appropriate incentives for manufacturers to conduct the type of research needed to support NICE guidance at launch. The use of OIR and AWR guidance, and its link to effective price, provides clear signals and an incentive to ensure that the type of evidence which would require research that cannot be conducted once a technology is approved for NHS use is sufficient at launch. Therefore, a predictable OIR and AWR policy signals what type of evidence is likely to be most important at an early stage. It offers manufacturers a choice: (1) accept OIR guidance at a higher price but restricted volume, (2) reduce the effective price to achieve approval, or AWR where that is possible or (3) conduct the evaluative research at an earlier stage so that additional evidence at launch is not required.

How the NHS and manufacturers are likely to share the value of evidence might inform whether manufacturers should be expected to conduct the research specified in AWR or OIR guidance or contribute to the costs of publically funded research that may ultimately benefit their product. Two issues need to be considered: (1) the resource constraints on publically funded research may mean that other research priorities (often without commercial interest) may be more valuable to the NHS and (2) the success of AWR recommendations when manufacturers are asked to conduct the research will depend on whether NICE and/or the Department of Health are able to establish contractual arrangements as part of an AWR recommendation, that is, arrangements that can be monitored and enforced with credible penalties to ensure that agreed research is conducted and in the way intended. At present, NICE does not have a credible mechanism because removing approval of a technology simply because recommended research has not been conducted is not considered ethically appropriate or a credible threat.

The assessments that need to be made can also be used to consider what would be the value of (1) being able to conduct research while a technology is approved, (2) making evidence that is needed by the NHS available at launch and (3) being able to acquire evidence more quickly. This might inform a range of policies, such as early advice, public investment in early transitional and evaluative research or better data collection or information systems, that might make AWR possible. Understanding the relationship between the time taken for research to report and the value of the evidence to future populations can also help to inform (1) investments that might make research findings more quickly available, (2) the trade-off implicit in the choice of alternative research designs and (3) identification of those areas where, if research is to be undertaken, there must be confidence that it can report quickly.

How should assessment be undertaken?

Although the NICE appraisal process may be well suited to identifying the *need for evidence* when assessing cost-effectiveness, these other critical assessments (*the type of research and its priority*) are not necessarily ones for which NICE and its Advisory Committees, as currently constituted, have particular expertise, not least because they reflect the decisions of those responsible for research design, prioritisation and commissioning. Therefore, more informed judgements and better decisions might be possible through greater involvement of the research community. A

Research Advisory Committee could be constituted that could consider provisional OIR or AWR guidance, translating the need for particular types of evidence into particular types of research, costs, ethics, relative priority, likelihood of success and when the research is likely to report. The committee might also make recommendations about whether research should be publically funded or undertaken by the manufacturer with appropriate contractual arrangements (which may require the involvement of the Department of Health at some stage).

The order of considerations in the checklist and algorithm means that all seven assessments do not necessarily need to be made. Therefore, one model for an efficient process of assessment would be to consider points 1–5 routinely. The Appraisal Committee would then be in a position to either rule out OIR or AWR and issue guidance in the usual way or indicate in the Appraisal Consultation Documents that OIR or AWR was provisionally recommended subject to advice from a Research Advisory Committee and subsequent analysis to support an assessment of points 6 and 7 of the checklist before Final Appraisal Determination. This model would avoid unnecessary analysis and incorporate the judgements of the research community without necessarily delaying appraisal.

What additional information and analysis might be required?

Cost-effectiveness was presented in terms of NHEs per patient treated and for the population of patients over time. This provides information in a way that is directly relevant to the assessments that need to be made. All of the information required to express expected cost-effectiveness in this way is commonly available from the type of analysis already conducted during appraisal.

An early indication of the potential importance of irrecoverable costs can be based on their scale relative to expected NHEs; the point at which any initial losses are expected to be compensated by later gains; whether treatment decisions are reversible; and what opportunities to improve health might be forgone by a delay to initiating treatment.

The question of whether or not further research might be worthwhile (point 3 of the checklist) requires some assessment of (1) how uncertain a decision based on expected cost-effectiveness might be and (2) what the consequences, in terms of population NHEs, are likely to be if an incorrect decision is made. The methods of analysis decompose this into a series of steps, presenting what is available within current appraisal but in ways that can more directly inform the assessment required. Commonly, there is also uncertainty about alternative assumptions or judgements that might be made, often represented by alternative scenarios.

An assessment of the type of evidence needed (point 4 of the checklist) requires judgements about (1) how important particular types of parameters (inputs to the economic model) are to estimates of costs and quality-adjusted life-years, (2) what values these parameters would have to take to change a decision based on expected cost-effectiveness, (3) how likely is it that parameters might take such values and (4) what would be the consequences if they did, that is, what might be gained in terms of population NHEs if the uncertainty in the values of these parameters could be immediately resolved. The methods of analysis take these steps in turn, presenting what is available within current appraisal but in ways that more directly inform the assessment required. It is only when assessing the consequences of uncertainty associated with particular parameters that additional analysis is required to provide quantitative estimates.

The information required to assess whether other sources of uncertainty will resolve over time (point 5 on the checklist) requires information that is not commonly sought as part of NICE appraisal. It requires information about (1) likely changes in prices of the technology and its

comparators, (2) the emergence of new technologies that might make existing ones obsolete or change their cost-effectiveness and (3) other relevant research reporting. A number of potential sources of information and evidence were examined; however, many sources were either proprietary or public access was restricted, making it difficult to inform these assessments. When information and estimates were available they were often not complete or directly relevant to a UK context. NICE many need to consider how AGs and manufacturers can be provided with access to this type of information.

Funding

Chapter 1

Introduction and overview

The National Institute for Health and Clinical Excellence (NICE) is increasingly making decisions about health technologies close to licence through the single technology assessment (STA) process. Inevitably these decisions are being made when the evidence base to support these technologies is least mature and when there may be substantial uncertainty surrounding their cost-effectiveness, including their effectiveness and potential for harms. In these circumstances further evidence may be particularly valuable as it would lead to better decisions that improve patient outcome and/or reduce resource costs. However, a decision to approve a technology will often have an impact on the prospects of acquiring further evidence to support its use. This is because, once positive guidance has been issued, the incentives for manufacturers to conduct research are limited. Also, the clinical community is unlikely to regard further randomised controlled trials (RCTs) to be ethical once positive guidance provides access with a funding mandate. Therefore, the decision to approve a technology should account for both the potential benefits of access to a cost-effective technology and the potential costs to future NHS patients in terms of the value of evidence that may be forgone by early adoption. The general issue of balancing the value of evidence about the performance of a technology and the value of access to a technology can be seen as central to a number of policy questions. Establishing the key principles of what assessments are needed for 'only in research' (OIR) or 'approval with research' (AWR) recommendations, as well as how these assessments should be made, will enable them to be addressed in an explicit and transparent manner.

The Medical Research Council (MRC) and National Institute for Health Research (NIHR) Methodology Research Programme recently funded the Universities of York and Brunel to undertake research to help inform when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development (see *Appendix 5*). The NICE *Guide to the Methods of Technology Appraisal*[1] states that 'the Appraisal Committee may recommend that particular interventions are used within the NHS only in the context of research'. It indicates that four issues should be considered by the Appraisal Committee when recommending further research. These are (1) whether or not the intervention is reasonably likely to benefit patients and the public, (2) how easily the research can be set up or whether or not it is already planned or in progress, (3) how likely the research is to provide further evidence and (4) whether or not the research is good value for money.

The aims of this research are to:

1. establish the key principles of what assessments are needed to inform an OIR or AWR recommendation
2. evaluate previous NICE guidance in which OIR or AWR recommendations were either made or considered, and examine the extent to which the key principles from (1) are evident
3. evaluate a range of alternative options to establish the criteria, additional information and/ or analyses that could be made available to help the assessments needed to inform an OIR or AWR recommendation
4. provide a series of final recommendations, with the involvement of key stakeholders, establishing both the key principles and associated criteria that might guide OIR and AWR recommendations, identifying what, if any, additional information or analyses might be

included in the NICE technology appraisal process and how such recommendations might be more likely to be implemented through publically funded and sponsored research.

The relevance of this work to NICE has been evaluated through a series of two workshops involving key stakeholders, including members of NICE and its Advisory Committees (including lay members and members of other NICE programmes), patient representatives, manufacturers, and research and NHS commissioners, as well as relevant academics. Establishing the key principles of what assessments are needed to inform OIR or AWR recommendations requires a critical review of a diverse literature on principles and policy and previous NICE guidance (see *Chapters 2* and *4*); the development of a coherent conceptual framework (see *Chapter 3, Key principles and assessments needed* and *Changes in prices and evidence*); and consideration of whether or not such principles conflict with established ethical principles and the social value judgements adopted by NICE (see *Chapter 3, Social value judgements and ethical principles*).

The first workshop took place in September 2010 and considered four main topics:

1. the relevance to NICE of existing literature
2. whether or not the key principles and assessment that have been identified provide useful guidance on when OIR and AWR might be considered
3. the insights from a detailed review of previous NICE guidance
4. whether or not the proposed methods to inform assessment and criteria to select case studies are suitable.

Relevant summaries of the key issues were provided in the form of briefing documents that covered these four topics. These documents, which formed the basis for the workshop presentations and related group discussions, as well as a summary of feedback and list of participants, are available at www.york.ac.uk/che/research/teehta/workshops/only-in-research-workshop/.

The primary output of this workshop was a set of principles and explicit criteria (a sequence of assessments and decisions) to support OIR and AWR recommendations. This sequence of assessments (an algorithm) can be summarised as a simple 7-point checklist, which was subsequently applied to a series of four case studies. Each case study examines how each of the assessments might be made based on the type of evidence and analysis currently provided in NICE technology appraisals and how these assessments might be better informed with a range of additional information and/or analyses.

The second workshop took place in June 2011 and considered:

1. whether or not the revised algorithm of assessments and the associated checklist has identified the key judgements that need to be made when considering OIR and AWR guidance
2. based on the application of this checklist to the series of case studies, whether or not such assessments can be made based on existing information and analysis provided to NICE and in what circumstances could additional information and/or analysis be useful
3. what implications this more explicit assessment of OIR and AWR might have for policy (e.g. NICE guidance and drug pricing), the process of appraisal (e.g. greater involvement of research commissioners) and methods of appraisal (e.g. should additional information, evidence and analysis be required).

Relevant summaries of the key issues were provided in the form of briefing documents that covered these three topics. These documents, which formed the basis for the

workshop presentations and related group discussions, as well as a summary of feedback and list of participants, is available at www.york.ac.uk/che/research/teehta/workshops/only-in-research-workshop/.

The primary output of this workshop was a list of possibilities that NICE might choose to take forward in the next revision of the *Guide to the Methods of Technology Appraisal*[1] and how this might inform the formulation of guidance in the other NICE programmes. It might also suggest how a consideration of uncertainty and the need for evidence might influence value-based pricing and the impact of patient access schemes on OIR and AWR guidance.

The main report is intended to be accessible to a wide audience, providing intuitive explanations of why certain assessments are important and illustrating how they might be informed using examples. Notes have been used extensively, especially in *Chapters 3* and *5*, to provide explanatory detail without adding undue complexity to the exposition in the main text (see *Notes*). The main report is accompanied by (1) supporting material for each chapter (see *Appendices 1–5*), (2) a technical appendix (see *Appendix 6*), which provides a more formal treatment of why, in principle, each type of assessment is important, (3) full details of the analysis undertaken for each of the four case studies referred to in *Chapter 5* of the main report (see *Appendices 7–10*) and (4) a series of technical notes (see *Appendix 11*) that deal with some conceptual and analytical details that are common to the case studies reported in *Appendices 7–10*.

The main report is organised as follows. The results of a critical review of policy, practice and literature in this area are presented in *Chapter 2*, supported by material in *Appendix 1*. The key principles and the sequence of assessments needed are presented in *Chapter 3*, supported by *Appendix 2* and critically the technical appendix (see *Appendix 6*), which provides more formal treatment of the issues. A review of NICE technology appraisal guidance is presented in *Chapter 4*, supported by material in *Appendix 3*. The checklist of assessment needed and its application to the four case studies using a range of additional information and analysis is reported in *Chapter 5*. *Chapter 5* is supported by material in *Appendix 4* and critically by the full details of the analysis undertaken for each case study, reported in *Appendices 7–10*, and the series of technical notes in *Appendix 11*. Finally, *Chapter 6* briefly draws together some of the possible implications for policy, process and methods of appraisal, distinguishing those issues directly relevant to the NICE remit and those that might be most relevant to other public bodies and stakeholders. *Chapter 6* draws on the feedback provided at the second workshop, which has been summarised and is available online (see above).

Chapter 2

Critical review of policies, practice and literature

Aims and objectives

There is a growing and diverse literature on OIR/AWR[2–5] and on when decisions to adopt a technology should be delayed.[6,7] There is also considerable variation in the terminology used across this literature, which reflects the different decision contexts, not all of which are relevant to NICE. There is a need to critically review this literature to distil any common themes and core principles relevant to the NICE context. The main purpose of the review is to help to inform the development of a unifying conceptual framework within which these themes and principles can be located and understood. This, in turn, will enable a consistent and clear terminology to be established which will assist clarity in subsequent policy debates and may provide NICE with useful definitions and terminology that can be used in communicating its guidance, considerations and methods. The specific aims of the review were:

- to review alternative terminologies and taxonomies used to describe and classify approaches to OIR/AWR and to establish their relevance to the NICE context
- to identify any common themes and principles discussed in relation to OIR/AWR.

Methods

The existing literature on OIR and AWR is only partly represented in traditionally published papers; much is located in policy and discussion documents. The diversity in these sources was reflected in the range of search strategies employed, covering (1) traditional published literature, (2) grey literature and (3) policy and discussion documents. In addition, relevant interest groups and policy websites were searched, reference lists of previous reviews[8] were checked, separate citation searches were performed using key references and discussions were held with our Advisory Group. In reviewing the results of the systematic search and selecting relevant studies for inclusion, a relatively inclusive approach was adopted. Case study examples of AWR/OIR were not searched for explicitly; instead the focus of the searches was to identify relevant methodological papers. Case study examples comprising a significant discussion of methodological issues were, however, included in the review.

A summary of the search strategies are discussed under the relevant subheadings below. Full details of the separate searches are reported in *Appendix 1*.

Traditional published literature

A traditional systematic search was undertaken in MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE, The Cochrane Library, EconLit and Cumulative Index to Nursing and Allied Health Literature (CINAHL) based on appropriate search terms informed by key publications to identify relevant literature.[8] An iterative approach was used in the development of search terms and strategies incorporating 'pearl-growing' techniques through additional citation and reference searching of key publications and discussions with our Advisory

Group. Searches were restricted to documents produced from 1999 onwards as the use of OIR and AWR policies is a relatively recent process and so it is not expected that any relevant references will be identified before this date.

To test the performance of the search strategy, records initially identified from the case for support and citations contained within those documents and articles were cross-referenced with the results identified from the searches (capture–recapture method). In doing this it was concluded that the final search strategy identified all of the records it would be likely to identify, that is, those accessible from traditional (non-grey literature) sources.

Grey literature searches

Additional searches of grey literature were also undertaken using a similar approach reported in the review by Stafinski *et al.*[8] The following grey literature databases were searched: the New York Academy of Medicine Grey Literature Report and the IDEAS database (Department of Economics, University of Connecticut). A similar set of key words as for the review of the traditional published literature was used for the grey literature searches, with some modifications because of the different recording mechanisms operated by the grey literature databases. Searches were again restricted to documents produced from 1999 onwards.

Policy and discussion documents

The third element of the search was based on policy and discussion documents. Although the focus of the search related to UK policy and discussion documents, documents from key international organisations were also included. Relevant international organisations and websites were identified in conjunction with our Advisory Group. A number of policy documents were also identified from those referenced in the case for support and from the reference lists of key papers.[9–12] A number of additional organisation websites were searched for relevant documents. Where possible, web searches were restricted to documents produced from 1999 onwards for consistency with the other search elements.

Other sources

The final element of the search considered a variety of other sources including relevant interest groups [e.g. the Health Technology Assessment international (HTAi) interest subgroup on conditional coverage and evidence development for promising technologies and the Programs for Assessment of Technology in Health (PATH) website containing publications relating to the programme of conditionally funded field evaluations (CFEEs) in Ontario], other reviews,[8] references suggested by our Advisory Group and additional searches based on citations of papers identified from earlier stages.

Literature search results

A summary of the search results is provided in the following sections. Full details of the results and a summary of included references are reported in *Appendix 1*.

Identified references

In total, 59 references were included in the review; 43 of these were journal articles,[2–5,8,13–50] 11 were policy documents (eight UK and three non-UK)[9–12,22,51–56] and five were based on presentation slides or discussion documents.[56–60]

Of the 43 journal articles, the majority are academic 'think pieces' on issues relating to OIR and AWR with additional responses to previous articles.[5,13–16] Although a number of the journal articles provide a discussion of many of the general issues relating to the use of OIR and

AWR,[2,3,17–24] a smaller proportion explicitly address the issue of terminology and/or provide a taxonomy of OIR/AWR types ($n = 11$). The majority of the journal articles discuss specific issues relating to OIR/AWR implementation, most commonly in relation to evidence collection and design ($n = 41$). These issues are often illustrated in the context of specific case study applications such as the positron emission tomography register or the National Emphysema Treatment Trial (NETT). A relatively large number of these articles also discuss ethical issues and social value judgements relating to the implementation of OIR/AWR policies ($n = 20$) and issues of investment and reversal costs ($n = 21$).

The 11 policy documents[91–112,22,51–56] identified are from UK and non-UK sources. From the UK, relevant documents are produced by HM Treasury,[9] the Department of Health,[52] the House of Commons Health Committee,[10] NICE and the Agency for Healthcare Research and Quality (AHRQ),[51] NICE Citizens Council,[11] NHS Scotland,[53] the Office of Fair Trading (OFT)[12] and the Office of Life Sciences.[54] Outside of the UK, relevant documents identified are all produced by US organisations: the Centers for Medicare & Medicaid Services (CMS),[55] the Health Industry Forum[56] and the National Health Policy Forum.[22] In many of these policy documents OIR and AWR are typically discussed as part of a wider general health policy theme as opposed to being the central issue under consideration. Only a small number[22,51,55] offer any insights into terminology and/or taxonomies relating to OIR/AWR. Over half ($n = 7$) present general issues of OIR/AWR and specific issues most commonly related to evidence collection. Issues of investment and reversal costs, changing prices and ethical issues/social value judgements are covered to a lesser extent in the policy documents.

The final five references included are presentation slides or discussion documents.[56–60] These are all widely cited in the OIR and AWR literature. None of these references discusses terminology or taxonomies of OIR/AWR types. Instead they again focus on general themes of OIR/AWR, with some discussing specific issues such as evidence collection.

Summary of the key issues for practice and policies for 'only in research' and 'approval with research' schemes

The following sections discuss the main findings in line with the key objectives of the review.

Review of terminology and taxonomies of 'only in research'/'approval with research' schemes

Multiple definitions of OIR- and AWR-type schemes are reported in the literature. These are commonly provided within a broader consideration of conditional coverage or risk-sharing schemes. Despite the variation in terminology that exists, a number of common themes emerge. Most notably, the use of OIR/AWR is commonly defined as providing an alternative to a binary accept/reject decision for policy-makers in situations in which the technology does not appear to meet the standard criteria for reimbursement, predominantly because of uncertainty surrounding the existing evidence base and when additional data collection could reduce this uncertainty.[21] The emphasis placed on uncertainty and the specific role that the collection/generation of additional evidence plays in reducing existing uncertainty is what distinguishes OIR/AWR schemes from the broader range of conditional coverage or risk-sharing approaches,[25] in which the focus is to shift the burden of uncertainty onto another party (usually the manufacturer) rather than to collect information to reduce this uncertainty for future decisions.[4,24]

In considering how to appropriately define and categorise OIR/AWR schemes, it is important to consider what particular terms mean in their various contexts. In the context of NICE, OIR is the term used when a recommendation is made to restrict an approval decision to only those patients who subsequently receive the intervention as part of a well-designed programme of research.

OIR has been available as a formal policy option since the inception of NICE[52] and provides an important additional option to accept and reject decisions.[3]

In contrast to an OIR recommendation, the use of AWR does not necessarily limit coverage to those participating in the clinical study or registry. Hence, the distinction between OIR and AWR is primarily the degree of coverage that each confers for reimbursement purposes. Although AWR is not currently a formal policy option available to NICE, it is able to issue specific research recommendations as part of any guidance and can link this to the timing of any reappraisal. Consequently, AWR represents a valid option for consideration. Importantly, both OIR and AWR strategies are distinct from general recommendations for further research made as part of the appraisal process, in which no link to generating evidence as a condition of coverage is made. However, given the current directives and remit of NICE, the research recommendations issued as part of an AWR decision are not a mandatory requirement of approval. As a result, inevitably there exists some uncertainty following an AWR recommendation over whether or not the stated research recommendations will actually be conducted.

Although slightly different definitions and terminology are applied in the broader literature relating to OIR and AWR schemes, they are relatively similar in their meaning. For example, in the USA, the term 'coverage with evidence development' (CED) is often used as a catch-all term for OIR- and AWR-type schemes. Sections 1862(a)(1)(A) and (1)(E) of the Social Security Act provide statutory provision for the CMS to issue both OIR- and AWR-type coverage decisions involving the collection of additional evidence in registries or clinical trials, through national coverage determinations (NCDs). CMS describes two related but distinct processes: coverage with appropriate determination (CAD) and coverage with study participation (CSP).[55] As with OIR and AWR, these separate forms of CED are also closely linked to the level of coverage. CMS may issue a CAD to determine that patients receiving the treatment meet the conditions specified in the NCD. As part of this they may request more data. CSP allows coverage of certain items or services for which the evidence is not adequate to support full coverage and for which further data would be of benefit. Coverage may be extended to patients enrolled in a clinical research study. To recommend a CSP the evidence should assure basic safety, there should be high potential to provide significant benefit and there may be significant barriers to conducting clinical trials. Consequently, CSP would fit with OIR as it currently exists in the UK and CAD is closer to an AWR-type scheme.

Another example of the use of CED as a catch-all term for OIR/AWR schemes is the conditionally funded field evaluations (CFFEs) conducted in Ontario, Canada. These are recommended on the basis of a Health Technology Policy Analysis undertaken by the Medical Advisory Secretariat for the Ontario Health Technology Advisory Committee, which may conclude that there is not enough evidence to support uptake and diffusion of the technology. In these circumstances, coverage for a technology is provided conditional upon additional data being collected to specifically address residual uncertainty to better inform evidence-based decision-making. Primarily, the impetus for conducting a CFFE is a lack of evidence on transferability of evidence to a particular jurisdiction and/or its effectiveness or cost-effectiveness within particularly subgroups.[17,20] There are, however, reasons to recommend a CFFE on the basis of any decision uncertainty regarding the quality of evidence, safety data and cost-effectiveness.[20] In practice, coverage is sometimes restricted during the course of the field evaluation to only those patients participating in the study (e.g. the use of positron emission tomography scanners), that is, akin to an OIR recommendation; however, in other instances, hospitals can still purchase technologies and provide services through global budgets during the period of evaluation, thus appearing less restrictive than an OIR recommendation. Although this implicitly suggests that separate coverage schemes are considered when a CFFE is commissioned, there appears to be no formal distinction made within existing policy documents and other published literature between

different types of schemes and also no discussion of the specific factors that might influence the degree of coverage. Instead, the type of scheme and degree of coverage appear to be determined on a case-by-case basis.

As well as any commonalities between CED and OIR/AWR schemes, there also appear to be similarities in terms of the challenges faced in successfully undertaking these schemes. In the UK, an OIR scheme can be recommended by NICE when appraising health technologies, but it also has potential in public health, diagnostics and devices. However, following an OIR recommendation, there are no formal arrangements to develop the research study required to reduce uncertainties.[3] NICE does not hold a budget to commission research so unless it is publically funded by research commissioners it will be undertaken only if manufacturers conduct it, with the NHS contributing excess treatment costs. The lack of co-ordination also makes it difficult to ensure an update of the recommendation following production of new evidence.[3] In the USA, although the CMS will cover the costs of a trial or registry associated with a CSP decision, there is currently no Medicare-specific funding mechanism for the additional data collection under CAD and hence there exists similar uncertainty concerning who will be responsible for paying for the additional data collection under CAD.

Although there have been several previous attempts to develop taxonomies,[8,21,23,25] none of these has been focused specifically on OIR/AWR schemes and typically these are presented as part of a broader categorisation of conditional coverage and risk-sharing schemes. For example, in the taxonomy developed by Carlson et al.,[25] conditional coverage schemes are divided into CED and conditional treatment continuation schemes. Within CED, two subtypes are presented – OIR and 'only with research' (OWR) – with OWR similar to the term AWR used in this report. However, a more detailed consideration of OIR/AWR schemes and the potential for further subtypes within these schemes has not been previously explored in existing taxonomies.

General issues of 'only in research'/'approval with research' schemes

There are many issues that need to be resolved to enable the successful implementation of an OIR or AWR scheme both generally and also within the specific constraints of NICE.[11] Central to this is the need to clarify the objectives of these schemes and the relevant criteria for their use. However, the critical review identified only limited discussion on the specific circumstances under which an OIR or AWR scheme may be an appropriate policy option.[9,10,51] The lack of any clear guidance has led to concerns expressed over ambiguity regarding their use[26,57] and that OIR is currently being used as a 'polite no' by NICE.[11] Importantly, these concerns are not restricted to the use of OIR/AWR schemes by NICE. The general lack of clarity on the principles and criteria for using these schemes is reflected in many commentators' views that the development and use of schemes internationally has appeared to be rather ad hoc to date.[57] Such concerns clearly highlight the importance of developing a clear set of principles for the use of OIR/AWR by NICE.

Although the majority of existing policy documents related to NICE are not explicit about the rationale and principles for the use of OIR and AWR in the UK, the report from the NICE Citizens Council[11] on the use of OIR does describe the particular circumstances that the Council thought should be taken into account when NICE considers whether or not to issue OIR recommendations. These circumstances include:

- Whether at least one appropriate, relevant study is:
  - planned (e.g. the study will definitely start within 6 months of the guidance publication date)
  - in progress (e.g. recruitment to the study is open and is expected to last at least 1 year beyond the guidance publication date) or

- could be established quickly (analysis in *Chapter 5* explores the impact of the time taken for research to report).
- Whether or not the question addressed by the study will contribute to reducing the uncertainties identified during the preparation of NICE guidance.
- Whether or not the research is feasible (in terms of numbers of patients, recruitment, etc.) and is likely to deliver results within an appropriate time period.
- Whether or not a fully supportive decision would lead to significant irrecoverable fixed costs of implementation (the impact of these types of cost are explored in *Chapter 3, Technologies with significant irrecoverable costs* and *Chapter 5, Point 2: Are there significant irrecoverable costs?*).
- Whether a fully supportive decision, instead of an OIR recommendation, would lead to the termination of research in progress or prevent new research from beginning and thus have a negative impact on future collection of relevant information.
- Whether or not it is realistic to hope that research can be carried out to the satisfaction of NICE. Factors to be considered include the timeliness of the research, potential number of patients able to participate in research, the pace of the current research and the precise nature of the questions to be answered.

In setting out a clear rationale and set of principles for the use of OIR/AWR, NICE will also be able to work towards identifying which technologies may be suitable for such policies. Ideally they should be those with potential net benefit but also some degree of uncertainty.[4] It has also been argued that these schemes could also be used to 'fast track' particular treatments.[26] However, in addition to their role in new and emerging technologies,[3] other commentators have also stressed their potential use for established interventions to inform recommendations for increased investment or for disinvestment.[5]

In addition to the rationale and principles, there are also numerous practical issues that need to be resolved for the successful use of such policies. The recent lung volume reduction surgery case study highlighted a number of challenges for OIR/AWR, in particular significant opposition from the clinical community, the significant level of funding required, the length of time required to complete data collection and limited access for patients in remote areas. As a result of the multiple sclerosis risk-sharing scheme, the importance of interagency collaboration, achieving consensus on acceptable quality of evidence, external peer review, predefined clinical benefit and determining who pays for treatment was also apparent. There also remain other important challenges, including the need to ensure that research is actually conducted and is fit for purpose, as well as ensuring that the process is undertaken in a legal, ethical and acceptable manner.[51] Another important consideration is that these schemes need to be designed in order to develop appropriate incentives to produce evidence in a timely fashion and strategies need to be put into place to ensure that the research is actually carried out.

Within the UK, the importance of interagency collaboration has been highlighted as a key issue in ensuring the success of OIR/AWR schemes. In particular, the Cooksey report[9] recommended that arrangements between the Health Technology Assessment (HTA) programme, the NIHR Service Delivery and Organisation programme and NICE should be formalised so that recommendations for the use of interventions in the context of clinical studies can be operationalised.

Specific issues of 'only in research'/'approval with research'
As well as the more general issues that need to be resolved to ensure the effective use of OIR/AWR policy options, there are a number of specific issues that need to be addressed.

Evidence collection

Acquiring appropriate evidence following an OIR or AWR policy is of paramount importance.[25] Without an appropriately designed and conducted study, it is likely that little will be achieved in terms of reducing the uncertainty that led to the use of such policies in the first instance. OIR and AWR policies allow evidence to be generated specifically to inform decisions, a role not intended for traditional regulatory trials.[5] This raises a number of issues and potential challenges related to the design and funding of further research studies. First, there is currently very little in the way of formalised arrangements following an OIR/AWR recommendation in many countries.[18,28] One exception to this are CFFEs conducted in Ontario, which have a specific funding stream (albeit modest) covering the evaluation by PATH and additional monies for the fieldwork itself. However, in many instances this budget is not sufficient to cover the full costs of a CFFE and therefore other avenues must be explored, such as cost-sharing.[17,20]

A key issue identified in determining the success of these schemes is the development of working partnerships between stakeholders (clinical community, decision-makers and manufacturers).[3] Related to this is the issue of obtaining funding for OIR/AWR studies and establishing who pays for the research. In relation to NICE, it has been recommended that the relevant study should be either planned or currently in progress, or alternatively that a new study could be established quickly.[11] Without secure funding the research may never be undertaken and thus the uncertainties leading to an OIR/AWR recommendation will remain.

The design of the OIR/AWR study will ultimately determine its success[23] and some of the failures of existing schemes have been attributed to inappropriately designed studies.[57] Perhaps the most important consideration emerging from the literature is the issue of which type of study is most appropriate for an OIR/AWR scheme.[51] OIR/AWR research (unlike licensing research) is not confined to RCTs and, depending on the source of uncertainties, other types of evidence may be sufficient.[29] The choice of study is ultimately context specific and related to the source of uncertainty; however, it may also be influenced by factors such as cost and availability of suitable patients, collaborating clinical centres and potential ethical considerations. Good routine data capture mechanisms have a potentially crucial role to play in the feasibility of any scheme[18] and the development of health informatics could greatly reduce the cost of evidence.[18,55] Whichever study type is chosen, Tunis and Whicher[30] argue that discussion regarding the study design should not take place when the decision is made over who should pay for the study as this imposes restrictions.[30] Clarification is also needed on how the evidence collected as a result of a OIR/AWR policy will be used in an updated coverage decision[31] and also how many data are enough to inform subsequent decisions.[29]

Important lessons can be learnt from previous studies commissioned as part of an OIR or AWR policy. The multiple sclerosis risk-sharing scheme has been heavily criticised by many.[26] Despite its intention to reduce uncertainty[23] associated with the use of the disease-modifying drug therapies beta-interferon and glatiramer acetate in the UK NHS, the exclusion of a control group from the study has meant that it is difficult to determine effectiveness. Similarly, the Implantable Cardioverter-Defibrillator Registry failed to provide the CMS with data required to answer questions regarding patient survival.[56] Another widely cited example, a trial of lung volume reduction surgery (NETT),[5,22,27,28,32,33,57,58] is looked on favourably by some[32,57] but has also been criticised for taking too long[5] and for not addressing the uncertainties relevant for decision-making. A number of implications for future studies were also noted, in particular the importance of interagency collaboration, achieving consensus on acceptable quality of evidence, external peer review, predefined clinical benefit, determining who pays for treatment and establishing longer-term follow-up.[27]

Investment and reversal costs

Investment and reversal costs have also been identified as relevant considerations in the existing literature. In particular, NICE needs to determine whether a fully supportive decision (as opposed to OIR) would lead to significant irretrievable costs of implementation and if it would lead to termination of ongoing research or prevent future research.[11] Gafni and Birch[14] also highlight the need to consider what an intervention that has been subjected to a CED scheme is displacing before any assessment of potential cost savings through such schemes can be made.

There is also an ongoing challenge of disinvesting in technologies that have previously been approved.[18] Withdrawing coverage is logistically and politically difficult and it is considered more difficult to reverse a 'yes' than a 'no'.[24] Although no clear consensus has emerged on how these costs could be factored into the decision-making process, it has been suggested that these could be based on formal options analysis.[29]

Changing prices

Although discounting list prices can be thought of as an example of a risk-sharing agreement,[23] depending on how the OIR/AWR system operates, it may also lead the manufacturer to reconsider the pricing of the technology. Allowing prices to change as part of an OIR/AWR scheme also further extends the options available to decision-makers. Evidence generated as part of an OIR/AWR clinical study may also lead to a change in price if NICE believes that there is significant new evidence that will affect a drug's value. Something similar was observed with the multiple sclerosis risk-sharing scheme. Depending on the results observed, potential adjustments to the price of the drugs will be made at intervals to achieve an agreed cost per quality-adjusted life-year (QALY) of no more than £36,000.[23] The wider coverage associated with the multiple sclerosis risk-sharing scheme meant that it was necessary to have an upfront agreement on price changes following provision of evidence. It is not clear, however, to what extent changing prices will reduce uncertainty regarding the coverage decision.

Ethical issues

The potential ethical issues arising from the use of OIR/AWR schemes is another important theme emerging from the existing literature. For OIR, the issue of compulsory participation is often raised as a concern. Also, because of practical arrangements under OIR, treatments may not be available in all areas, causing geographical inequalities.[34] If a RCT is commissioned following an OIR recommendation, this raises a greater issue in terms of participation than a simple registry. These access issues in relation to an OIR policy linked to a clinical trial can be somewhat remedied by a large-scale geographically diverse trial with broad inclusion criteria.[22]

In addition, it has been argued that denying access to a treatment demonstrated to be effective (however uncertain) is unethical. Patient advocacy groups may also be unwilling to accept this policy especially if the treatment is considered to be safe and efficacious.[24] These issues have important implications for both the design and the successful conduct of research and hence are considered in more detail in later sections.

Summary

The critical review identified a number of important themes and principles in relation to the use of OIR/AWR schemes. However, much of the existing literature is relatively discursive and there is a need to provide a set of principles and to establish an analytical framework to help guide and develop appropriate criteria for the use of OIR/AWR schemes by NICE.

Chapter 3

What assessments are needed?

Since an important objective of the NHS is to improve health outcomes across the population it serves, a technology can be regarded as valuable if its approval is expected to increase overall population health. The resources available to the NHS must be regarded as fixed (certainly by NICE) and so it is not sufficient to establish that a technology is more effective (the health benefits compensate for any potential harms) than the alternative interventions available, because approving a more costly technology will displace other health-care activities that would have otherwise generated improvements in health for other patients.[62] Therefore, even if a technology is expected to be more effective, the health gained must be compared with the health expected to be forgone elsewhere as a consequence of additional NHS costs, that is, a cost-effective technology will offer positive net health effects (NHEs).[63–65] A social objective of health improvement and an ethical principle that all health impacts are of equal significance, whether they accrue to those who might benefit from the technology or other NHS patients, is an established starting point for the NICE appraisal process (see *Social value judgements and ethical principles*).[1]

An assessment of expected cost-effectiveness or NHEs relies on evidence about effectiveness, impact on long-term overall health and potential harms, as well as the costs that fall on the NHS budget together with some assessment of what health is likely to be forgone as a consequence (the cost-effectiveness threshold).[66] Such assessments are inevitably uncertain and, without sufficient and good-quality evidence, subsequent decisions about the use of technologies will also be uncertain and there will be a chance that the resources committed by the approval of a new technology may be wasted if the expected positive NHEs are not realised. Equally, rejecting a new technology will risk failing to provide access to a valuable intervention if the NHEs prove to be greater than expected. Therefore, if the social objective is to improve overall health for both current and future patients then the need for and value of additional evidence is an important consideration when making decisions about the use of technologies.[67–69]

This is even more critical once it is recognised that the approval of a technology for widespread use might reduce the prospects of conducting the type of research that would provide the evidence needed.[70] In these circumstances there will be a trade-off between the NHEs for current patients from early access to a cost-effective technology and the health benefits for future patients from withholding approval until valuable research has been conducted. A key ethical question arising from this trade-off is whether or not the health impacts for future patients should be considered and regarded as of similar significance to impacts on current patients (see *Social value judgements and ethical principles*).[24]

Because publically funded research also consumes valuable resources that could have been devoted to patient care, or other more valuable research priorities, there are a number of trade-offs that must be made. In making these trade-offs consideration also needs to be given to uncertain events in the near or distant future, which may change the value of the technology and the need for evidence.[71] In addition, implementing a decision to approve a new technology is, in general, not a costless activity and may commit resources that cannot subsequently be recovered if the guidance changes in the future.[6,7,72] For example, there may be costs associated with implementing guidance or training health-care professionals, or other investment costs associated with equipment and facilities.[73,74] The irrecoverable nature of these costs can have

particular influence on a decision to approve a technology if new research is likely to report or other events may occur in the future (e.g. the launch of new technologies or changes in the price of existing technologies).

The primary purpose of this chapter is to provide a non-technical exposition of the conceptual framework, developed more formally in *Appendix 6*, which identifies the key principles and assessments that are needed when considering both approval and research decisions. The first section outlines the key principles and the different types of assessment needed and how each sequence might lead to different categories of guidance. The next section, *Changes in prices and evidence*, examines how guidance might change if there are changes in the effective price of the technology or evidence. The following section, *Social value judgements and ethical principles*, highlights the social values and ethical principles associated with OIR and AWR.[a] Importantly, within this chapter we do not presuppose how the assessments ought to be made because there are a range of different types of additional information, evidence and methods of analysis that might be useful. These alternatives are examined in *Chapter 5* where they are more fully explored and evaluated through four case studies.

Key principles and assessments needed

The key principles and assessments fall into four broad areas:

1. expected cost-effectiveness and population NHEs (including benefits, harms and NHS costs)
2. the need for evidence and whether or not the type of research required can be conducted once a technology is approved for widespread use
3. whether or not there are sources of uncertainty that cannot be resolved by research but only over time
4. whether or not there are significant (opportunity) costs that will be committed and cannot be recovered once the technology is approved.

Guidance will depend on the combined effect of all of these assessments because they influence whether the benefits of research are likely to exceed the costs and whether any benefits of early approval are greater than withholding approval until additional research is conducted or other sources of uncertainty are resolved.

This can be complex because these different considerations interact. For example, the effect of irrecoverable costs will depend on the need for additional research and will also influence whether research is worthwhile. The sequence of assessments, decisions and resulting guidance can be represented by a flow chart or algorithm. Although such a representation is an inevitable simplification of the necessary trade-offs it helps to (1) identify how different guidance might be arrived at, (2) indicate the order in which assessments might be made, (3) identify how similar guidance might be arrived at through different combinations of considerations and (4) identify how guidance might change (e.g. following a reduction in price) and when it might be reviewed and decisions reconsidered. The complete algorithm is complex (reported in *Appendix 2* for completeness), representing the sequences of assessments and associated decisions, each leading to a particular category and type of guidance. However, the key decision points in the algorithm, reflecting the main assessments and judgements required during appraisal, can be represented as a simple 7-point checklist (see *Chapter 5*, *A checklist of assessment*).

Four broad categories of guidance are represented within the algorithm and include 'approve', 'AWR', 'OIR' and 'reject'. Each of the categories is further subdivided and numbered to indicate the different types of apparently similar guidance that could arise from different

considerations. 'Delay' is not considered a particularly useful category because NICE always has the opportunity to revise its guidance, that is, a decision to reject can always be revised but it is only with hindsight that reject might appear to be delayed approval. The distinction made between assessment and decision reflects the NICE appraisal process: first, critically evaluate the information, evidence and analysis (an assessment), which can then assist the judgements (decisions) that are required in appraisal when formulating guidance.

Technologies without significant irrecoverable costs

Some element of cost that once committed by approval cannot be subsequently recovered is almost always present. However, the significance of these types of costs depends on their scale relative to expected population NHEs associated with the technology as well as the nature of subsequent events (see *Technologies with significant irrecoverable costs* and *Chapter 5, Point 2: Are there significant irrecoverable costs?*).[75] In this section we consider the relatively simple sequence of assessments and decisions which lead to guidance for those technologies that are not judged to have 'significant' irrecoverable costs associated with approval.

Technologies expected to be cost-effective

The sequence of assessments and decisions, which ultimately leads to guidance, starts with cost-effectiveness and the expected impact on population NHEs, that is, where existing NICE appraisal currently ends. This is an assessment of expected cost-effectiveness, that is, 'on average', based on the balance of the evidence and analyses currently available. It will include an assessment of effectiveness and potential for harms as well as NHS costs (see NICE's *Guide to Methods of Technology Appraisal*[1]). Any assessment may be very uncertain with the scale and consequences of uncertainty assessed subsequently when considering the need for additional evidence. The sequence of assessments and decisions (judgements required) is illustrated in *Figure 1*. This demonstrates that an assessment of cost-effectiveness is only a first step and does not itself inevitably lead to a particular category of guidance. For example, a technology that might on balance be expected to be cost-effective might nevertheless receive OIR guidance if the additional evidence that is needed cannot be acquired if the technology is approved.

Need for evidence

Some initial assessment of the need for further evidence and a decision about whether or not further research might be potentially worthwhile is important because a 'no' at this point can avoid further and complex assessments, for example a technology offering substantial and well-evidenced health benefits at modest additional cost is likely to exhibit little uncertainty about whether or not the expected population NHEs are positive. In these circumstances, further research may not even be potentially worthwhile (i.e. the opportunity costs of conducting this research exceed its potential value) and so guidance to approve could be issued on the basis of existing evidence and at the current price of the technology (e.g. 'Approve[4]' in *Figure 1*). If additional evidence is needed and further research might be worthwhile, then further assessments and decisions are required before guidance can be issued. Critically, some assessment is required of the type of evidence that is needed and whether or not the type of research required to provide it is likely to be conducted if approval is granted.[29]

Research is possible with approval

If research is possible with approval, some further assessment of the long-term benefits of research is required, including (1) the likelihood that the type of research needed will be commissioned by research funders or conducted by manufacturers, (2) how long until such research will be commissioned, recruit and successfully report and (3) how much of the uncertainty might be resolved by the type of research that is likely to be undertaken.[70] An assessment of other sources of uncertainty that will resolve only over time is also needed, for example changes in prices or the launch of new technologies.[71] These sources of uncertainty
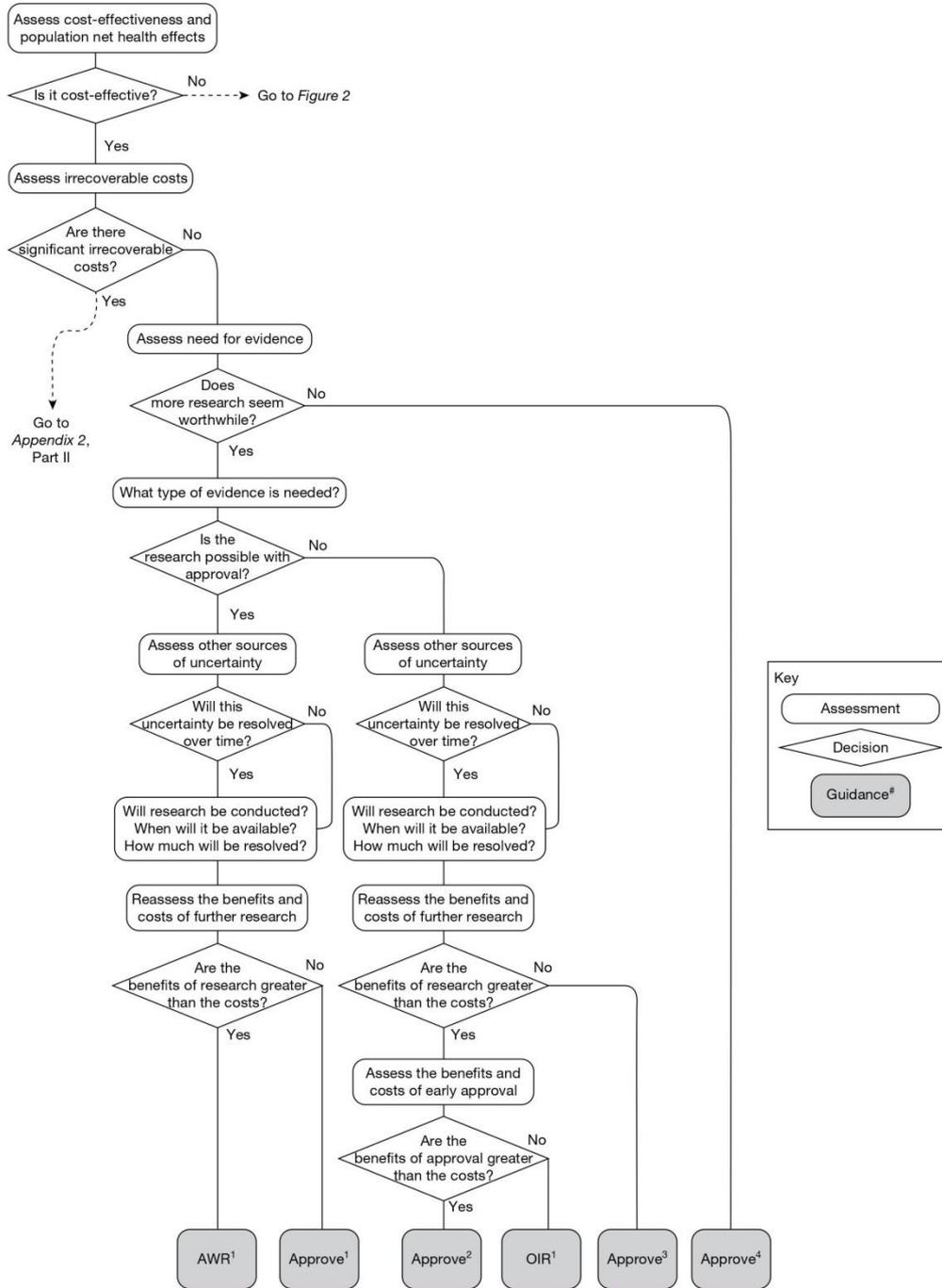
FIGURE 1 Algorithm for technologies expected to be cost-effective.

will influence the future benefits of research that could be undertaken as part of AWR. For example, even if the current benefits of research, which might be very likely to be undertaken, are considerable, if the price of the technology is likely to fall significantly before or shortly after the research reports, or if future innovation makes the current technology obsolete (or more effective), then the future benefits, once the research reports, might be very limited. In these circumstances it might be better to approve (rather than use a policy of AWR) and reconsider whether and what type of research is needed at a later date once these uncertainties have resolved. The judgement of whether the long-term benefits of research are likely to exceed its expected costs determines guidance, with 'AWR[1]' and 'Approve[1]' in *Figure 1* dependent on 'yes' and 'no', respectively.

Research is not possible with approval

The type of research needed, for example RCTs, may not be possible once a technology is approved for widespread NHS use (because of ethical concerns, recruitment problems and limited incentives for manufacturers). In these circumstances the expected benefits of approval for current patients must be balanced against the benefits to future patients of withholding approval to allow the research to be conducted. Initially, the same assessment of the long-term value of the type of research that might be conducted if approval is withheld is still required. Similarly, the impact of other sources of uncertainty on the longer-term benefits of research is also needed. If the benefits of research are judged to be less than the costs (i.e. research is not worthwhile anyway), the technology can be approved based on current evidence and prices ('Approve[3]' in *Figure 1*). However, judging that research is worthwhile at this point is not sufficient for OIR guidance. In addition, an assessment of whether the benefits of early approval (expected population net benefits for current patients) are greater than the opportunity costs (the net benefit of the evidence likely to be forgone for future patients as a consequence of approval) is required. If the expected benefits of early approval are judged to be less than the opportunity costs then OIR guidance would be appropriate ('OIR[1]' in *Figure 1*). Alternatively, if the expected benefits of early access for current patients are judged to be greater than the opportunity costs for future patients then approval would be appropriate ('Approve[2]' in *Figure 1*). All of these assessments, including the benefits of early approval and the value of evidence, will change if the effective price of the technology is reduced (see *Changes in effective prices*).

Technologies not expected to be cost-effective

A technology that is not expected to be cost-effective will, on balance, impose negative population NHEs if it is approved. These negative NHEs can arise because the technology may not be effective, the potential for harm exceeds any benefits and/or the additional NHS costs are not justified by the magnitude of the expected health benefits offered. In these circumstances approval can be ruled out, but which of the other categories of guidance might be appropriate will depend on subsequent assessments and decisions (*Figure 2*).

Need for evidence

Any assessment will be uncertain, so it remains possible that a technology that is not expected to be cost-effective on the balance of existing evidence might offer positive NHEs. Therefore, the scale and consequences of this uncertainty must be considered to make an initial assessment of the need for additional evidence and whether additional research might, in principle, be worthwhile. If it is not then the technology can be rejected based on existing evidence and its current price ('Reject[4]' in *Figure 2*). Alternatively, if further research might be worthwhile then an additional assessment is required of whether the type of evidence and research that is needed can be conducted without approval.
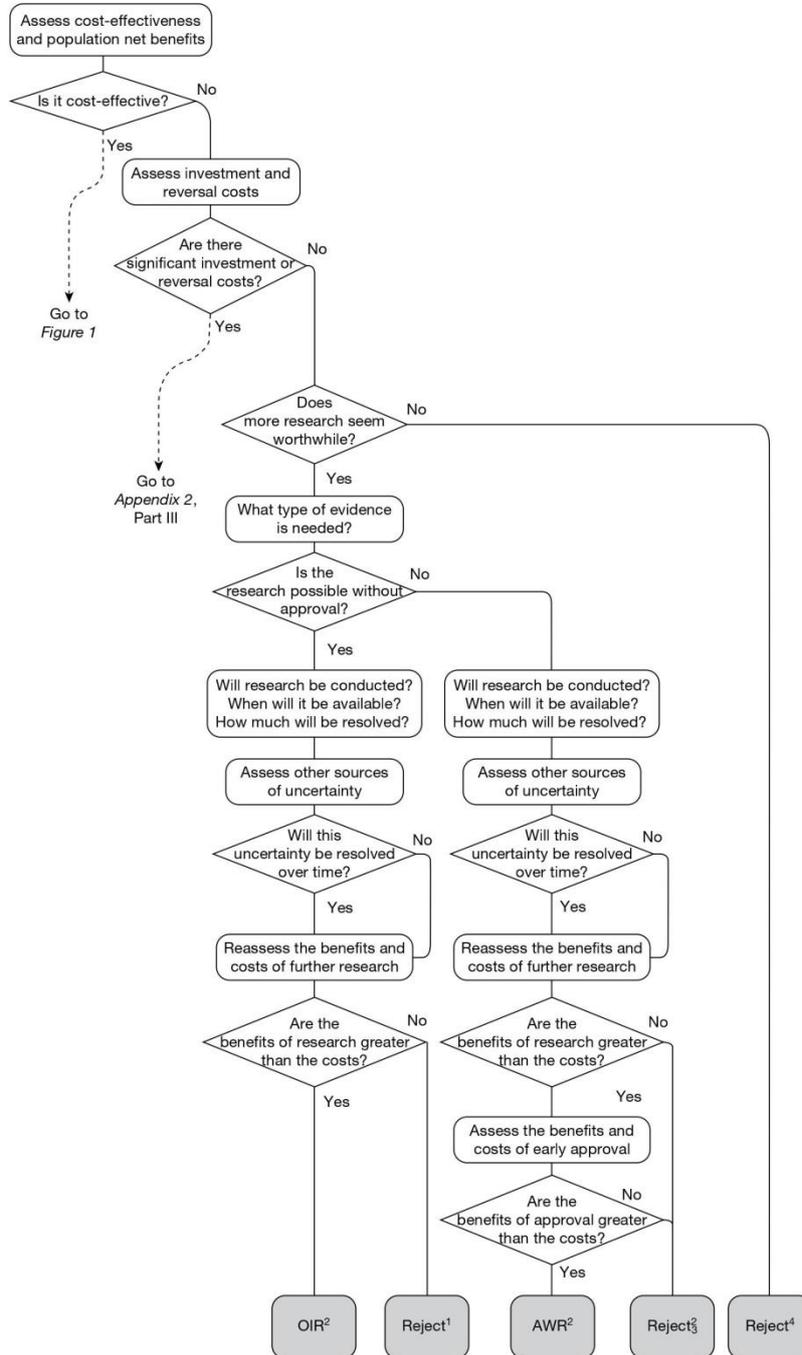
FIGURE 2 Algorithm for technologies not expected to be cost-effective.

Research is possible without approval

Generally, most types of research (including RCTs) will be possible without approval. Further assessment of the longer-term benefits of the type of research that is likely to be conducted, and when it might report, is required, including the impact of other sources of uncertainty that will resolve over time. If, following this reassessment, the expected benefits of research are judged to exceed the associated costs then OIR would be appropriate ('OIR² in *Figure 2*). Alternatively, if the costs of research are likely to exceed the longer-term expected benefits then the technology should be rejected at this point ('Reject¹' in *Figure 2*).

Research is not possible without approval

In some circumstances it is possible that certain types of evidence might be acquired only, or be more easily acquired (more quickly and at lower cost), once a technology is in widespread use, for example linking surrogates (specific to the technology) to longer-term health outcomes, longer-term and/or rare adverse events, greater understanding of learning and incremental improvements in the use of a technology, or identifying the particular types of patients that might benefit most.[76] In this less common situation, in which the type of research needed is not possible (or is significantly more costly) without approval, the same assessment of the longer-term benefits of research is required. If further research is judged not to be worthwhile following this reassessment, the technology can be rejected ('Reject²' in *Figure 2*). Alternatively, if research is judged worthwhile, an additional assessment of whether the benefits of approval exceed the costs is required. In this case, approval of a cost-ineffective technology would make the research possible, but will impose opportunity costs (negative expected population NHEs). The key question is whether the net benefits of the research exceed these opportunity costs. If they do not then the technology should be rejected even though research, had it been possible without approval, would have been worthwhile ('Reject³' in *Figure 2*). Alternatively, if the net benefits of research more than offset the opportunity costs then AWR would be appropriate even though the technology is expected to be cost-ineffective ('AWR²' in *Figure 2*).

Therefore, AWR guidance for technologies not expected to be cost-effective is certainly possible but is appropriate only in certain circumstances: (1) the type of research needed is not possible without approval, (2) the long-term benefits of the research are likely to exceed the expected costs and (3) the additional net benefits of such research exceed the opportunity costs of approving a cost-ineffective technology. More commonly, research might be possible but more costly without approval. In these circumstances, AWR guidance could be considered only if the additional costs of research without approval exceed the opportunity costs of approving a cost-ineffective technology.

Technologies with significant irrecoverable costs

Irrecoverable costs are those that, once committed, cannot be recovered should guidance be revised at a later date. In most NICE appraisals these are included in the expected (per patient) cost of a technology. However, rarely is their potential additional impact explored when future events, such as research reporting or other sources of uncertainty resolving, might mean that guidance will be revised in the near or distant future.[6,72] These types of cost are commonly thought of as capital expenditure on equipment or facilities which have a life expectancy that extends beyond the current patient population. They might also include the resources required to implement guidance or to train staff to use a new health technology or a period of 'learning' during which NHEs are lower. Although these costs are incurred up front, they tend to be included in NICE assessments as if they are paid per patient treated over the lifetime of the equipment or facility. This common assumption will have no effect as long as guidance is certain not to change during this period. However, if it is possible that initial approval might be withdrawn at some point, then, although future patients will no longer use the technology, these upfront costs cannot be recovered (see *Issues specific to 'only in research' recommendations*).

Therefore, in these circumstances it would be inappropriate to include these costs as if they were paid per patient treated (see *Point 2: Are there significant irrecoverable costs?*). The possibility that approve or AWR might be reconsidered after research reports, for example, and the impact that this would have on expected costs need to be considered before committing these types of capital costs, that is, it may be better to withhold approval and avoid commitment of resources until the uncertainty is resolved.

However, irrecoverable costs may be much more common. Even in the absence of capital investment in equipment and facilities, most new technologies offer a 'risky investment profile' for each patient treated. Generally they impose initial per patient treatment costs that exceed the immediate health benefits (see *Point 1: Is it expected to be cost-effective?*). These irrecoverable treatment costs are offset only by cost savings and health benefits in the longer run, that is, initially negative NHEs (losses) are only gradually compensated by later positive ones (gains). Therefore, a technology expected to be cost-effective may be expected to break-even, that is, when accumulated 'gains' compensate earlier 'losses', after some considerable time. If guidance is likely to change it is possible that initial losses will not be compensated by later gains and the expected additional NHEs will not be realised.[75] This type of investment profile becomes significant (has some influence on a decision to approve) if the decision to treat a presenting patient can be delayed until uncertainty is resolved (e.g. research reports or other events occur) because the commitment of irrecoverable opportunity costs (negative NHEs) can be avoided (see *Point 2: Are there significant irrecoverable costs?*). In these circumstances, OIR or reject avoids this commitment and preserves the option to approve the technology at a later date when its purchase by the NHS represents a 'less risky investment'.[75,b]

Although aspects of irrecoverable cost are almost always present, their potential significance also depends on their scale relative to expected population NHEs of the technology. Critically, their impact depends on the chance that guidance will be revised in the near or distant future as a result of new evidence becoming available or changes in prices and technologies. The full algorithm becomes more complex (see *Figures 27* and *28* in *Appendix 2*) so here we focus on the key differences from the section on technologies without significant irrecoverable costs.

Technologies expected to be cost-effective

The presence of irrecoverable costs associated with a technology that is expected to be cost-effective will influence guidance and be regarded as 'significant' only if there are future events (research reporting or other sources of uncertainty resolving) that might change guidance. For example, if research is possible with approval and is expected to be worthwhile, AWR does not necessarily follow as previously (e.g. see 'AWR[1]' in *Figure 1*) because the impact of irrecoverable costs must also be considered. Now OIR may be more appropriate than AWR (e.g. the choice between 'OIR[4]' or 'AWR[4]' in *Figure 27*), even though the research would be possible with approval, because OIR avoids the commitment of irrecoverable costs until the results of research are known. This is especially so when there are also other sources of uncertainty that might resolve while the research is being conducted because they increase the chance that guidance will be revised (e.g. 'OIR[3]' or 'AWR[3]' in *Figure 27*).

If research is not possible with approval but is expected to be worthwhile, then OIR will be appropriate if the opportunity costs of early approval are judged to exceed the benefits (e.g. 'OIR[6]' rather than 'Approve[9]' in *Figure 27*). These opportunity costs will now also include the impact of irrecoverable costs when guidance might be changed as well as the value of evidence that will be forgone by early approval. Therefore, irrecoverable costs will tend to make OIR rather than approval more likely, particularly when there are other sources of uncertainty that might resolve while the research is being conducted (e.g. 'OIR[5]' rather than 'Approve[7]' in *Figure 27*).

If research is not judged worthwhile, approval does not necessarily follow as previously (e.g. 'Approve[1,3,4]' in *Figure 26*). Now the technology should be approved only if there are no other sources of uncertainty. If there are other sources of uncertainty, then an assessment of the benefits and costs of early approval is needed that takes account of irrecoverable costs and the risk that guidance might change in the future. Therefore, reject rather than approval is possible, even though a technology is expected to be cost-effective, because the decision to commit the irrecoverable costs can be reconsidered once the other sources of uncertainty have resolved (e.g. 'Reject[5,6]' in *Figure 27*).

Technologies not expected to be cost-effective

The presence of irrecoverable costs for technologies not expected to be cost-effective does not change the categories of guidance, or how they might be arrived at. However, it does mean that reject is more likely to be appropriate than AWR when research is not possible without approval (see 'AWR[5]' in *Figure 28*). This is because a decision to reject, although it may be revised to approve, generally does not commit irrecoverable costs. Although there may be resources associated with making sure subsequent approval is properly implemented, these costs are properly considered as an irrecoverable cost associated with approval (rather than a reversal cost of reject). There may be circumstances when implementing guidance to reject a technology also requires resources if it has already diffused into clinical practice. If these are significant they should be taken into account in the same way as other irrecoverable costs, tending to make AWR more likely to be appropriate.

Different types of guidance

Each sequence of assessment and decision leads to different categories and 'types' of guidance for technologies with differing characteristics, indications and target populations. The different 'types' of guidance illustrate how similar guidance might be arrived at in different ways, helping to identify the particular combinations of considerations that might underpin guidance, contributing to the transparency of the appraisal process. The possible categories and types of guidance are summarised in *Table 1*; the numbers in the body of the table refer to the numbered guidance in *Figures 1* and *2* and the algorithm in *Appendix 2* (see *Figures 26–28*).

The categories of guidance available to NICE have wider application than is reflected in previous guidance (see *Chapter 4*). For example, there are five different types of OIR that may be appropriate when a technology is expected to be cost-effective. Indeed, OIR may be appropriate even when research is possible with approval if there are significant irrecoverable costs. AWR can be considered only when research is possible with approval, but reject remains a possibility even for a cost-effective technology if there are irrecoverable costs. Therefore, the full range of categories of guidance (OIR and reject as well as AWR and approve) ought to be considered for technologies that, on the balance of existing evidence and current prices, are expected to be cost-effective.

It is only approval that can be ruled out if a technology is not expected to be cost-effective, that is, cost-effectiveness is necessary but not sufficient for approval but lack of cost-effectiveness is neither necessary nor sufficient for rejection. Although likely to be uncommon, there are circumstances when AWR may be appropriate even when a technology is not expected to be cost-effective (see *Technologies without significant irrecoverable costs* and *Technologies expected to be cost-effective*). More commonly the choice of appropriate guidance will be either reject or OIR. Importantly, which category of guidance will be appropriate depends only partly on an assessment of expected cost-effectiveness and hence this assessment should be regarded only as an initial step in formulating guidance. Guidance will depend on a number of other key assessments including (1) the need for evidence, (2) whether or not the type of research required

TABLE 1 Different types of guidance

Type of guidance	No significant irrecoverable costs					Significant irrecoverable costs				
	Research not needed	Research possible with approval		Research not possible with approval		Research not needed	Research possible with approval		Research not possible with approval	
		Benefits >costs	Benefits <costs	Benefits >costs	Benefits <costs		Benefits >costs	Benefits <costs	Benefits >costs	Benefits <costs
Technologies expected to be cost-effective										
Approve (12)	4		1	2	3	11, 12		5, 6	7, 9	8, 10
AWR (3)		1					3, 4			
OIR (5)				1			3, 4		5, 6	
Reject (3)						7		5		6
Technologies not expected to be cost-effective										
Approve (0)										
AWR (2)				2					5	
OIR (2)		2					7			
Reject (8)	4		1	2	3	11		8	9	10

The numbers in the body of the table refer to the numbered guidance in *Figures 1* and *2* and the algorithm in *Appendix 2*.

is possible with approval, (3) the expected longer-term benefits and costs of the type of research likely to be conducted, (4) the impact of other sources of uncertainty that will resolve over time and (5) the significance of any irrecoverable costs.

Changes in prices and evidence

The type of guidance that might be appropriate will be influenced by changes in the effective price of the technology, the type of evidence available to support its use and whether or not further research is likely to be undertaken, either by manufacturers or research commissioners, as a result of OIR or AWR guidance.

Changes in effective prices

Any change in the effective price of a technology, either through patient access schemes (which offer some form of discount that reduces NHS costs) or direct price changes (possibly negotiated though a future value-based pricing scheme), will affect key assessments and decisions, leading to different 'paths' through the algorithm, consequently changing the category of guidance that would be appropriate (Simon Walker, Centre for Health Economics, University of York, UK, July 2011, personal communication).[4,23] For example, provisional OIR guidance for a technology that is expected to be cost-effective might be revised to approve with a sufficient price reduction because the benefits of early approval will be greater and uncertainty about its cost-effectiveness and therefore the value of additional evidence will tend to be lower (e.g. from 'OIR[1]' to 'Approve[2]' in *Figure 1*).[c] Similarly, AWR might be revised to approve if the benefits of early approval now exceed the value of additional evidence (e.g. from 'AWR[1]' to 'Approve[2]' in *Figure 1*).[d]

Equally, provisional guidance to reject a technology that is not expected to be cost-effective might be revised to OIR if the reduction in price is not sufficient to make it cost-effective but makes the costs associated with a reject decision more uncertain and hence the value of research worthwhile (e.g. from 'Reject[1]' to 'OIR[2]' in *Figure 2*).[e] If the reduction in price is greater and sufficient to make the technology cost-effective, then guidance might be revised to AWR if research remains worthwhile and possible with approval (e.g. from 'Reject[1]' or 'OIR[2]' in *Figure 2* to 'AWR[1]' in *Figure 1*). Clearly, with an even greater reduction in price, it is possible that provisional guidance to reject could be altered to early approval (e.g. 'Approve[1]' in *Figure 1*). Even if research is not possible with approval, a sufficient reduction in price could also lead to early approval (e.g. from 'Reject[1]' or 'OIR[2]' in *Figure 2* to 'Approve[2,3,4]' in *Figure 1*).[f]

Therefore, consideration of the effect of price changes on OIR and AWR is needed when assessing the potential impact of patient access schemes and more direct price negotiation through value-based pricing.[65,77–79] It should be noted that, all other things being equal, the presence of significant irrecoverable costs will require greater reductions in effective price to achieve the same revision to a more permissive category of guidance.

Threshold prices and value-based prices

The price at which the technology would just be expected to be cost-effective is commonly regarded as the value-based price for the technology, that is, the maximum price that the NHS can afford to pay without imposing negative health effects.[65,80] This single price describes the threshold for approve/reject decisions and would be the relevant threshold price when (1) OIR or AWR guidance is not available to the decision-maker or there is no uncertainty in cost-effectiveness or (2) the research, if needed, can be conducted with approval and (3) there are no irrecoverable costs. In all other circumstances there are a number of other threshold (value-based) prices. The number and value of these thresholds depends on the characteristics of

the technology (the path through the algorithm); however, the threshold prices for approval will always be lower than the single approve/reject price based on expected cost-effectiveness.

For example, for a technology (without significant irrecoverable costs) for which research could be conducted without approval but not with it, there are two threshold prices: (1) the threshold that would move guidance from reject to OIR and (2) the threshold that would move guidance from OIR to approve. The latter will always be lower than the price that would move the same technology from reject to approve if OIR was excluded from consideration. If a technology also imposes significant irrecoverable costs then there may be more threshold prices. For example, when research can be conducted with or without approval there are three thresholds: (1) reject to OIR, (2) OIR to AWR and (3) AWR to approve. Again, (3) will be lower than the approve/reject threshold for the same technology if AWR was excluded from consideration. All other things being equal the presence of irrecoverable costs will tend to reduce the threshold price for approval.

Even in circumstances in which price negotiation becomes possible alongside NICE appraisal, it will be important to retain the OIR and AWR as available categories of guidance for two reasons. First, there is no guarantee that manufacturers will always agree to the lower price threshold that would lead to approval rather than OIR or AWR. Second, and possibly more importantly, there may be many circumstances when there is no effective price reduction that would make approval appropriate.[g] For example, reject or OIR guidance may still be appropriate even if the effective price of a technology is zero if there is substantial uncertainty about its effectiveness and/or potential for harms.

Incentives for evaluative research

These threshold prices represent the maximum effective price at launch to achieve a particular category of guidance when the results of any subsequent research that might be undertaken are not yet known. This is different to the type of flexible pricing agreements described in the current Pharmaceutical Price Regulation Scheme (PPRS) in which price is revised once the research reports and the results are known, with prices increasing if the evidence suggests that benefits were originally underestimated and reducing if they were overestimated.[81] This means that manufacturers retain an incentive to conduct further evaluative research if they believe that there are additional benefits that could not be evidenced at launch. Publically funded evaluative research, however, will still be required when these incentives are insufficient and especially in those cases in which the original evidence is likely to have overestimated the benefits or underestimated the potential for harm. However, it should be noted that linking effective prices to the results of publically funded research means that the NHS will benefit (realise the value of evidence) only if the results lead to a lower price or more restrictive guidance because the technology is found not to be cost-effective (thus avoiding the losses associated with negative NHEs). Manufacturers will, however, be able to appropriate the value of evidence if research results suggest that NHEs were originally underestimated because prices will increase within the flexible pricing scheme (or a value-based pricing scheme that might replace it). Even under current arrangements this value can be appropriated when the technology is reappraised by NICE (e.g. any patient access scheme could be withdrawn or less restrictive positive guidance issued). Consideration of how the NHS and manufacturers are likely to share the value of evidence might inform whether manufacturers should be expected to conduct the research specified in AWR or OIR guidance. As long as incentive-consistent contractual arrangements can be set in place, that is, those that can be monitored and enforced with credible penalties, it should ensure that any agreed research is conducted in the way intended. Alternatively, manufacturers might be expected to make some contribution to the costs of publically funded research that may ultimately benefit their product (see *Chapter 5, Point 7: Are the benefits of approval greater than the costs?*).

It is important that policy provides (or at least does not undermine) appropriate incentives for manufacturers to conduct the type of research needed to support NICE guidance at launch. The use of OIR and AWR guidance, as described in the algorithm, could provide clear signals and incentives. For example, the threshold price for reject/OIR and reject/AWR will be higher than that for OIR/approve and AWR/approve. Guidance restricted to OIR also offers very limited NHS volumes and revenue to manufacturers. This provides a strong incentive to ensure that the type of evidence which would require research that cannot be conducted once a technology is approved for NHS use is available and is sufficient at launch (e.g. relative effectiveness and subtle but important differences in side-effect profiles). Therefore, a predictable OIR and AWR policy signals what type of evidence is likely to be most important at an early stage.

The use of OIR and AWR, as described in the algorithm, offers manufacturers a choice to (1) accept OIR and AWR guidance at a higher price, (2) reduce the effective price to achieve approval, when that is possible or (3) conduct the evaluative research at an earlier stage so that cost-effectiveness is not uncertain at launch. Other things being equal, those new technologies that are supported during the NICE appraisal process by more, better-quality and relevant evidence will be more likely to be approved (rather than receiving OIR or AWR guidance) and at higher prices than those that are not, because additional evidence is less likely to be needed. Therefore, greater consideration of OIR and AWR will tend to reward those manufacturers who have invested in good-quality and relevant evidence with earlier approval of their technology. In addition, the effect of price on OIR and AWR recommendations suggests that those technologies supported by better evidence will tend to get approval at higher effective prices, providing an incentive for manufacturers to invest in the type of evidence needed earlier in the development process.

Assessing the prospects of research

When considering OIR or AWR guidance there must be some assessment of (1) the type of research needed to address the key uncertainties, (2) whether or not this will be regarded as ethical and can be undertaken while the technology is approved for use, (3) whether or not it is likely to be a priority for public funding and be commissioned and (4) when it is likely to report.

Although the NICE appraisal process may be well suited to identifying the need for evidence when assessing cost-effectiveness, these other critical assessments are not necessarily ones for which NICE and its Advisory Committees, as currently constituted, have particular expertise, not least because they reflect the decisions of those responsible for research design, prioritisation and commissioning.[3,82] Without sufficient co-ordination between these communities there is a danger that OIR or AWR could be issued when the type of research required would not be regarded as either ethical or feasible or of sufficient priority compared with other competing research needs to be commissioned. Because publically funded research is also budget constrained, it is perfectly possible that research that might be valuable from a wider NHS perspective might nevertheless not be a priority if other more valuable research might be displaced. This might be a particular concern if there is a possibility that the research could be undertaken by the manufacturer rather than displacing other research without proprietary interest. Therefore, a decision of whether OIR or AWR research should be undertaken by the manufacturer or through publically funded research is one that NICE cannot properly take alone.

Although some judgement about how the research community might respond to OIR or AWR recommendations when NICE is formulating guidance is clearly possible, more informed judgements and better decisions might be possible through greater involvement of the research community. For example, a Research Advisory Committee could be constituted that could consider provisional OIR or AWR guidance, making recommendations during the consultation period about the type of research needed and its ethics, feasibility and likely priority before final

appraisal and guidance. It might also make recommendations about whether research should be publically funded or undertaken by the manufacturer with appropriate contractual arrangements. There are of course many different ways in which greater co-ordination might be achieved. However, because some of the assessments that NICE must make in formulating OIR or AWR guidance are, in fact, research decisions that fall outside its remit, it would seem sensible to draw on the expertise of those involved in, and responsible for, these types of research decisions to help make these assessments.

Social value judgements and ethical principles[h]

This section clarifies and discusses ethical issues arising from OIR and AWR decisions. There are some important preliminary points. First, the question considered here is whether or not OIR and AWR recommendations are consistent with the values and principles that currently underpin standard NICE practices.[1] It is not in the remit of this report to evaluate those underpinning values and principles themselves. In particular, it is assumed here that the health budget is necessarily limited; also that, generally speaking – and although also taking into account issues of need and equity as discussed in the NICE values statements[83] – scarce health-care resources ought to be broadly allocated so as to maximise health outcomes of the population as a whole and hence that treatments that benefit one group of patients will be funded at an opportunity cost to other patients. Given these assumptions, the emphasis in this discussion is on new ethical challenges created by OIR and AWR decisions[11] as distinct from issues shared with standard NICE recommendations.

This section is in four parts: first, ethical issues raised by both OIR and AWR are discussed; second, ethical issues raised specifically by OIR are addressed; third, ethical issues specific to AWR are considered; and, finally, a summary of the ethical analysis is provided. It is important to have a working definition of OIR and AWR. The defining characteristic of an OIR decision is that a necessary condition of a patient receiving the treatment in question is that he or she participates in the relevant research; typically, NICE recommends that the innovative treatment is available under the NHS only as part of a RCT (although other types of research study might be relevant). The defining characteristic of an AWR decision is that receipt of the new treatment by patients is not conditional on their participation in the relevant research programme; typically, NICE approves the treatment but stipulates that further evidence must be collected, for example on long-term outcomes and adverse events, often in patient registries. This distinction between OIR and AWR is not always clear-cut. For example, an AWR decision limiting coverage to a subpopulation – members of a geographical region, say – whose access to the treatment is explicitly conditional on their contributing to evidence development by submitting their health information to a registry is, on the above definitions, more like an OIR than an AWR decision. Nonetheless, the working definitions of OIR and AWR suggested here suffice for the purposes of this ethical analysis.

Issues common to 'only in research' and 'approval with research' recommendations

Generally speaking, the benefit of attaching research conditions to NICE recommendations is an improved evidence base for resource allocation decisions in the future. The beneficiaries of the research are members of future populations who will profit from better informed allocation decisions. But achieving this benefit can impose significant opportunity costs on current patients. This is true for some OIR decisions and some AWR decisions.

To be clear about when a research condition does impose an opportunity cost on the present population, and how significant this is, two issues need to be carefully considered. The first is

what is meant by present and future populations. For the purposes of this discussion the present population comprises people whose interests are directly affected by a NICE recommendation (e.g. they receive an innovative treatment approved by NICE, or benefit from resources made available because NICE rejects an innovative treatment). Future populations comprise people whose interests are indirectly affected by decisions, in particular by the subsequent research results that improve the evidence base for future NICE judgements. It is important to note that on these definitions some people – specifically patients with a chronic condition who live sufficiently long – will be members of both the present and the future populations. Thus, some individuals in the present population may benefit from the research condition because they will also be members of the future population. This will not be true of all, so the issue of balancing the interests of some individuals in the present population against some individuals in the future population remains.

The second issue is under what circumstances the present population is disadvantaged by the research condition compared with the alternative recommendation that NICE might make. This will depend both on what the alternative recommendation would be and on the level of current evidence about cost-effectiveness for the intervention. Three general types of situation might arise:

1. The evidence concerning the cost-effectiveness of a new treatment compared with standard treatment is genuinely balanced such that the likelihood of a new treatment turning out to be less cost-effective than standard treatment is the same as the likelihood of that new treatment turning out to be more cost-effective than standard treatment. Under these conditions it makes no difference to the present population whether a research-conditional recommendation is made or not. Indeed it makes no difference whether the new intervention is accepted or rejected. Future populations, however, will benefit from the research. Some current patients will be members of both present and future populations and will therefore benefit from research.
2. The balance of evidence as judged by NICE suggests that the new treatment is less cost-effective than standard treatment but NICE wishes strongly to encourage further research, hence it is considering a research-conditional recommendation. Assuming that NICE's judgements on this issue are generally better than chance then the present population is disadvantaged by a research-conditional recommendation compared with rejection but advantaged by a research-conditional recommendation compared with acceptance. Future populations are always advantaged by effective research.
3. The balance of evidence is that the new treatment is more cost-effective than standard treatment. In this situation the present population is disadvantaged by a research-conditional recommendation compared with acceptance but advantaged by a research-conditional recommendation compared with rejection. Again, future populations will be advantaged by effective research.

There is currently little empirical evidence about whether or not NICE's judgements about the relative cost-effectiveness of two interventions, in situations (2) and (3) above, are better than chance. If research-conditional recommendations are allowed then it will be important to collect data to examine whether or not further research, on average, confirms NICE's initial judgement about expected cost-effectiveness. There is some evidence which suggests that relevant authorities do accurately judge when evidence favours neither one intervention nor the other.[84] This provides some grounds for expecting that when NICE judges that the evidence on cost-effectiveness favours an intervention, such a judgement will be correct more often than it is false.

In summary, in judging whether or not it is right to make a research-conditional recommendation there are four key issues to consider:

1. What is the likely effect on the current population of a research-conditional recommendation compared with whichever would be the alternative recommendation?
2. What is the likely benefit to the future population from the research?
3. What proportion of individuals in the present population is also likely to be in the future population?
4. How should we weigh up the disadvantages to individuals in the present population in relation to the advantages to individuals in the future population (some of whom will be in the present population)?

The first three of these considerations are essentially empirical issues and NICE will have to make judgements in the setting of uncertainty. The fourth issue is an ethical one. Do the interests of members of the future population count? If so, how are they to be weighed against the interests of members of the present population? Which set of interests should prevail?

One way of addressing this question is to consider how radical a departure from current NICE values, principles and practices it would be to accord weight to the interests of future populations. Arguably, doing so would not be much of a departure at all. For one thing, taking the interests of future populations into account is consistent with fundamental NICE assumptions about how to make allocation decisions. Specifically, NICE currently considers how the requirements of maximising health gain, need, equity and so on are to be balanced against each other; in doing so, NICE takes the identifiability of patients who benefit from an intervention to be irrelevant. So, the fact that beneficiaries of putting a research condition on approval are unidentifiable because they will exist only in the future does not seem to add anything to considerations that NICE currently recognises and weighs. Similarly, some consideration of future populations is implicit in standard NICE judgements. Consider, for example, new treatments for which evidence is poor but, nonetheless, what evidence there is suggests that they are cost-effective. If NICE took only the interests of the present population into account then these should be funded on the basis of the poor but positive evidence. In fact, such innovations are often rejected to ensure that the evidence attains the standard that NICE requires. The beneficiaries of this decision are not in the present population; what NICE has in mind is future populations of patients who will benefit from allocation decisions made on better evidence.

It would seem, then, that decisions that take into account the interests of future populations are consistent with NICE's values, principles and practices. Nonetheless, there are some issues that require further thought. Although trading the interests of present and future populations coheres with the general NICE framework, quite how these sets of interests should be weighted will be important. Consider, for example, what might be called the 'bird in the hand' argument. Suppose that NICE is balancing the health benefit that will accrue from early approval against the value of further evidence. That the latter is considered greater than the former has to be offset against the fact that the former is more secure than the latter simply because one can be more confident about events in the near future (in this case that patients will enjoy health benefits from early approval) than about events in the further future (in this case that the research will report and improve the evidence base for allocation decisions). At this point, the discussion of the ethics of research-conditional judgements segues into practical considerations familiar from other sections of this report, such as the likelihood of research being conducted.

Issues specific to 'only in research' recommendations

To tease out the ethical issues raised by OIR it will be useful to construct an illustrative case. Suppose that NICE appraises a new treatment for which there is strong evidence of effectiveness – that is, the innovation is known to be clinically superior to existing treatments – but there is considerable uncertainty over its cost-effectiveness. Whether or not the new treatment would prove to be cost-effective depends not only on how expensive it is but also on how much health

benefit it would produce – that is, quantification of the health benefit – compared with standard treatment. NICE might consider an OIR judgement in these circumstances to establish more exactly the size of benefit that, in turn, is deemed necessary to establish cost-effectiveness. For example, NICE could approve the new treatment only in the context of a RCT comprising two trial arms, the innovative treatment arm and the standard treatment arm. Crucially, on this decision, patients outside the trial, and participants randomly allocated to the standard arm of the RCT, would be denied what is almost certainly the better treatment for their condition. This scenario creates a number of important ethical issues.

Equipoise

A criterion established in research ethics for the legitimacy of carrying out a RCT is that there is substantial uncertainty as to which of the treatments being compared – that is, an innovative treatment and a standard treatment – is the more effective. This is sometimes known as the principle of equipoise. The principle is meant to capture the intuition that no one – patient or participant – should knowingly be offered less than the best treatment for their condition. OIR decisions may be made when there is such substantial uncertainty. In such cases, the principle is respected and ethical review of the relevant research poses only issues already considered as standard by researchers and Research Ethics Committees (RECs). However, in the situation envisaged here, NICE is considering recommending OIR when the intervention in question is clearly superior to alternatives, but the degree of its superiority remains uncertain, and so its cost-effectiveness is uncertain. Evidently, in this scenario researchers are not in equipoise about the relative effectiveness of the two interventions. The substantial uncertainty relates to whether the more effective, but more expensive, treatment produces sufficient extra benefit compared with alternatives for it to be recommended by NICE, but it does not relate to whether it is more effective. Is it permissible to flout the principle of equipoise concerning effectiveness and give an OIR decision in such circumstances?

An intuitive response is that patients are harmed by an OIR decision that denies them the best-known treatment for their condition in the interests of research. Theoretical support for this intuition is provided by the well-known principle of maleficence: above all, do no harm. But harm-based objections to OIR are inconclusive for two reasons. First, the concept of harm is contested. The three main accounts define someone being harmed by contrast with, respectively, (1) their state before the harm was perpetrated, (2) the state they could have been in and (3) a minimum or baseline standard of well-being. When clinicians and patients are not sufficiently uncertain about the effectiveness of the treatments being compared, patients denied a better treatment by OIR are harmed according to the definition of harm based on (2) because they are put in a worse state than was possible. But the patients are not harmed according to the understanding of harm based on (1) and (3) because they will receive the standard NHS treatment (i.e. the same treatment as patients who do not take part in the research, or the same treatment as all NHS patients would have received had NICE rejected the treatment rather than approved it as OIR). So, harm-based objections to OIR are as inconclusive as the current debate on the definition of harm. Second, harm-based objections to OIR are question-begging, in the following way. Suppose that there was a consensus on the nature of harm and, further, on the fact that OIR without equipoise harms some patients. To conclude that this makes OIR impermissible is to assume that the harm in question outweighs the benefits of research. But this is precisely what is in dispute, namely, the relative values of the benefits of early approval and of further research. Given NICE values, principles and practices, it is perfectly feasible to conclude that the harm perpetrated by OIR is justified by the benefits to future patients of a better evidence base for allocation decisions.

There are more fruitful lines of thought about OIR without equipoise. First, the principle of equipoise itself is under considerable strain from pressures that have nothing to do with OIR

and AWR. For example, it has been argued that it is permissible to trial less than the best-known treatment for human immunodeficiency virus in developing countries unable to afford the most effective interventions.[85] It is generally accepted that, as a result of such pressures, the principle of equipoise needs to be refashioned. NICE can trade on this by asking, 'About what must researchers be uncertain?'. The traditional requirement is that researchers must be in equipoise about the best-known treatment; but, NICE might argue, the really salient uncertainty is not about effectiveness per se, but about the extent of effectiveness, and hence about the cost-effectiveness. Because, in the scenario envisaged here, the OIR recommendation is made precisely because of uncertainty about the extent of effectiveness, the refashioned principle of equipoise is respected. This argument is part of a larger research ethics question about what is required in terms of equipoise, so will not be addressed here. In the present context there is a more important practical consideration to emphasise. A RCT required by OIR would have to be reviewed by a REC. RECs are used to requiring traditional equipoise (i.e. substantial uncertainty about which is the better treatment). So, if RECs are to approve research of the kind considered here, in which researchers are not in equipoise as traditionally understood, they will have to be informed of, and agree with, the rationale for conducting these distinctive studies.

Another way of looking at this issue is as follows. If NICE does not advise OIR then it must advise either to approve or to not approve. If it advises not to approve, because although the treatment is clinically the more effective it is judged not to be sufficiently effective to be cost-effective, then no one receives this treatment on the NHS. Compared with that situation, an OIR decision would benefit some patients and harm none. If NICE was to approve the treatment then all patients for whom the treatment is relevant would benefit and, by comparison, an OIR decision would harm some patients [on definition (2) above]. But other patients might be unfairly harmed by the decision to approve because if, in fact, the treatment is not sufficiently effective to be cost-effective the opportunity costs of providing the treatment outweigh the benefits. And because the relevant research is not being carried out, it will remain unknown that this is the case. In any event it would be problematic for a REC to refuse to sanction the type of research we are considering once NICE had made an OIR decision because that condemns both present and future populations to receive only the inferior treatment, whether or not the more effective treatment is sufficiently effective.

Coercion

Another research ethics principle relevant to the sort of OIR decision under discussion is that competent patients have the right to consent to participate in, and withdraw from, a research project. This is akin to the competent patient's right to consent to treatment, both rights being underpinned by the principle of respect for individual autonomy. Conversely, it is impermissible to coerce competent patients to participate in research. In OIR the patient can have the more effective intervention on the NHS only if he or she agrees to be a research participant. Does this coerce patients to participate in the trial?

Any research study involving care clinically superior to that available on the NHS provides an incentive to enrol in that study. Whether or not such an incentive constitutes coercion depends on whether or not the patient is being presented with a threat as opposed to an offer to participate. Importantly, in the sort of OIR decisions under discussion, patients who do not receive the new treatment – that is, patients not enrolled on the trial, and participants not allocated to the new treatment arm – will receive standard NHS care. Arguably, then, the trial provides an offer, namely, the chance of receiving better than standard treatment, as opposed to any threat, and so does not constitute coercion. To clarify, suppose, by contrast, that patients will be refused access to normal NHS care unless they agree to participate in the research; this would present a threat as opposed to an offer to participate, and thereby constitute coercion. In

sum, providing that standard research ethics requirements are met – principally, that prospective participants are properly informed about, and give valid consent to participate in, the trial, and that if they do not participate their access to standard NHS care will not be affected – the research required by OIR is not coercive and respects the principle of individual autonomy because patients retain the right to choose whether or not to accept the chance of better than standard treatment offered by the trial.

Equity

A further ethical worry is that OIR decisions result in inequity because participants in one arm of the trial receive better treatment than both those in the other arm and those not participating in the research. It is important to distinguish two versions of this inequity charge. The first is that it is always wrong to allocate health resources in ways that will lead to an unequal distribution of health benefits. This version of the inequity worry is bound to founder in this context because NICE's values and principles, which are taken as granted here, entail that limited health resources will and should be allocated to maximise benefit to the whole population, even at the expense of subgroups within it. So, the fact that OIR results in an unequal distribution of health benefits can be justified by NICE's principles if the research will provide sufficiently valuable evidence.

A second version of this inequity worry is more involved. Under OIR, some patients will receive the intervention (paid for by the NHS) and others will not, simply because the former happen to be in a position to participate in the research. Many of the factors that determine access to treatment through participation in research – such as geographical location, socioeconomic status and patient characteristics – should not be considered relevant to whether patients have access to the treatments being studied. This is a significant consideration but one that can be overridden. Specifically, in deciding whether or not an OIR recommendation is ethically acceptable, a judgement would need to be made whether or not the benefits of such a recommendation outweigh the lack of equity (although, in fulfilling an OIR recommendation, it would always be important to minimise such inequities).

Issues specific to 'approval with research' recommendations

'Approval with research' raises a problem about consent, which is related to the discussion of coercion above. AWR also raises a further distinctive issue centring on the likelihood of the research taking place.

Consent

Two established principles of medical ethics are that competent patients have a right to confidentiality (including a right to decide whether or not to disclose their personal medical information) and a right to informed consent to participate in research (including a right to decline to participate in, or to decide to withdraw from, a study with impunity). It might be argued that some AWR decisions transgress these rights because the required research may involve collecting data on long-term outcomes and adverse events on patient registries (or some similar system of epidemiological data collection) without the explicit consent of patients. This argument is unsound. For one thing, in many cases a patient's decision to have a treatment that was approved with research may imply consent to the relevant data collection. Furthermore, although it comprises personal medical information, the data collected will be anonymous, ameliorating concerns about breaching confidentiality. Finally, this type of data collection that may be required by an AWR decision is equivalent to current large-scale epidemiological research studies that are considered ethically permissible. So, concerning consent, and on grounds of consistency, AWR should be permitted as long as the research is conducted to the ethical standards normally required of data collection of this sort.

Incentivising research

A further ethical issue raised specifically by AWR recommendations involves the mechanisms used to give 'teeth' to the research requirement: how will NICE ensure that the relevant research is carried out? One option has a subtle ethical dimension. NICE might threaten that, if the research is not satisfactorily completed (e.g. by the relevant manufacturing company), the intervention would cease to be made available on the NHS. Although this would provide an incentive for the manufacturer to carry out the research, it raises the following problem. At time T(1), the AWR decision is made, that is, the intervention is funded by the NHS. Suppose that, at time T(2) – the time when NICE reconsiders the decision to fund – the research has not been carried out (or has failed to provide any further relevant information). At time T(2) NICE will be making a decision on exactly the same information and evidence as at T(1). In this case it would seem that NICE should make exactly the same decision, namely, to provide the intervention on the NHS. But if NICE decides to reject the intervention (on the grounds, for example, that the manufacturer had failed to carry out the relevant research) then patients could claim unfairness. The unfairness is that they were provided with the treatment on the NHS between T(1) and T(2) but not after T(2) even though the evidence is exactly the same in both situations.

This is a serious problem for AWR decisions but it is not essentially an ethical matter. Rather, it is another point at which the ethical discussion segues into practical considerations familiar from other sections of this report. Specifically, various problems will result from the research condition put on an AWR decision not being met, and not just the ethical quandary outlined here. This situation should be avoided by doing everything reasonable to ensure that the relevant research will be conducted and reported. In fact, the risk that the ethical issue described here will arise if the research condition is not met can be used to put further pressure on whoever is responsible for conducting the research.

Summary of ethical analysis

Although they create numerous ethical challenges, OIR and AWR judgements are ethically permissible given the values and principles that underpin standard NICE practices. Nonetheless, this ethical analysis has identified some issues requiring further thought, which are highlighted in this summary.

Futurity

The fact that OIR and AWR trade off the interests of present and future populations is justified given current NICE values, principles and practices. Nonetheless, such decisions rely on research in the future being conducted, reported and useful. Future events are less predictable the more distant they are from the present and so the benefits of early approval are more secure than the value of future research. This might justify approving clinically beneficial treatments even though further research to determine their degree of effectiveness would be valuable.

Research Ethics Committees

'Only in research' decisions when clinicians and researchers would not be in equipoise about the effectiveness (as opposed to the cost-effectiveness) of an intervention are consistent with current NICE values, principles and practices, and further justified by the pressure brought on the principle of equipoise from elsewhere. Nonetheless, RECs are used to requiring traditional equipoise, that is, uncertainty about effectiveness. So, RECs may disallow trials required by an OIR decision unless they are informed of, and agree with, the rationale for contravening the traditional principle of equipoise.

Informing participants

Participants in trials required by an OIR decision are not coerced. Nonetheless, patients are strongly incentivised to participate by the prospect of being allocated to the new treatment arm of a trial. Two things must be made very clear to prospective participants: that participation does not guarantee that the new treatment will be received and that standard NHS care will be neither denied nor compromised if a patient declines to participate in, or decides to withdraw from, a study.

Unfair access to a trial

The inequity involved in an OIR decision is justifiable given current NICE values, principles and practices. Criteria for access to the treatment trial have to be established. Nonetheless, thought should be given to how to minimise unfairness. For example, if geographical location is a criterion of access, over-representation of certain regions at the expense of others should be avoided.

Registries

Evidence development required by AWR may involve a health data registry. These evidence development mechanisms should comply with standard epidemiological research ethics practices. For example, explicit consent to provide data should be collected whenever this is feasible, data collection and storage should comply with current data protection legislation and data should be anonymous.

Incentivising research

There is an important practical question as to whether or not an AWR research requirement will be met. One option is to incentivise researchers by threatening to withdraw approval if the research fails to report. This has a subtle ethical dimension: to carry out the threat would be unfair on future patients because the evidence base has not changed since the original AWR judgement. This is a form of inequity that strengthens the requirement to ensure that research conditions on decisions to approve treatments are met.

Chapter 4

A review of 'only in research'/'approval with research' recommendations in NICE technology appraisal guidance

Introduction

The National Institute for Health and Clinical Excellence issues technology appraisal guidance on the use of new and existing health technologies in the NHS. The recommendations are formulated by independent Appraisal Committees following a review of evidence and submissions from interested parties on the technology of interest. Since its inception, NICE has occasionally issued guidance that includes recommendations for further research as an alternative to binary 'accept' or 'reject' decisions. Previous studies that have considered the use of NICE research recommendations have tended to focus on the OIR-type recommendations.[3] It remains unclear whether the remit of NICE includes provision for issuing guidance in which approval is conditional on research being conducted (AWR). For example, NICE has no specific budget for research funding to accompany its recommendations to the NHS. In addition, the organisations that are responsible for implementing NICE guidance also do not have dedicated funding for NICE-recommended research. Nevertheless, AWR-type recommendations appear to have been issued by NICE in the past.

The National Institute for Health and Clinical Excellence provides its committees with general guidance on the health technology assessment methodologies and social value judgements it considers to be most appropriate for the formulation of NICE guidance.[1,83] These documents refer to the possibility of issuing recommendations in the context of research. The document describing the social value judgements that NICE expects its Advisory Committees to consider phrases the possibility of issuing research recommendations quite broadly in a way that could capture both OIR- and AWR-type recommendations: 'NICE's advisory bodies may recommend the use of the intervention within a research programme if this will provide more information about its effectiveness, safety or cost'.[83] However, the *Guide to the Methods of Technology Appraisal* refers more specifically to OIR-type recommendations and states that 'the Appraisal Committee may recommend that particular interventions are used within the NHS only in the context of research'.[1] Also, NICE has recently categorised its guidance into four groups (recommended, optimised, only in research and not recommended) and does not include an AWR category.

In its guidance to Advisory Committees, NICE specifies four issues that should be considered when recommending further research within the guidance. These are (1) whether or not the intervention is reasonably likely to benefit patients and the public, (2) how easily the research can be set up or whether it is already planned or in progress, (3) how likely the research is to provide further evidence and (4) whether or not the research is good value for money.[1] However, it is not clear to what extent these criteria have been considered in the formulation of NICE guidance. In addition, there is no guidance on when NICE advisory bodies should consider recommending research rather than 'accept' or 'reject' decisions.

The main aim of this review was to identify where OIR/AWR recommendations were made or considered in the development of NICE guidance. Secondary aims were to identify the considerations that led to the recommendations for further research and to assess the implementation of the OIR/AWR recommendations based on reviews of published guidance.

Methods

Inclusion and exclusion criteria

All NICE technology appraisal draft and final guidance up to January 2010 was considered for inclusion in the review. NICE guidance documents are published in a standardised format, with the guidance to the NHS presented in section 1. The rest of the document provides an overview of the evidence (sections 2 and 3), an explanation of how the evidence was interpreted by the committee (section 4), additional information to assist the implementation of the guidance (section 5) and a list of recommended related research (section 6). The research recommendations in section 6 do not form part of the formal guidance to the NHS. The committee's final recommendations [Final Appraisal Determinations (FADs)] are made publicly available and can be appealed by specific stakeholders before becoming final guidance to the NHS. In 2002 the NICE process was amended to also publish draft guidance [Appraisal Consultation Documents (ACDs)] for public consultation. This review included both final published guidance and draft guidance. In some cases multiple ACDs or FADs were issued by NICE when guidance changed following consultation; in these cases all ACDs and FADs, along with the final guidance, were reviewed.

The criterion for inclusion was guidance from the Technology Appraisal Programme that recommends research, or refers to ongoing or planned research, in the guidance section of the documents. The research recommendations could be framed either as OIR or AWR based on the following definitions for this review:

- OIR – a recommendation which states in the guidance section that the technology should not be used routinely unless it is in the context of further research
- AWR – a recommendation which states in the guidance section that the technology should be used routinely and which recommends further research is conducted.

Only documents that have been made publicly available are included [specifically, ACDs for Technology Appraisals (TAs) TA1–43, except TA32, were not made publicly available]. Documents that have been publicly released but later removed from the NICE website are included in the review (e.g. guidance that has been replaced by a subsequent review) and have been obtained directly from NICE where appropriate. Draft recommendations that request further clarification or analysis from the sponsor of the technology (sometimes referred to as 'minded no' recommendations in the STA process) are excluded as they usually require the reanalysis of existing data rather than additional data collection. The documents containing OIR/AWR recommendations were cross-checked with a list of appraisals including OIR recommendations complied by NICE to check for potential omissions (Sarah Garner, Associate Director of Research and Development, 4 March 2010, personal communication).

Data extraction and analysis

Data from each draft and final guidance document that included OIR and/or AWR recommendations were extracted using a template developed for the project (see *Appendix 3*). The template was developed following the review of the literature described in *Chapter 2*, and iteratively alongside the development of the framework described in *Chapter 3*. Data extracted

included information on the appraisal process, information on the technology, reported and accepted incremental cost-effectiveness ratios (ICERs) and the reasons for including an OIR or AWR recommendation in the guidance.

Data were extracted by one reviewer (JY) and a sample cross-checked by another reviewer (LL). The data were analysed to identify common characteristics of appraisals that included OIR and/ or AWR recommendations. When OIR/AWR recommendations changed between draft and final guidance, explanations for the change were reviewed and assessed. In addition, the extracted data were analysed to assess the extent to which the assessments proposed in the framework in *Chapter 3* were evident in informing the OIR/AWR recommendation in the draft or final guidance. Specifically, the following assessments were considered:

- significant irrecoverable costs
- cost-effectiveness of the technologies
- need for further research and the type of evidence requested
- possibility of conducting research with and without approval
- impact of price on the considerations
- resolution of uncertainties over time
- relative costs and benefits of research
- other considerations.

Significant irrecoverable costs

The guidance was reviewed for consideration of significant irrecoverable costs and explanations of their impact on the recommendations. It became clear early on that there would be little or no mention of these costs in the guidance, therefore an additional step was taken to assess whether or not the technologies were likely to incur irrecoverable costs. A possible indication of whether or not a technology has irrecoverable costs is if the directive to make NHS funding available for the treatment has been extended. Usually the directive is for funding to be available within 3 months of guidance being issued; however, this period can be extended if there are issues relating to implementation (e.g. a need to amend the NHS infrastructure or a need for significant training of staff to deliver the technology). The OIR/AWR guidance was cross-checked against the appraisals that have received extensions to the 3-month directive, and associated documents from the Department of Health website were reviewed for information on the reason for the extension.

Cost-effectiveness of the technologies

The NICE *Guide to the Methods of Technology Appraisal*[1] states that all appraisals should include an assessment of cost-effectiveness as a standard part of the NICE appraisal process. ICERs reported as considered most plausible by the committee and ICERs reported by the Evidence Review Group (ERG) or Assessment Group (AG) were extracted and classified according to the NICE threshold range for cost-effectiveness: < £20,000; £20,000–30,000 and > £30,000 per additional QALY gained.[1] In addition, consideration was given to whether or not reference was made to if the committee considered the technology to be a cost-effective use of NHS resources.

Need for further research and the type of evidence requested

The reasons for OIR/AWR recommendations and the type of evidence requested were reviewed from the 'Committee's Considerations' section in the guidance documents. Potential reasons for inclusion of OIR/AWR recommendations were identified at the outset after reviewing the preliminary findings of the systematic review and the draft framework for assessment. The following categories for which a need for further evidence was considered necessary were included:

- clinical effectiveness:
  - relative clinical effectiveness
  - data on the natural history/progression of disease
  - long-term data
  - relative clinical effectiveness for the OIR/AWR population
  - adverse effects
  - data to substantiate mechanism of treatment action
- cost-effectiveness:
  - to resolve uncertainty in cost-effectiveness estimates
  - cost-effectiveness data with an appropriate comparator
  - quality-of-life data
  - cost data
- other uncertainties:
  - budget impact
  - investment and irreversible costs
  - potential impact on ongoing research.

Possibility of conducting research with and without approval

The data were reviewed for reference to the possibility of conducting research and the potential impact of the recommendations on ongoing or future research. In addition, the type of research design requested (experimental or observational) was recorded. Although an imperfect proxy, this can give an indication of the possibility of conducting research because, as discussed in *Chapter 3*, it may be more difficult to obtain data on relative effectiveness if a technology has already been recommended for routine use as patients or their clinicians may be unwilling to participate in randomised trials that include the previous ('old') standard care.

Impact of price on the considerations

Standard NICE technology appraisal methodology is to consider the list price of technologies [e.g. as reported in the *British National Formulary* (BNF) for drugs] but not to take account of possible future price changes or local discounts.[1] However, national discounts such as those offered through patient access schemes may be considered.[86] The aim of these schemes is to 'improve the cost effectiveness of a medicine and therefore allow NICE to recommend treatments it would otherwise have found too costly'. Manufacturers may formally offer a reduction in the price of the technology to the NHS or may offer other schemes that reduce the overall cost of the technology to the NHS (e.g. by providing some courses of treatment at no cost). Although a formal process for patient access schemes was introduced in 2009, 'access' or 'risk-sharing' schemes have previously been adopted, for example the Department of Health risk-sharing scheme for beta-interferon.[87] However, unlike the multiple sclerosis scheme, data collection is not necessarily required for NICE patient access schemes, which can take many different forms (e.g. a simple reduction in price). Information on patient access schemes for OIR/AWR guidance was collected as part of the review.

Resolution of uncertainties over time

References to possible changes over time that could resolve the key uncertainties identified by the committee were sought from the guidance documents as well as the impact of these considerations on the guidance.

Relative costs and benefits of research

Reference in the guidance documents to consideration of the relative costs and benefits of research were sought. This could include formal value of information-type analyses or qualitative assessments.

Review of implementation of the 'only in research'/'approval with research' recommendations

Each piece of NICE guidance is considered for review at a specified length of time after publication (usually 3 years). To examine the impact of OIR/AWR recommendations on evidence collection and future guidance we considered guidance that included OIR/AWR recommendations and that had been updated by NICE. The documents were examined for changes in the evidence base available between original appraisal and review that were considered by the committee, and for changes to the decisions reflected in the guidance.

Results

Use of 'only in research' and 'approval with research' recommendations in NICE technology appraisal guidance

Of the 184 appraisals conducted up to January 2010, 40 included OIR/AWR recommendations in the draft and/or final guidance. In total, 29 FADs and 31 ACDs were issued relating to these 40 appraisals and which included OIR/AWR recommendations. Multiple ACDs were released for some appraisals; the 31 ACDs with OIR/AWR recommendations arise from 25 appraisals. All of the 29 FADs relate to the final guidance to the NHS (i.e. there were no changes to the recommendations between publication of the FAD and it becoming guidance). *Table 2* shows the frequency of OIR and AWR recommendations in the guidance documents. A list of all appraisals including OIR and AWR recommendations is provided in *Appendix 3*.

The research recommendations in the guidance most commonly took the form of OIR guidance. The terminology used in the OIR guidance differed between appraisals. Some guidance was specific about the type of research recommended whereas other guidance was much more general. For example, TA5[88] on the introduction of liquid-based cytology for cervical cancer screening is very specific in the research it recommends; it includes a whole host of recommendations and refers to 'a programme of pilot implementation projects' that should evaluate a range of impacts including 'the effect on test results', 'the extent to which productivity improvements in cytology laboratories are realised in routine practice' and 'the impact in the primary care setting'. However, the guidance for a subgroup of patients in TA163[89] simply states that 'infliximab should only be used for the treatment of acute exacerbations of active ulcerative colitis in clinical trials'. It could be inferred that the more specific recommendations are recommending the research in a more positive way, whereas some of the more general recommendations could be seen to be akin to a diluted reject decision sometimes referred to as a 'polite no'. Some guidance also refers to research ongoing at the time that the guidance was published. For example, TA8[90] on hearing aids states that 'There is insufficient robust scientific evidence to support the nationwide introduction of digital hearing aids at present. Evidence regarding the benefits of digital devices as compared with the current NHS range and to more sophisticated analogue devices, is expected to be available after the completion of research projects currently being undertaken in the UK.'

TABLE 2 Frequency of OIR/AWR recommendations in NICE guidance

Recommendation	Draft guidance	Final guidance
OIR	26	25
AWR	5	4
Total	31	29

Despite not being a formal category of guidance used by NICE, a handful of appraisals took the form of AWR guidance, that is, they recommended the technology but also recommended further research in the guidance section. One example of an AWR recommendation in final guidance comes from TA113[91] on the use of inhaled insulin. The guidance section recommends the use of inhaled insulin for a specific subgroup of patients and also includes the statement that 'Data on the use of inhaled insulin according to this guidance should be collected as part of a coordinated prospective observational study.' Another example of AWR guidance included very specific recommendations for further research. TA36[92] on treatments for rheumatoid arthritis recommended within the guidance that 'All clinicians prescribing etanercept or infliximab should (with the patient's consent) register the patient with the Biologics Registry established by the British Society for Rheumatology (BSR) and forward information on dosage, outcome and toxicity on a 6-monthly basis.'

Changes to the inclusion or exclusion of OIR/AWR recommendations between draft and final guidance were more common than suggested by the summary numbers in *Table 2*. Eleven appraisals included OIR/AWR recommendations in the draft guidance but not in the final guidance. Three appraisals included OIR/AWR recommendations in the final guidance but not in the draft guidance (ACDs were unavailable for a further 12 appraisals). The reported issues that led to the changes are explored further later in this chapter.

Most pieces of NICE guidance included several recommendations. These related to multiple technologies, multiple indications or different settings for the use of the technology. Over half of the OIR/AWR recommendations specified the need for further research in specified subgroups of patients (52% of OIR/AWR recommendations in final guidance documents). In approximately a quarter of cases, the OIR/AWR recommendations targeted a subset of the technologies included in the appraisal.

Table 3 shows the frequency of OIR/AWR recommendations by year since the beginning of the Technology Appraisal Programme, and the total number of final OIR/AWR recommendations as a proportion of the total number of pieces of final guidance that year. Sixteen per cent of all final guidance included an OIR/AWR recommendation. It appears that OIR/AWR guidance has been issued less frequently over the last few years. No final guidance included OIR/AWR recommendations in 2007, which is also the year that the STA process was introduced. Differences in the frequency of OIR/AWR recommendations were observed between the two NICE appraisal processes. Of appraisals issued through the multiple technology appraisal (MTA) process, OIR or AWR recommendations were included in draft guidance of 23 appraisals and final guidance of 28 appraisals. These 28 appraisals account for 19% of all final guidance issued within the MTA process. In the STA process only two ACDs and one piece of final guidance contained OIR/AWR recommendations. This accounts for just 2% of all final guidance issued through the STA process up to the time that the review was conducted.

The data were examined for differences in the use of OIR/AWR recommendations according to general disease area and the type of technology under appraisal. In absolute terms OIR/AWR recommendations were more common for cancer treatments ($n = 10$), accounting for over a third of all of the OIR/AWR recommendations in final guidance, followed by musculoskeletal conditions ($n = 7$), which accounted for almost a quarter of cases identified (*Table 4*). However, NICE has appraised a large number of treatments for cancer: 28% of all published appraisals over the review period. Only 7% of all NICE technology appraisal guidance has related to musculoskeletal conditions and so it appears that a disproportionate amount of this guidance has included OIR/AWR recommendations compared with appraisals for other conditions.

TABLE 3 Numbers of OIR/AWR recommendations by year of publication

Publication year	Draft guidance			Final guidance			Total no. of pieces of final guidance	% of final guidance including an OIR/AWR recommendation
	OIR	AWR	Total	OIR	AWR	Total		
2000	N/A	N/A	N/A	6	0	6	17	35
2001	N/A	N/A	N/A	2	0	2	14	14
2002	4	2	6	3	3	6	23	26
2003	3	0	3	4	0	4	19	21
2004	2	0	2	1	0	1	13	8
2005	6	1	7	3	0	3	7	43
2006	5	1	6	3	1	4	19	21
2007	2	1	3	0	0	0	21	0
2008	4	0	4	2	0	2	32	6
2009	0	0	0	1	0	1	19	5
Total	26	5	31	25	4	29	184	16

N/A, not available.

TABLE 4 Use of OIR/AWR recommendations by therapeutic area

Disease area	Draft guidance, n	Final guidance, n
Musculoskeletal	6	7
Cancer	8	10
Mental health and behavioural conditions	3	3
Eye	1	1
Ear and nose	1	1
Infectious diseases	1	1
Mouth and dental	0	1
Urogenital	0	1
Endocrine, nutritional and metabolic	2	1
Digestive system	1	1
Cardiovascular/central nervous system	2	2
Total	25	29

Just over half of the final guidance with OIR/AWR recommendations related to the appraisal of drugs ($n = 16$; 55%). However, taking into account the total number of pieces of drug guidance published, the use of OIR/AWR appears to be on average less common for drug appraisals: 11% of all drug appraisals within the period contained OIR/AWR compared with 47% of all guidance on therapeutic or surgical procedures and 27% of all guidance on devices.

Consideration of irrecoverable costs

Investment and irrecoverable costs were not explicitly quantified in any of the guidance documents and there were no documented considerations relating to these costs in the OIR/ AWR guidance. This suggests either that the technologies reviewed by NICE do not have significant irrecoverable costs or that such costs do not explicitly influence decision-making. Three of the technology appraisals that included OIR/AWR guidance (TA60,[93] TA68,[94] TA97[95]) had their funding direction extended and, although the rationale for this is not explicit on the

Department of Health website, the characteristics of the technologies suggest that it could be due, at least in part, to a need for significant additional training for staff. In addition, one appraisal explicitly noted some concerns around training although this was not directly cited as a reason for issuing the OIR guidance. TA51[96] (an earlier version of TA97[95]) on computerised cognitive behavioural therapy (CCBT) stated that 'Further information is required about the extent of training needed and circumstances under which different staff could provide support for users of CCBT.' Thus, it would appear that some of the technologies considered in the Technology Appraisal Programme are associated with significant irrecoverable costs even if these have not previously been explicitly incorporated into the decision-making process.

Cost-effectiveness of technologies

As expected given the NICE process, all OIR/AWR appraisals included a consideration of the cost-effectiveness of the technologies. Most of the guidance documents reported several different estimates of incremental cost-effectiveness based on analyses submitted by different stakeholders relating to different uses of the technology or based on different sets of assumptions or evidence. However, a formal assessment of cost-effectiveness was not always conducted or reported in the ACD or FAD for the use of the technology specified in the OIR/AWR recommendation. *Table 5* shows the ICERs (incremental cost per QALY gained) for the overall population and for the specific OIR/AWR indication where this differs. The table distinguishes between the base-case estimates of the ICER submitted by the AG/ERG in their original analysis and the estimate of the ICER considered most plausible by the Appraisal Committee following appraisal of all of the evidence submitted to NICE and noted in the guidance documents.

Most documents did not cite the ICER considered by the Appraisal Committee to be most realistic. The ERG/AG estimates were more frequently reported and it is likely that they were available in supporting documents and were not directly referred to in the ACD or FAD. ICERs for technologies with OIR/AWR recommendations were frequently >£30,000 (the upper bound of the NICE threshold range). The 'other' category includes ICERs that were reported using non-QALY-based outcome metrics and ICERs presented as a range that could not be classified into the categories. For example, TA5[88] on the use of liquid-based cytology reported ICERs of £1100 and £2500 per life-year gained depending on the length of the screening interval. When ICERs were not directly reported, there was usually an indication of whether the technology was considered to be cost-effective. For example, TA65[97] states that 'The clinical and cost effectiveness of rituximab in patients with localised disease has not been established' and TA44[98] states that 'Appraisal Committee believed that metal on metal hip resurfacing arthroplasty was likely to be of similar cost effectiveness to conventional total hip replacements in people who were expected to outlive the device'.

Table 6 shows the numbers of technologies stated to be cost-effective when used in the context of the OIR/AWR recommendation according to the final guidance document. In most cases an OIR recommendation was issued and the technology was on average not cost-effective. There were two examples of an OIR recommendation being used when the technology was probably cost-effective based on the accepted analyses. Both of these appraisals (TA5[88] on liquid-based cytology and TA51[96] on CCBT) requested that pilot implementation programmes be undertaken before routine introduction of the technologies in the NHS. The technology was considered likely to be cost-effective in three of the four cases in which an AWR recommendation was issued. In the remaining appraisal including an AWR recommendation [TA36[92] – etanercept (Enbrel®, Wyeth) and infliximab (Remicade®, Schering-Plough) for rheumatoid arthritis], the ICERs were £27,000–35,000, close to, or above, the upper end of the cost-effectiveness threshold range.

TABLE 5 Number (%) of OIR/AWR recommendations by category of incremental cost-effectiveness

Incremental cost per QALY	OIR/AWR indication		Total population	
	Committee's preferred estimate	AG/ERG's estimate	Committee's preferred estimate	AG/ERG's estimate
Not reported	30 (68)	22 (50)	23 (52)	10 (23)
Dominates	0	1 (2)	0	1 (2)
ICER <£20,000	0	2 (5)	1 (2)	4 (9)
ICER £20,000–30,000	4 (9)	1 (2)	5 (11)	3 (7)
ICER >£30,000	9 (20)	15 (34)	12 (27)	22 (50)
Dominated	0	2 (5)	0	1 (2)
Other	1 (2)	1 (2)	3 (7)	3 (7)
Total[a]	44 (100)	44 (100)	44 (100)	44 (100)

a　Total from 40 appraisals with OIR/AWR recommendations (29 FADs and 11 ACDs). Two appraisals reported ICERs separately for two technologies with OIR/AWR recommendations and one reported ICERs separately for three technologies.

TABLE 6 Type of recommendation and conclusion regarding cost-effectiveness (in FADs only)

Cost-effectiveness	OIR, n	AWR, n	Total, n
Considered cost-effective	2	3	5
Not considered cost-effective	23	1	24
Total	25	4	29

Need for further evidence and type of evidence requested

The rationales for issuing the OIR/AWR recommendations as stated in the guidance documents are summarised in *Table 7* (note that multiple reasons were cited in some cases). Three of the four appraisals that did not provide a rationale were issued before the committee's considerations were routinely described in the documents. The OIR recommendation in the other appraisal (TA75[99] on hepatitis C) related to three specific subgroups of patients; two subgroups were not referred to in the committee's considerations at all and no evidence was noted for the third subgroup.

A need for further evidence on the relative effectiveness of the intervention in the overall population or the OIR/AWR subgroup was the most commonly cited reason for issuing the OIR/AWR recommendation. A need for longer-term data was also frequently cited. Uncertainty in the cost-effectiveness estimates was also a common consideration; however, in all cases this was coupled with a need for further clinical evidence. Concern about the budget impact of introducing the technology, investment and reversal costs and the potential impact on ongoing research did not lead to the OIR/AWR recommendation in any of the appraisals.

In some appraisals the stated rationale for the OIR/AWR recommendation changed between draft and final guidance. For example, the use of infliximab for the treatment of acute exacerbations of ulcerative colitis (TA163[89]) was recommended OIR in both the ACD and the FAD. However, the final guidance cited only poor relative effectiveness evidence as a rationale for the OIR recommendation whereas the draft guidance also cited a need for other types of evidence such as information on adverse effects and cost-effectiveness data.

Possibility of conducting research with and without approval

No appraisals cited a concern for the impact of recommendations on ongoing trials as a direct rationale for the OIR/AWR guidance. In addition, the possibility of conducting research was not explicitly noted in most appraisals. Decision-makers may be more likely to use OIR

TABLE 7 Reasons for including research recommendations within the guidance

Reason for requesting further research	Draft guidance, n	Final guidance, n
None stated	1	4
Clinical effectiveness		
Need for more evidence on relative effectiveness in the whole population	19	16
Need for data on relative effectiveness in the target OIR population	15	9
Need for longer-term data	13	7
Need for information on adverse effects	6	4
Need for data on natural history/progression of disease	2	0
Need further evidence to support mechanism of treatment action	4	3
Cost-effectiveness		
Uncertainty in cost-effectiveness estimates	13	6
Need for cost-effectiveness data with an appropriate comparator	2	2
Need for more data on quality-of-life impact	6	3
Need for more data on costs	1	1
Other uncertainties		
Budget impact	0	0
Investment and reversal costs	0	0
Potential impact on ongoing research	0	0

TABLE 8 Type of research recommended

Type of research	Draft guidance, n (%)	Final guidance, n (%)
Experimental	21 (68)	14 (48)
Observational	5 (16)	7 (24)
Unclear (or both)	5 (16)	8 (28)
Total	31 (100)	29 (100)

recommendations than AWR when data on relative effectiveness are required. A lack of sufficient evidence on relative effectiveness was cited in 19 pieces of final guidance identified in the review and most of these included OIR recommendations (i.e. including those that stated a need for more evidence on relative effectiveness for the whole population *or* for evidence on relative effectiveness for the target OIR population in *Table 7*). The exception to this was TA113[91] (inhaled insulin), which included an AWR recommendation for a highly selective subgroup of patients, and which noted the committee consideration that data on relative effect would be most appropriately collected through a disease register.

Experimental research designs, such as RCTs, were more frequently referred to, and were slightly more common in draft guidance (*Table 8*). Two appraisals cited a need for further evidence on relative effectiveness in the final guidance but recommended observational studies because of anticipated difficulties in conducting RCTs in the specific OIR patient population (TA37[100]) or indication (TA167[101]). There were changes in the recommended type of research between draft and final guidance, which were mainly due to changes in the target OIR/AWR population (e.g. TA68[94]) or changes to be less specific about the research design (e.g. TA89[102]).

Overall, reference to the likelihood of research being conducted was rare in the guidance documents. However, there was one notable exception in which the likelihood of further research being conducted led to a change in the guidance: the ACD for TA129[103] stated that bortezomib monotherapy was not recommended except for use in well-designed clinical studies; however, this recommendation was removed in the FAD, which noted that no further studies of bortezomib were planned and highlighted a difficulty of conducting further research in the group of patients to which the guidance related (this guidance was later further amended following the offer of a patient access scheme).

The impact of price on 'only in research'/'approval with research' recommendations

Price changes or discounts featured in two of the appraisals including OIR/AWR recommendations. A patient access scheme was incorporated into TA129[103] and the guidance changed from reject to approval. This appraisal had previously included an OIR recommendation in draft guidance but was changed to reject prior to the offer of the access scheme as described in the previous section. In the second appraisal, an OIR recommendation was revised to approval after the committee revised their estimates of cost-effectiveness based on discounted prices of the technology along with further information on quality-of-life improvements (TA166[104] on cochlear implants).

Other considerations

Considerations around whether uncertainties in the evidence base would resolve over time were not explicitly mentioned as reasons for issuing OIR or AWR recommendations. In addition, the relative costs and benefits of conducting research were not reported as considerations of the committee when formulating its research recommendations.

Review of implementation of the 'only in research'/'approval with research' recommendations

Among the appraisals with OIR/AWR recommendations in the final guidance, 10 were later reviewed by NICE, including two that were incorporated into NICE clinical guidelines. *Table 9* provides details of the appraisals and whether or not additional evidence was provided and the change to the OIR/AWR recommendation (new evidence for other recommendations included within the guidance is not noted in the table).

In the majority of reviewed appraisals (*n* = 7), new evidence informing the OIR/AWR recommendation was available for the review. In four of these reviews, the OIR or AWR restriction was removed and the technology was recommended routinely. In two cases the additional evidence was considered insufficient to warrant a change in the OIR recommendation. In the remaining appraisal the OIR was revised so that some technologies within the class were recommended routinely whereas OIR recommendations were issued for others (TA51[96] on CCBT). In all cases the changes in the guidance were owing to the new data relating to the evidence gap identified in the OIR/AWR recommendation.

In three cases no new evidence was provided on the OIR/AWR indication. For the review of TA6,[105] no new RCT data were available for the OIR recommendation, which was made more restrictive in the review guidance. New evidence on clinical effectiveness was not available for the review of TA33,[106] but further information on adverse effects was provided. This was considered inadequate and no change was made to the OIR recommendation. The OIR recommendation was removed from the review of TA37[100] despite a lack of new evidence presented. The documents state that the reasons for this were a reduction in demand for the drug in this setting (it had since become licensed and NICE approved for treatment of an earlier stage of disease) and concerns about the feasibility of future data collection.

TABLE 9 Details of appraisals that underwent review

Original	Review	Additional evidence provided for the OIR/AWR indication	Summary of change to OIR/AWR guidance
TA5[88]	TA69[107]	New evidence available. Pilot implementation programmes were requested in the OIR. A Scottish implementation study and other evidence became available	OIR removed: technology recommended
TA6[105]	TA30[108]	No additional evidence presented	OIR amended: TA30[108] includes an OIR recommendation for a more restricted indication
TA16[109]	TA89[102]	Updated RCT data and new non-RCT evidence	OIR unchanged (some amendments to types of evidence required)
TA17[110]	TA105[111]	New evidence (RCTs) available	OIR removed: technology recommended
TA30[108]	CG81[112]	New evidence (RCT and registry data) available	OIR removed: CG81[112] did not include the OIR indication in the scope of the guideline
TA33[106]	TA93[113]	No new RCTs but updated adverse effect data	OIR unchanged
TA36[92]	TA130[114] (only in ACD)	New RCT and registry data available	AWR removed: technology recommended. A new OIR for another use of the drugs in ACD, but this was removed in the FAD (no guidance provided for this use of the drugs)
TA37[100]	TA137[115]	No new evidence presented	OIR removed: technology recommended
TA51[96]	TA97[95]	New evidence (RCT and non-RCT) available	OIR amended: original OIR was for CCBT as a class; amended OIR was for specific CCBT packages
TA72[116]	CG79[117]	New evidence (RCTs) available	OIR unchanged

Conclusions

The National Institute for Health and Clinical Excellence has issued OIR/AWR recommendations in 16% of its published technology appraisal guidance. These recommendations have most frequently taken the form of OIR recommendations. OIR recommendations have mainly been issued for technologies considered not to be cost-effective by the Appraisal Committee; however, there has been occasional use of OIR for technologies for which the best available evidence suggests that they may be cost-effective: liquid-based cytology for cervical cancer screening and CCBT. In both of these cases, the implementation of routine use of the technologies in the NHS could have required fairly substantial infrastructure or training requirements and possibly significant irreversible costs. In addition, in both cases the recommended research referred to 'pilot implementation projects'. Changes in the evidence base of reviewed appraisals show some limited success in the conduct of research recommended in OIR/AWR guidance; however, it is unclear if the research can be directly attributable to the NICE recommendations.

The most common reason cited for the OIR/AWR recommendations was the need for further evidence on relative effectiveness. Although uncertainty over the cost-effectiveness of the OIR/AWR technologies was cited in several appraisals, this was always accompanied by concern about the uncertainty around the clinical effectiveness. Some of the assessments recommended in the proposed framework do not appear to be currently considered when formulating guidance. Consequently it is not possible to categorise the recommendations into the 35 possible decision pathways described in *Chapter 3*. The lack of explicit consideration given to irrecoverable costs when formulating OIR/AWR guidance is notable. It seems unlikely that such costs do not occur in the provision of the technologies considered by NICE, particularly given that the introduction of some of the technologies has required changes to the existing provision of NHS infrastructure that have been expected to take longer than the standard 3 months.

The National Institute of Health and Clinical Excellence has recently issued a categorisation of all of its technology appraisal guidance. There are some differences between the NICE categorisation

and the results of this review because of differences in the definitions employed. The most notable differences relate to the classification of AWR guidance. NICE does not use the terminology of AWR in its classification system. However, this review has identified a small number of appraisals that apparently fall into this category of approving a technology for use, and also recommending research within the guidance to the NHS. In all of these cases, observational studies and/or data collection through disease registers were recommended. The NICE categorisation refers only to final guidance. Of the four pieces of final guidance identified in this review, NICE categorised one as recommended, two as optimised and one as OIR. The lack of a formal category of AWR guidance from NICE most likely reflects its remit, which is to make recommendations on the best of technologies within the NHS rather than to make recommendations on research to research funders. Despite this, there is clearly ambiguity in the terminology used in the guidance and differences in the interpretation of recommendations for research made within the guidance section of the documents.

One striking finding from this review is the decline in use of OIR/AWR recommendations over the past 5 years. The decline in the use of OIR/AWR recommendations coincides with the introduction of the STA process in 2006. Only one appraisal conducted through this process – which is now the most commonly used route for new technologies – included an OIR recommendation in the final guidance and none included AWR recommendations. At first glance this may appear to be counterintuitive. Technologies appraised through this process are usually new and therefore have a more limited evidence base than technologies appraised through the MTA process. However, it could also be that the STA process has started to shift the burden of proof of effectiveness and cost-effectiveness onto manufacturers and sponsors of technologies and recommendations to the NHS regarding the research of these technologies are seen as less relevant. It could also be linked to an increased opportunity to negotiate on the costs of technologies through patient access schemes or to tighter time and resource constraints in the production of STA guidance. The rarity of OIR/AWR recommendations in the STA process does not fully account for the reduction over time and there has been a decline in OIR/AWR recommendations within the MTA process. There is no clear reason for this from the documentation included in this review. However, insights gained from discussion with NICE stakeholders at workshops held as part of this research have suggested that it may be in line with other developments since the establishment of NICE – including refinements to its processes and the phrasing of the guidance. In addition, a possible explanation may include a realisation of the difficulties of developing and implementing OIR/AWR recommendations – perhaps because of difficulties in getting research funders to implement the recommendations.

One limitation of this analysis has been the reliance on the documented considerations of the Appraisal Committee in formulating its recommendations and whether these, in some cases fairly brief, summaries fully reflect all of the considerations that led to the recommendations and specifically the recommendations for research. In addition, identifying the ICERs related to the OIR/AWR indication and considered by the committee to be the most plausible was difficult as it was not always clearly stated and/or available from the documentation, although the clarity of reporting of ICERs has improved in the more recent documentation. Finally, focusing on the reviews as an indication of the success of the research recommendations could bias towards a positive finding as a lack of new evidence could have led to the postponement of planned reviews. However, information from the NICE website suggests that research to potentially inform a review is being conducted in most of the appraisals including OIR/AWR recommendations. Sixteen OIR/AWR appraisals have been considered for review: six have been postponed pending the reporting of ongoing research and a further six are ongoing or scheduled. Only two reviews have been cancelled because of a lack of new evidence and a further two have been cancelled because of the technology becoming obsolete.

Recent developments at NICE have increased the potential for further research alongside approval, although the uptake remains limited. One opportunity has arisen through the supplementary guidance to NICE committees about technologies used to treat patients at the end of life.[118] This guidance describes criteria for when the committees should consider departing from the usual criteria for cost-effectiveness. In addition, the guidance states that when recommending a treatment under the end-of-life criteria, NICE 'will normally recommend to the Department of Health that it should give consideration to a data collection exercise for treatment recommended for use on the basis of the criteria set out in Section 2. The purpose of this will be to assess the extent to which the anticipated survival gains are evident when the treatments involved are used in routine practice. The outcome of this exercise will be evaluated when the guidance for that treatment is reviewed.' However, in practice the uptake of this recommendation appears to be limited and in an early review of the policy it was noted that implementation of such schemes had proven problematic and was likely to be particularly difficult for non-cancer treatments.[119] Another opportunity has arisen through the introduction of patient access schemes in 2009; however, although these allow for additional data collection, it is not a requirement.[86]

The National Institute of Health and Clinical Excellence has recently issued a methods manual for the identification and prioritisation of research recommendations.[120] Although the guidance is aimed at the research recommendations provided outside of the main guidance to the NHS in the NICE documentation, the process also briefly refers to research recommendations in the guidance. A process is described whereby the Advisory Committees identify the 'key' evidence gaps, which are then translated into brief research questions by NICE staff in collaboration or consultation with other stakeholders. Prioritisation is conducted by NICE staff based on the 'existing state of the evidence' and reviewed at an annual meeting with the NIHR Evaluation, Trials and Studies Coordinating Centre. No further recommendations on when research recommendations should be included in guidance are described.

This review has revealed that NICE has used OIR/AWR since its inception, but these recommendations appear to be on the decline. Consideration of cost-effectiveness is routine within the appraisal process, and consideration of whether further research is needed and the type of evidence required appears to be frequently considered for OIR/AWR appraisals. However, some of the assessments outlined in the proposed framework do not appear to be currently considered, most notably whether or not significant irrecoverable costs are likely and whether or not the benefits of the proposed research outweigh the costs.

Chapter 5

Informing the assessments

The key principles and assessments that are needed when considering OIR or AWR guidance have been outlined in *Chapter 2*. How these assessments might be made and whether the existing methods of appraisal are sufficient, or whether additional information, evidence and analysis might be useful, were not addressed. In this chapter we outline additional information and evidence that might be useful and a range of methods of analysis that could be used to inform each of the assessments and decisions within the algorithm. We take existing methods of NICE appraisal as an accepted starting point and focus instead on exploring what additional information and analysis might feasibly be included in appraisal and how it might be interpreted to inform the judgements required using a series of case studies. We also consider whether this type of additional information and analysis might be routinely required within appraisal or only when OIR or AWR appears to be particularly relevant, for example more sophisticated additional analysis might be required only if it is established that further research might in principle be worthwhile.

Full details of each of the case studies are reported in *Appendices 7–10*. A separate technical appendix (see *Appendix 6*) is also provided that describes in detail how to carry out the calculations that can inform each of the assessments described in this chapter.

A checklist of assessment

The possible sequences of assessment and decision that lead to particular categories and types of guidance were represented as an algorithm in *Figures 1* and *2* and *Figures 26–28* in *Appendix 2*. The sequence of judgements required can be summarised as a simple checklist that could be considered by the Technology Appraisal Review (TAR) team/ERG and Appraisal Committee as well as manufacturers during appraisal. There are two checklists: one for technologies expected to be cost-effective (*Table 10A*) and one for those not expected to be cost-effective (*Table 10B*) based on the balance of existing evidence and current effective prices. The only difference between the checklists is at point 4, where, for technologies expected to be cost-effective, the judgement is whether the research is possible *with* approval whereas a judgement of whether research is possible *without* approval is required if the technology is not expected to be cost-effective.

Each of the seven points on the checklist relate to the sequence of decision nodes that fully describe the algorithm in *Appendix 2*. Therefore, each sequence of 'yes' or 'no' judgements defines a single pathway leading to a particular type and category of guidance (the type and category of guidance implied by each combination is described in *Table 32*, *Appendix 4*). However, all seven assessments do not necessarily need to be undertaken because sometimes earlier decisions will lead directly to guidance. For example, a 'no' at point 3 always leads directly to either approve or reject and hence further assessment is unnecessary. Similarly, a 'no' at point 6 also leads directly to approve or reject if there are no significant irrecoverable costs associated with the technology (see *Table 32*, *Appendix 4*).

TABLE 10A Checklist for OIR and AWR: technologies expected to be cost-effective

Point	Assessment	Judgement Yes	No
1	Is it cost-effective?	Yes	
2	Are there significant irrecoverable costs?		
3	Does more research seem worthwhile?		
4	Is the research possible with approval?		
5	Will other sources of uncertainty resolve over time?		
6	Are the benefits of research greater than the costs?		
7	Are the benefits of approval greater than the costs?		

TABLE 10B Checklist for OIR and AWR: technologies not expected to be cost-effective

Point	Assessment	Judgement Yes	No
1	Is it cost-effective?		No
2	Are there significant irrecoverable costs?		
3	Does more research seem worthwhile?		
4	Is the research possible without approval?		
5	Will other sources of uncertainty resolve over time?		
6	Are the benefits of research greater than the costs?		
7	Are the benefits of approval greater than the costs?		

Introduction to case studies

The objective of developing a series of case studies was to (1) demonstrate how the key principles and assessments might inform the development of guidance through application of the checklist and (2) establish whether existing methods of appraisal are sufficient, or whether (and when) additional information and analysis might be useful.

Selection of case studies

Case studies were selected to ensure that the full range of possible analysis was feasible within the time and resource constraints of this research project, while exploring situations in which OIR or AWR might be particularly relevant and challenging. Therefore, de novo analysis or substantial reanalysis of original assessments is not possible. Nor would it be necessary or informative as one of the objectives is to explore what *additional* information and analysis might be required. For this reason candidate case studies that met the following feasibility criteria were considered: (1) the economic analysis was regarded as a suitable basis for developing guidance, (2) an analysis of uncertainty in expected cost-effectiveness [probabilistic sensitivity analysis (PSA) as specified in the NICE reference case] was conducted and (3) ready access was available to the electronic versions of the models that informed guidance.

There are three groups of potential case studies in which the key principles and assessment described above might have influenced guidance: (1) when OIR or AWR was included in the FAD, (2) when OIR or AWR was considered during appraisal (e.g. included in the ACD or section 6 of the technology appraisal guidance) and (3) when OIR or AWR was not obviously

considered at any stage. As well as examples of AWR for technologies expected to be cost-effective and OIR for those not, there are also a number of particularly interesting ways in which guidance might be influenced by these additional considerations. For technologies expected to be cost-effective these include (1) OIR rather than approve when research *is not possible* with approval and (2) OIR or even reject rather than AWR or approve even if research *is possible* with approval because there are significant irrecoverable costs.

To fully explore the implications of these principles and assessments it is useful to select case studies that reflect the range of possible and interesting characteristics. For example, technologies (1) that are and are not expected to be cost-effective, (2) with and without irrecoverable costs, (3) for which other sources of uncertainty are and are not present, (4) for which the research needed is and is not possible with approval, (5) that are non-pharmaceutical interventions and (6) that are appraised under the MTA and STA process. Four case studies will not be enough to demonstrate the full range of possible combinations of interesting characteristics or illustrate all of the potential impacts of interest. Therefore, in selecting case studies there was a need to balance feasibility and coverage of those characteristics of greatest interest.

Background to the case studies

The following four case studies were selected. A range of additional information was sought and further analysis conducted to inform the sequence of assessment and judgements required when completing the OIR/AWR checklist in *Tables 10A* and *10B*.

Enhanced external counterpulsation for chronic stable angina

The NIHR HTA programme identified enhanced external counterpulsation (EECP) as an important topic and commissioned a short report to examine the clinical effectiveness and cost-effectiveness of EECP as an adjunct to standard therapy in patients with chronic stable angina. Although the topic was not ultimately considered by NICE it was commissioned in the same way and with the same resources as other assessment reports that inform NICE guidance. The assessment followed the NICE reference case and is consistent with the type of analysis that would have been required in a MTA appraisal. Like other MTA TARs it was published in full as a HTA monograph.[75,121]

Enhanced external counterpulsation is a non-invasive procedure (adjunct to standard therapy) used to provide symptomatic relief from stable angina. The analysis compares EECP with standard therapy alone. RCT evidence suggests an improvement in health-related quality of life (HRQoL) with EECP at 12 months.[121] To characterise the uncertainty associated with possible longer durations of treatment effect, formal elicitation of expert clinical judgement was undertaken. This provided an estimate of the probability, with uncertainty, of continuing to respond to treatment with EECP in subsequent years.

The possible pathways through the algorithm that the EECP case study illustrates are reported in *Figure 30* in *Appendix 4*. In this case study the new technology is expected to be cost-effective but with potentially significant irrecoverable costs. These irrecoverable costs include both (1) long-lived costs associated with the purchase of equipment and (2) large initial per-patient treatment costs, combined with a chronic condition in which a decision not to treat a particular patient with EECP can be changed at a later date (decisions are not irreversible) when research reports or other events occur. Consequently, these irrecoverable costs might influence the category of guidance, for example OIR rather than approve. This case study also provides an opportunity to explore the impact of research design (length of follow-up) on guidance and to examine the potential role of elicitation rather than extreme scenarios in characterising uncertainty.

Clopidogrel for the management of patients with non-ST segment elevation acute coronary syndromes

The use of clopidogrel (CLOP) (for up to 12 months) in combination with low-dose aspirin was recommended by NICE following a MTA appraisal for patients with non-ST segment elevation acute coronary syndrome (NSTE-ACS) presenting with a moderate to high risk of ischaemic events (TA80[122] in 2004 and updated in 2010 in CG94[123]). In TA80 the Appraisal Committee considered 12-month or lifetime treatment with CLOP but recommended research to inform optimal treatment duration. The original assessment report had included an analysis of shorter treatment durations (< 12 months) and the NIHR HTA programme subsequently commissioned additional reanalysis based on this original work to inform this research recommendation in 2009. This case study is based on the reanalysis of TA80 undertaken in 2009, which included standard therapy compared with four alternative treatment durations of CLOP of 1, 3, 6 and 12 months. Importantly, although the case study is based on the later reanalysis of TA80, the analysis considered here has been undertaken from the standpoint of the original TA80 appraisal, asking what assessments might have been made at that time when standard therapy was low-dose aspirin.

The research recommendation was made in section 6 of TA80; therefore, this case study is not an example of AWR at FAD but an example of AWR considered during appraisal. The possible pathways through the algorithm that the CLOP case study illustrates are reported in *Figure 29* in *Appendix 4*, with the new technology expected to be cost-effective and have no significant irrecoverable costs. The CLOP case study also illustrates a number of other important characteristics, including (1) the impact that other sources of uncertainty (price change following patent expiry) can have on the value of further research, (2) the interpretation of analyses in which there are multiple alternatives and (3) the use of scenarios to represent alternative but credible assumptions.

Omalizumab for the treatment of severe persistent allergic asthma in children aged 6–11 years

The use of omalizumab (OMAL) for the treatment of severe persistent allergic asthma in children aged 6–11 years was not recommended by NICE following a STA appraisal (TA201[124] in 2010). The analysis compared OMAL as an add-on to standard care compared with standard care alone. The primary analysis was based on a prespecified severe asthma population within an international, multicentre, placebo-controlled RCT. However, a high-risk subgroup within this population (recent hospitalisation for an asthma exacerbation) was also identified post hoc.

Omalizumab was not found to be cost-effective in either the severe or severe/high-risk population. However, a RCT was recommended comparing OMAL with oral corticosteroids (OCS) in children to establish reduction in OCS use. This was made in section 6 of TA201; therefore, OMAL is not an example of OIR at FAD but an example of OIR considered during appraisal. The possible pathways through the algorithm that the OMAL case study illustrates are reported in *Figure 31* in *Appendix 4*, with the new technology expected not to be cost-effective and to have no significant irrecoverable costs. OMAL also illustrates assessment in small patient populations (rare disease) and how subgroup analysis can be considered.

Etanercept, infliximab and adalimumab for patients with active and progressive psoriatic arthritis

Following a MTA appraisal (TA199[125] in 2010) the use of biologic treatment with etanercept, infliximab and adalimumab (Humira®, Abbott) was recommended by NICE for patients with active and progressive psoriatic arthritis (PsA) and who have an inadequate response to standard treatment, including two conventional disease-modifying antirheumatic drugs (DMARDs). However, the guidance also recommended that treatment should start with the least expensive

drug, taking account of dose, route of administration and price. This guidance updated an earlier MTA appraisal in 2006 (TA104[126]) that had recommended etanercept and restricted guidance on the use of infliximab to only those patients shown to be either intolerant or contraindicated to etanercept.[a] The analysis in this case study is from the standpoint of TA199, using the updated model that included new evidence and adalimumab as an additional comparator. At this point NICE guidance recommended etanercept and so the first question posed in the checklist can be interpreted as whether or not the other technologies available (infliximab, adalimumab or palliative care) are expected to be cost-effective compared with etanercept.

In section 6 of TA199 the importance of data from patient registries on long-term outcomes and adverse events was highlighted; therefore, the PsA case study is not an example of AWR at FAD but an example of AWR considered during appraisal. The possible pathways through the algorithm that the PsA case study illustrates are reported in *Figure 32* in *Appendix 4*. In this case study the alternatives to etanercept are not expected to be cost-effective. However, etanercept as well as infliximab and adalimumab have potentially significant irrecoverable costs because of the high initial per-patient treatment costs, combined with a chronic condition in which treatment decisions are not irreversible. The PsA case study, like the EECP case study, also provides an opportunity to examine the potential role of elicitation in the appraisal process.

Is it cost-effective and what are the risks?

The judgements made at points 1 and 2 of the checklist are critical because, although neither leads directly to a particular category of guidance, they determine the subsequent path that might be taken, sometimes avoiding further and potentially complex assessments. For example, the absence of significant irrecoverable costs means that only four out of the 12 possible pathways require all seven assessments to be made (see *Table 32, Appendix 4*).

Point 1: Is it expected to be cost-effective?

The sequence of assessments starts with cost-effectiveness and the expected impact on population NHEs, that is, at the following point in the algorithm:

This requires an assessment of expected cost-effectiveness based on the balance of the evidence and analysis currently available. Methods to estimate expected cost-effectiveness are well established within the NICE appraisal process and are extensively described in the *Guide to Methods of Technology Appraisal*.[1,b] Commonly, expected cost-effectiveness is summarised and presented using ICERs. Equivalently, but more usefully in this context, cost-effectiveness can be expressed in terms of expected NHEs, which can be expressed per-patient treated or for a population of patients. This is especially important when later assessments require a comparison of benefits for current and future patient populations and when assessing the significance of irrecoverable costs (see *Point 2: Are there significant irrecoverable cost?*). All of the information

required to express expected cost-effectiveness in these ways is available from the type of analysis already undertaken during appraisal.

Cost-effectiveness at the patient level

Estimates of the expected NHS costs and QALYs for each patient treated over an appropriate time horizon – the 'patient time horizon' – can be summarised as an ICER, which must be compared with a cost-effectiveness threshold to judge cost-effectiveness too.[c] Equivalently, this can be expressed as the per-patient NHE of each intervention, that is, the difference between any health gained and health forgone elsewhere.[d]

Technologies expected to be cost-effective

The results for EECP are summarised in *Table 11*. There are only two alternatives (EECP and standard care) and so only one ICER. EECP is just expected to be cost-effective at a threshold of £20,000 per QALY.[e] Consequently, the NHEs of EECP are greater than those of standard care but the difference per patient treated (the incremental NHE) is small.

It is also important to consider how NHEs accumulate over time or the investment profile per patient treated with EECP. *Figure 3* illustrates the cumulative incremental NHE over the patient time horizon. The initial per-patient costs of EECP are high and are far in excess of the immediate health benefits in the initial period of treatment. These negative NHEs are gradually offset by positive NHEs in later periods. In this case, it is only after 14 years that the initial losses are compensated by later gains, that is, EECP is not expected to break-even until 14 years from

TABLE 11 Expected cost-effectiveness of EECP per patient treated

Treatment	Cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£)	Incremental NHE, QALYs (£)	NHE, QALYs (£)	Incremental NHE, QALYs (£)
EECP	4744	7.6045	19,391	7.3673 (147,346)	0.0074 (149)	7.4464 (223,391)	0.0865 (2595)
Standard care	–	7.3598	–	7.3598 (147,197)	–	7.3598 (220,795)	–
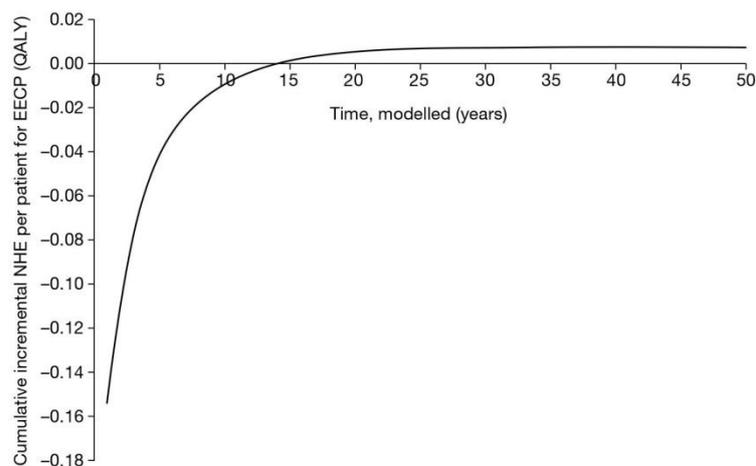
FIGURE 3 Cumulative incremental NHEs of EECP over the patient time horizon.

initial treatment. It is only beyond 30 years that the modest incremental NHEs reported in *Table 11* are eventually achieved.[f]

Multiple alternatives

Similar analysis can be conducted when there are more than two alternatives. For example, in the CLOP case study four treatment durations as well as current NHS treatment (aspirin alone) were considered at the time of TA80.[122] The results in *Table 12* indicate that 12-month treatment with CLOP is expected to be cost-effective, although the difference in NHE between 12 and 6 months' treatment is small.

The investment profile of CLOP, per patient treated, is illustrated in *Figure 4*. The per-patient costs of CLOP are in excess of the health benefits during the period of treatment. These negative NHEs are eventually offset by positive NHEs in later periods. In this case, it is only after 5 years that 12 months of treatment with CLOP breaks-even against current NHS care and it is not until 21 years that it is better than a shorter treatment duration of 6 months. Notice that shorter treatment durations with CLOP offer a much less 'risky profile', for example the break-even point for 1 month of treatment is 2 years against current NHS care.

Technologies not expected to be cost-effective

The ICER for OMAL in *Table 13* is greater than the threshold and so it is not expected to be cost-effective compared with standard care alone. Consequently, the incremental NHE of OMAL is negative.

TABLE 12 Expected cost-effectiveness of CLOP per patient treated

Treatment	Cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY NHE, QALYs (£)	Cost-effectiveness threshold £30,000 per QALY NHE, QALYs (£)
CLOP12	20,127	8.122	18,663	7.115 (142,307)	7.451 (223,525)
CLOP6	19,860	8.107	10,477	7.114 (142,288)	7.445 (223,362)
CLOP3	19,712	8.093	9396	7.108 (142,154)	7.436 (223,087)
CLOP1	19,598	8.081	4961	7.101 (142,025)	7.428 (222,837)
NHS	19,502	8.062	–	7.087 (141,734)	7.412 (222,353)

FIGURE 4 Cumulative incremental NHEs of CLOP over the patient time horizon.

The per-patient investment profile for OMAL is illustrated in *Figure 5* and shows that it is always expected to offer negative NHEs compared with standard care over the entire patient time horizon, that is, the high costs of treatment are never compensated by future health gains. In this example, the initial treatment costs with OMAL continue for 10 years (10 years is assumed to represent the duration that a patient would continue to receive treatment with OMAL) with health effects predominately while on treatment. Therefore, OMAL is not so much a 'risky purchase' but one that is simply not cost-effective at its current price.

Multiple alternatives

Psoriatic arthritis offers an example in which the alternatives to the treatment already recommended by NICE (etanercept at the time of TA199[125]) are not expected to be cost-effective, that is, the results in *Table 14* indicate that etanercept is expected to be cost-effective. Notice that although adalimumab is less effective than etanercept it is also cheaper; however, the resource savings it offers do not compensate for the reduction in health benefits.

Consequently, the investment profiles of the alternatives to etanercept, illustrated in *Figure 6*, differ in appearance. However, all of the biologic treatments for PsA have high initial costs, which are only gradually compensated for by later health benefits. Adalimumab, etanercept and infliximab all ultimately offer positive NHEs compared with palliative care but only break-even at 17, 17.5 and 34.5 years, respectively. Adalimumab offers a slightly less risky profile than etanercept and so it is only at 21.25 years that etanercept is expected to offer the highest NHE.

TABLE 13 Expected cost-effectiveness of OMAL per patient treated

Treatment	Cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£)	Incremental NHE, QALYs (£)	NHE, QALYs (£)	Incremental NHE, QALYs (£)
OMAL + standard care	94,992	16.64	93,844	11.8861 (237,721)	−2.1908 (−43,815)	13.4693 (404,078)	−1.2627 (−37,882)
Standard care	39,310	16.04	–	14.0768 (281,536)	–	14.7320 (441,960)	–

FIGURE 5 Cumulative incremental NHEs of OMAL over the patient time horizon.

TABLE 14 Expected cost-effectiveness of PsA treatments per patient treated

Treatment	Cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY	Cost-effectiveness threshold £30,000 per QALY
				NHE, QALYs (£)	NHE, QALYs (£)
1: Infliximab	90,343	7.269	60,965	2.752 (5504)	4.258 (8516)
2: Etanercept	78,150	7.069	17,733	3.161 (6322)	4.464 (8928)
3: Adalimumab	72,972	6.777	14,622	3.129 (6258)	4.345 (8690)
4: Palliative care	51,800	5.329	–	2.739 (5478)	3.602 (7204)
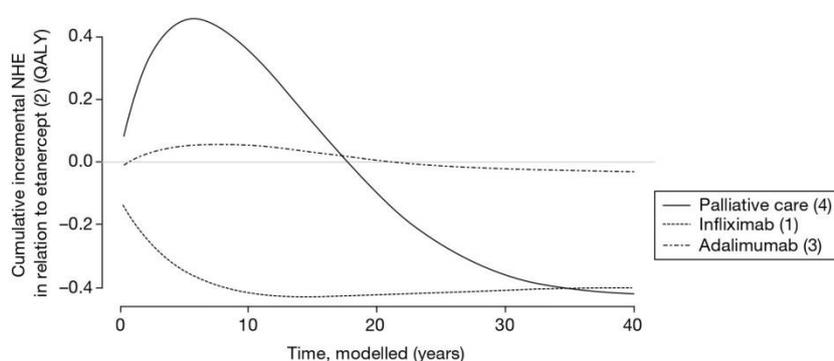
FIGURE 6 Cumulative incremental NHEs in PsA over the patient time horizon.

Cost-effectiveness at the population level

Per-patient NHEs can also be expressed for the population of current and future patients. This requires information about prevalence and future incidence of the target population (already required in appraisal). It also requires a judgement about the time horizon over which the technology will be used. This 'technology time horizon' ought to reflect the period over which the technology is likely to be part of clinical practice and generate the expected NHEs.[g] An estimate of the scale of the total population NHEs and how they cumulate over time is important for subsequent assessments, including (1) when the NHE for current patient populations must be compared with the benefits to future patients and (2) when the treatment decision can be changed so that the irrecoverable opportunity costs of initially negative NHEs become significant, that is, might influence the category of guidance.

For example, there is a large prevalent population ($n = 109,800$) eligible for EECP relative to future incident populations ($n = 9500$ per annum) in this chronic condition. The total population NHEs, assuming that the technology will be used to treat prevalent and incident patients over 10 years, are reported in *Table 15*. The expected cost-effectiveness is unchanged (the ICER is the same as in *Table 11*) but the incremental NHE, although small per patient, is more significant at a population level because of the size of the target population.

The investment profile for EECP when used to treat patients over 10 years is illustrated in *Figure 7*. At a population level it is not until 17 years (rather than 14 years at a patient level) that initial losses are compensated for by later gains and EECP breaks even. In other words, EECP appears a more risky investment when evaluated at a population rather than an individual level. This is because, although each patient treated with EECP is expected to offer the same profile of NHEs as shown in *Figure 3*, the negative NHEs associated with patients incident and treated

TABLE 15 Expected cost-effectiveness of EECP for the population

| Treatment	Costs (£M)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£M)	Incremental NHE, QALYs (£M)	NHE, QALYs (£M)	Incremental NHE, QALYs (£M)
EECP	896	1,435,787	19,391	1,391,001 (27,820)	1405 (28)	1,405,930 (42,177)	16,334 (490)
Standard care	–	1,389,596	–	1,389,596 (27,792)		1,389,596 (41,688)	

FIGURE 7 Cumulative incremental NHEs of EECP for the population.

in year 10 will not be offset by later gains until year 24. The population-level investment profile would exhibit greater risk (break-even later) if the prevalent population was smaller relative to the incident population and/or the technology time horizon was longer. For example, the break-even point extends to 23 years when the technology time horizon is increased to 20 years.

The effect on the other case studies of assessment at the population level is similar to the effect on the EECP case study. It simply increases the magnitude of differences in per-patient NHE (to a greater extent for longer technology time horizons) but leaves expected cost-effectiveness unchanged. However, the investment profiles at a population level also differ, exhibiting greater 'risk' indicated by later break-even points for the same reasons as for EECP. For example, the break-even points for CLOP when evaluated at a population level are reported in *Table 16*. At a technology time horizon of 10 years it is only at 11 years, rather than 5 years for a single patient, that 12 months of CLOP treatment breaks even against current NHS care and not until 27 years (rather than 21 years) that 12 months of CLOP treatment is better than a shorter treatment duration of 6 months. Even the shorter durations of treatment offer a 'risky profile', for example the break-even point for 1 month of treatment against current NHS care is 4 years (rather than 2 years).

TABLE 16 Expected cost-effectiveness of CLOP for the population

Technology time horizon	Treatment	Incremental NHE, QALYs (£M)	Break-even point (years)		
			12 months vs 6 months	12 months vs NHS	1 month vs NHS
5 years	CLOP12	269 (5.4)	24	8	4
	CLOP6	1881 (37.6)			
	CLOP3	1804 (36.1)			
	CLOP1	4073 (81.5)			
	NHS	–			
10 years	CLOP12	495 (9.9)	27	11	4
	CLOP6	3465 (69.3)			
	CLOP3	3324 (66.5)			
	CLOP1	7502 (150)			
	NHS	–			
15 years	CLOP12	686 (13.7)	30	12	4
	CLOP6	4799 (96)			
	CLOP3	4603 (92.1)			
	CLOP1	10,389 (207.8)			
	NHS	–			
20 years	CLOP12	846 (16.9)	33	12	4
	CLOP6	5921 (118.4)			
	CLOP3	5680 (113.6)			
	CLOP1	12,820 (256.4)			
	NHS	–			

Point 2: Are there significant irrecoverable costs?

The second point on the checklist requires (1) an assessment of whether there are irrecoverable costs and (2) a judgement of their potential significance, that is, at the following point in the algorithm:
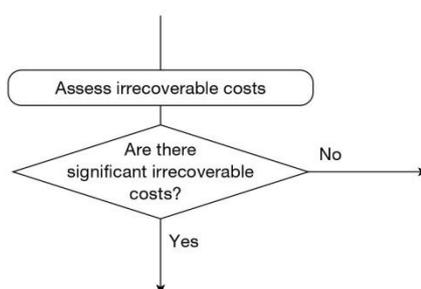
Irrecoverable costs are those that once committed cannot be recovered if guidance is changed at a later date (see *Appendix 6, Are there significant irrecoverable costs?*). Irrecoverable costs are most commonly thought of as upfront or capital costs of new facilities or equipment with long life expectancy (they might also include any practitioner training and the costs of implementation efforts). In NICE appraisal these types of cost are first annuitised[h] and then allocated pro rata to the number of patients likely to be treated during the lifetime of the equipment. That is, capital costs are treated as if they are paid per patient treated over the lifetime of the equipment.

If guidance remains unchanged throughout this period (i.e. research does not report or other sources of uncertainty do not resolve) then this common assumption has no influence. However, should guidance change (initial approval is withdrawn) before the end of the lifetime of the equipment then, although future patents will no longer use the technology, the cost of the equipment allocated to them cannot be recovered. The possibility that initial guidance might change and its impact on expected costs needs to be considered before costs are made irrecoverable through approval or AWR. The impact of irrecoverable costs will tend to be greater if they represent a greater proportion of the total costs, if guidance is more likely to change and if guidance is more likely to change in the near rather than distant future.

Enhanced external counterpulsation is the only case study in which these types of cost are present to any great extent because treatment requires capital investment in the EECP machines themselves. The expected per-patient and population costs reported in *Tables 11* and *15* allocated this capital cost in the usual way (i.e. annuitised over the 10-year lifetime of the machines and allocated to the number of patients treated each year). The irrecoverable costs are reported separately in *Table 17* and represent 19% of the total. However, this will have no influence on expected cost-effectiveness as long as guidance does not change during the lifetime of the equipment.[i]

Investment profile of net health effects

Even in the absence of capital costs of equipment and facilities, NHEs accumulate over time at both a patient and a population level. With the possible exception of OMAL[j] the analysis in the previous section indicates a common pattern of initially negative NHEs that are only gradually offset by positive NHEs in later periods. Therefore, approval or AWR commonly commits opportunity costs of negative NHE that are irrecoverable. For OMAL, the profile of NHEs at a patient level did not exhibit significant irrecoverable opportunity costs. Assessment at a population level and for longer technology time horizons simply increases the magnitude of the expected negative NHE; therefore, there are no irrecoverable costs in this case study.

Are irrecoverable costs likely to be significant?

Whether or not irrecoverable costs are significant, that is, might influence guidance, depends critically on whether guidance is likely to change and whether that is more likely in the near or in the distant future. That will depend on whether research is likely to be undertaken and when it is likely to report, as well as on other events that might occur, for example a change in price following patent expiry. These are assessed later, at points 5 and 6 in the checklist. However, the *potential* significance of any irrecoverable costs can be assessed at this point. For example, capital costs can be judged based on the proportion of total population costs that are irrecoverable for this reason as well as their scale relative to the additional population NHE offered (e.g. see *Table 17*).

Judging the potential significance of the investment profiles of NHEs is more nuanced. It depends whether treatment decisions for individual patients are irreversible, which in part depends on the nature of the disease. For example, in an acute condition the decision to treat a particular presenting patient with a technology cannot be reconsidered at a later date – it is irreversible.

TABLE 17 Capital costs associated with EECP

Treatment	Capital cost (£)	Non-capital cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY	
					NHE, QALYs (£M)	Incremental NHE, QALYs (£M)
EECP	170,304,591	725,408,798	1,435,787	19,391	1,391,001 (27,820)	1405 (28)
Standard care	–	–	1,389,596	–	1,389,596 (27,792)	–

Of course, it is possible that the later benefits are not realised but it is also possible that they will realise more (the profiles of NHEs in *Figures 3–6* are the average over these possibilities). Similarly, the possibility that guidance might change in the future (e.g. research suggests that the longer-term benefits will not offset initial losses) will not influence the irreversible decision to treat a presenting patient with a technology that is expected to be cost-effective before the research reports.

Implication for the case studies

Clopidogrel is a treatment for acute coronary syndromes and, although decisions about treatment and its duration are not irreversible in the very short run, over the time scales more likely for research being conducted (and reporting) or other events occurring that would change guidance they can be regarded as such. Therefore, although the investment profile of CLOP (at a patient and more so at a population level) exhibits irrecoverable costs these should not be judged significant in the sense that they have little potential to influence guidance. There are also no significant irrecoverable costs associated with OMAL but for different reasons; although treatment decisions are reversible in this chronic condition, any irrecoverable costs appear very limited (see *Figure 5*).

Both EECP and the biologics in PsA are for chronic conditions in which the decision to treat a particular patient can be changed at some later date (decisions are not irreversible). Therefore, the type of investment profile of NHEs at a patient and population level is significant because, instead of committing irrecoverable costs by deciding to use technologies expected to be cost-effective now, the decision and commitment of costs can be made later, after research reports, other events occur and/or guidance changes. Of course, proper account must be taken of the impact of withholding initiation of treatment on expected health benefits and costs (see *Point 7: Are the benefits of approval greater than the costs?*), for example some patients who might have been treated may not survive to benefit from the results of the research or disease may have irreversibly progressed so that the expected health benefits are lower.[75]

Enhanced external counterpulsation is the only case in which both types of irrecoverable costs are potentially significant. *Figure 7* illustrates the impact of accounting for the actual timing of expenditure on EECP machines rather than treating it as if it was paid when each patient was treated, that is, expenditure is treated like a consumable cost by spreading the capital cost over 10 years.[k] If approval of EECP might be withdrawn before 10 years, the potential losses in NHEs will be greater than initially indicated in *Figure 7* because the equipment costs allocated to treating future patients cannot be recovered. The earlier such a change might occur the greater the additional loss. The impact of these possibilities should be considered at point 7 of the checklist before guidance to approve or AWR commits both types of irrecoverable costs.

Pricing and irrecoverable costs

The importance of irrecoverable treatment costs when they may be potentially significant should also consider the scale of initially negative NHEs and the duration of such losses, that is, how long until the technology breaks even for an individual patient and for the population of patients who are likely to be treated if it is approved. Health technologies with patent protection are more likely to be priced close to the point at which the expected incremental NHEs are close to zero, that is, the ICER is close to or equal to the threshold. A value-based pricing scheme would formalise these existing incentives. A technology that is only just expected to be cost-effective will not break-even until close to the end of the patient time horizon and for much longer for the population of patents likely to benefit from its use (up to the technology time horizon *plus* the patient time horizon – less if patent expiry and cheaper generic enter before the technology time horizon). Therefore, those technologies already priced close to the threshold, and all new technologies considered in a value-based pricing scheme, will tend to increase the scale

of irrecoverable costs committed by approval, making OIR or reject more likely even when a technology is just expected to be cost-effective at point 1 of the checklist.[l]

Is further research required?

The judgements made at points 3 and 4 of the checklist are critical because if more research is not judged to be worthwhile no further assessments are required (unless there are significant irrecoverable costs; see *Table 32* in *Appendix 4*). If research is worthwhile, the type of evidence needed and whether or not the research required to generate it can be conducted, while the technology is approved will determine whether or not AWR or OIR is a possibility.

Point 3: Does more research seem worthwhile?

The third point on the checklist requires an assessment of the potential benefits of conducting further research, that is, at the following point in the algorithm:

This requires judgements about (1) how uncertain a decision to approve or reject a technology might be based on the estimates of expected cost-effectiveness and (2) whether or not the scale of the likely consequences of this uncertainty might justify further research. Some assessment of the potential consequences of uncertainty is important because it indicates the scale of the population NHEs that could be gained if the uncertainty surrounding this decision could be resolved immediately, that is, it represents an expected upper bound on the benefits of more research.[m] If the potential benefits of further research are unlikely to justify the costs, then a judgement that more research does not seem worthwhile will lead directly to guidance in the following circumstances (extracted from *Table 32* in *Appendix 4*):

	Assessment	1	2	3	4	5	6	7	Guidance
Pathway number	6	Yes	No	No	–	–	–	–	Approve[4]
	12	No	No	No	–	–	–	–	Reject[4]
	35	No	Yes	No	–	–	–	–	Reject[11]

Assessing the consequences of uncertainty

Some assessment is required of (1) how uncertain a decision based on expected cost-effectiveness might be and (2) what the consequences, in terms of population NHEs, are likely to be if an incorrect decision is made.

Enhanced external counterpulsation is expected to be cost-effective compared with standard care (*Table 18* and see *Table 11*), but the estimates of cost and QALYs are uncertain and so there is a chance that a decision to approve EECP based on existing evidence will be incorrect, that is, standard care might offer greater NHEs. Some assessment of the likely consequences

of approving EECP when standard care might be better could be based on the difference in expected NHEs, that is, the expected incremental population NHEs reported in *Tables 15* and *18*). This is illustrated in *Figure 8* in which a judgement about the probability that a decision based on expected cost-effectiveness is correct translates into expected consequences based on the expected incremental population NHE. For example, if the decision was judged to be 100% certain then there are no consequences and so there would be nothing to be gained by more research; however, as the probability that the decision is correct becomes less certain, the expected consequences (and hence potential value of more research) increase.

This judgement of how uncertain a decision might be can be informed by the PSA already used to estimate costs and QALYs and required as part of the NICE reference case.[127,128] The probability that EECP is cost-effective is 0.428 (see *Table 18*),[n] which would translate into approximately 800 QALYs (see *Figure 8*) over the technology time horizon,[o] based on the expected or average difference between NHEs. However, the difference in NHEs when EECP is not the correct decision is not necessarily the average. In fact, it is very unlikely to be the average and such estimates may substantially under- or overestimate the expected consequences of uncertainty.[p]

TABLE 18 Expected consequences of uncertainty for EECP

Treatment	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY			Cost-effectiveness threshold £30,000 per QALY		
		Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)	Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
EECP	19,391	1405 (28.1)	0.428	9287 (185.7)	1,405,930 (490)	0.700	2774 (83.2)
Standard care		–	0.572		–	0.300	

FIGURE 8 Probability that EECP is cost-effective and the consequences of uncertainty.

The same probabilistic analysis can be used to record the difference between the NHEs of EECP and the NHEs of standard care and the frequency of such errors (i.e. the proportion of simulations in which standard care offers greater NHEs). The distribution of consequences associated with these errors is illustrated in *Figure 9*. Commonly there are no consequences, because EECP is the correct decision (i.e. there is a 42.8% chance of zero consequences). However, when EECP offers lower NHEs than standard care (a 57.2% chance) there are consequences in terms of NHEs forgone. *Figure 9* illustrates that the consequences of errors may be relatively small, for example 9% of the simulations have consequences of < 5000 QALYs. However, consequences may also be very large, for example there is a small chance (5.7%) that they are > 30,000 QALYs. The average over this distribution provides the correct estimate of the expected consequences of uncertainty, which in this case is 9287 QALYs.[q]

These expected consequences can be interpreted as an estimate of the population NHEs over the technology time horizon that could be gained if the uncertainty surrounding this decision could be resolved immediately, that is, it indicates an expected upper bound on the benefits of more research.[127,129,r] The consequences can also be expressed as the equivalent NHS resources required to generate the same population NHEs (£185.7M in *Table 18*). They will increase with the size of the patient population and the technology time horizon. In the case of EECP the consequences fall with the cost-effectiveness threshold because a decision to approve EECP will be less uncertain (see *Table 18*). A judgement at this point that more research might be worthwhile seems reasonable because the upper bound on its potential benefits exceeds the likely costs.

Multiple alternatives

Similar analysis can be conducted when there are more than two alternatives but greater difficulties are encountered unless the results of the PSA are used to assess both uncertainty and its consequences. For example, in the CLOP case study, 12 months' treatment with CLOP is expected to be cost-effective but this is also uncertain. A judgement is required about the chance that 12 months of treatment is incorrect and if so which of the other four alternatives are likely to offer a higher NHE and how much higher. In other words, for decisions involving multiple alternatives, a judgement is required on the level of uncertainty surrounding the decision, how this uncertainty is distributed across the various alternatives and what the consequences are likely to be. The results of the PSA can inform this judgement. The probabilities that each of the five alternatives is cost-effective are reported in *Table 19*. This indicates that 12 months' treatment is uncertain (probability that it is incorrect is 0.476). However, much of this probability of error is allocated to 6 months' treatment with CLOP (0.18) for which the difference in NHEs is likely to be relatively modest.

The distribution of consequences is illustrated in *Figure 10*. Most commonly (52.4%) there are no consequences because 12 months' treatment with CLOP is the correct decision. When it is not, there is a greater chance of relatively small consequences of error (30% are < 10,000 QALYs), which occur predominantly when 6 months' treatment duration offers the highest NHE (18% chance). There is a small chance of much larger consequences (< 5% chance that they are > 30,000 QALYs). These occur only when standard NHS treatment offers the highest NHE. Overall, there is an almost 20% chance that NHS treatment offers a higher NHE, that is, there remains important uncertainty about the cost-effectiveness of CLOP itself, not just the duration of treatment. The expected consequence of uncertainty (5194 QALYs) is simply the average over this distribution. Again, this can be interpreted as an estimate of the population NHE that could be gained over the time horizon of this technology if the uncertainty about treatment and its duration could be immediately resolved. Therefore, like EECP, a judgement at this point that more research might be worthwhile seems reasonable because the potential benefits exceed the likely costs.
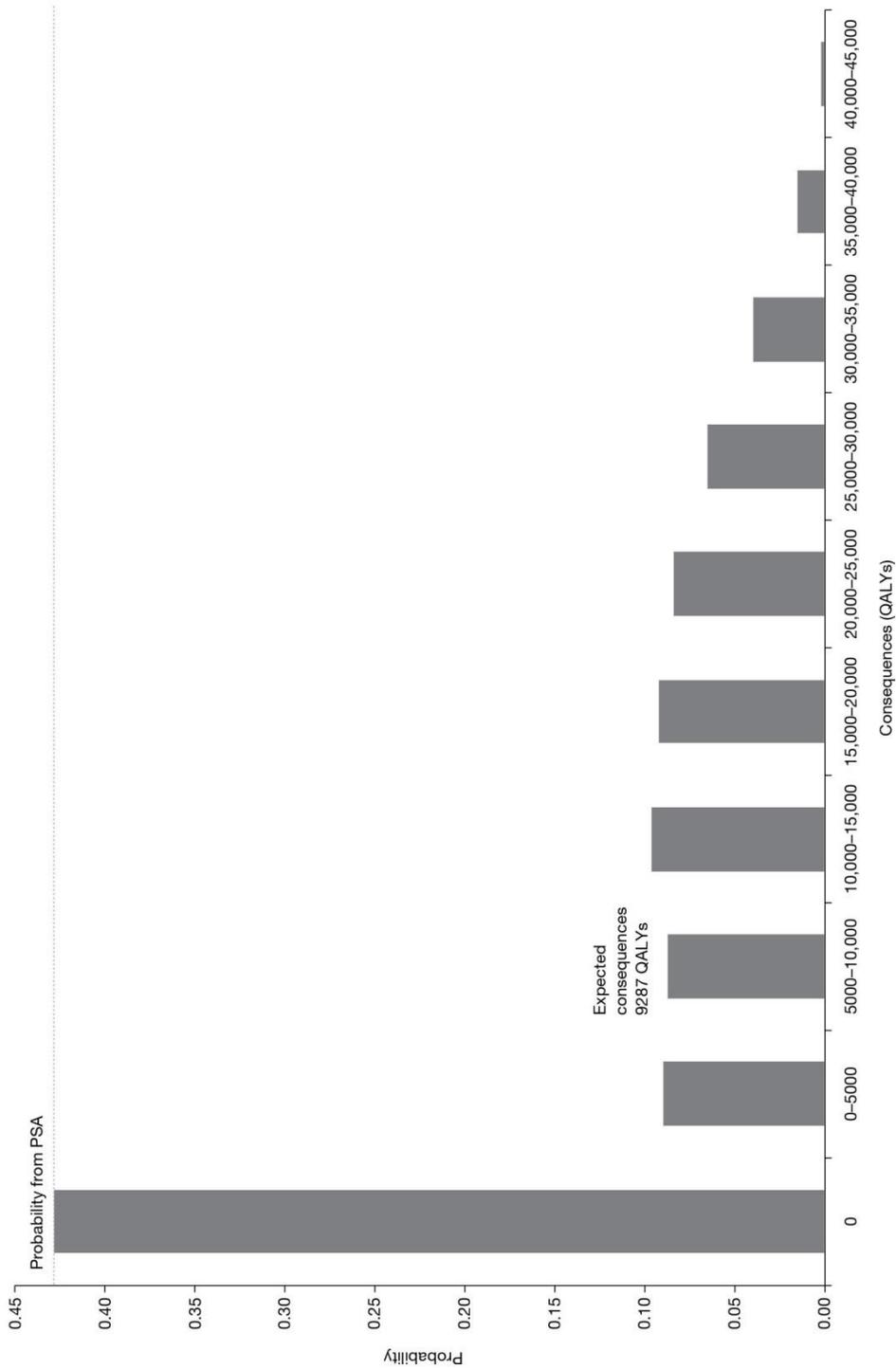
FIGURE 9 Distribution of the consequences of uncertainty for EECP.

TABLE 19 Expected consequences of uncertainty for CLOP

Treatment	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY			Cost-effectiveness threshold £30,000 per QALY		
		Incremental NHE,[a] QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)	Incremental NHE,[a] QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
CLOP12	18,663	495 (9.9)	0.524	5194 (103.9)	2798 (56.0)	0.677	3657 (109.7)
CLOP6	10,477	3465 (69.3)	0.180		4736 (94.7)	0.092	
CLOP3	9396	3324 (66.5)	0.018		4305 (86.1)	0.009	
CLOP1	4961	7502 (150.0)	0.075		8327 (166.5)	0.052	
NHS	–	–	0.202		–	0.170	

a The mean additional population NHE of moving from the least to the most effective alternative, that is, the incremental NHE of 12 months' treatment compared with NHS care is the sum of these increments (14,786 QALY or £295.7M at £20,000 per QALY)

FIGURE 10 Distribution of the consequences of uncertainty for CLOP.

Psoriatic arthritis provides a similar picture to that of CLOP, in which approval of the alternative (etanercept), which is expected to be cost-effective, is uncertain (probability that approval is incorrect is 0.557), but in this case most of this probability of error is associated with palliative care (probability of 0.4 that it is cost-effective). Again, there is a greater chance of relatively small consequences (19% are < 28,000 QALYs), most of which occur when adalimumab has the highest NHE, and a smaller chance of very large consequences (4.7% chance that they are > 138,000 QALYs), which occur when palliative care offers the greatest NHE. The expected consequences of uncertainty and the upper bound on the population NHE that might be gained by immediately resolving uncertainty (35,342 QALYs or £707M over the technology time horizon) supports a judgement that more research may be worthwhile.

Analysis of subgroups

Omalizumab was not expected to be cost-effective based on existing evidence. The ICER in *Table 13* was substantially greater than the threshold and a decision to reject this technology does not appear uncertain. This judgement is supported by the results of PSA (the probability that OMAL is cost-effective is zero in *Table 20*). Therefore, a decision to reject ('Reject[4]' in the algorithm) is not uncertain; there are no consequences of uncertainty and nothing to be gained by more research. However, it is possible to consider a high-risk subgroup within this population. Subgroups, once credibly defined, need to be considered in the same way, starting at point 1 on the checklist, that is, entering at the top of the algorithm (see *Appendix 6, Is further research*

TABLE 20 Expected consequences of uncertainty for OMAL

Treatment	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY			Cost-effectiveness threshold £30,000 per QALY		
		Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£)	Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£)
Severe population							
OMAL + standard care	93,844	−5789 (−116)	0.0	0	−3337 (−100)	0.0	0.0
Standard care		−	1.0		−	1.0	
High-risk subgroup							
OMAL + standard care	69,463	−3851 (−77)	0.0	0	−2048 (−61)	0.013	10.61 (0.32)
Standard care		−	1.0		−	0.987	

required?, Subgroups). Although the ICER for this high-risk subgroup is somewhat lower, it is still significantly higher than the threshold. The results of PSA suggest that even at a threshold of £30,000 the probability that OMAL is cost-effective is very small and the upper bound on the gains from more research are very limited (10.61 QALYs). Therefore, even after an analysis of subgroups OMAL is not expected to be cost-effective and more research does not seem worthwhile. OMAL can be rejected at this point and no further assessment is required.

Alternative scenarios

There are often alternative views about the quality and relevance of evidence as well as other assumptions that might be made when estimating expected costs and QALYs. These are commonly presented as separate scenarios, with estimates of costs and QALY presented for each. Much of the deliberation by the Appraisal Committee often surrounds the scientific value judgements required to judge the credibility of the alternative assumptions represented by the scenarios. The type of probabilistic analysis reported represents the uncertainty within each scenario and will be sufficient to indicate the potential benefits of research when only one scenario is regarded as credible. However, when more than one scenario might be credible and carry some 'weight', there will be uncertainty *between* as well as *within* scenarios. The 'weighting' of scenarios can be made explicit by assigning probabilities to represent how credible each is believed to be. The weighted average of costs and QALYs across scenarios can easily be calculated (see *Appendix 6, Is further research required?, Alternative scenarios*). It is also tempting to take a simple weighted average of the expected consequences of uncertainty across these scenarios as well. However, a simple weighted average may under- or overestimate the combined consequences of uncertainty within and between scenarios (see *Appendix 11, Why averaging scenarios may be misleading*).[130] The correct estimate requires the probabilities (weights) to be applied directly to the simulated output from PSA rather than to the mean values. Although this does not require additional simulation and is quick and easy to implement, it does require either that the probabilities are made explicit in advance or that estimates be presented for a range of probabilities that might represent the judgement of the Appraisal Committee following deliberation.

For example, the CLOP analysis presented above assumes a constant relative treatment effect for different durations of treatment (scenario A). An alternative assumption (scenario B) was that the relative treatment effect also differed by duration based on the data reported in the Scottish Intercollegiate Guidelines Network (SIGN) guidelines. This alternative assumption made longer durations less cost-effective and reduced the expected consequences of uncertainty from 5195 to

3969 QALYs. Although scenario A was regarded as more credible by the Appraisal Committee, scenario B might nevertheless carry some weight or have some probability associated with it. In this case the simple weighted average of expected consequences (linear combination of mean estimates) is very similar to the correct estimate based on weighting the output of PSA in *Figure 11*. This also shows how these estimates can be presented for a range of probabilities.

An alternative assumption of a common class effect across the three biologics was considered in the PsA case study (scenario B) but was judged less credible than the analysis that allowed differential effects (scenario A). The alternative scenario made etanercept less likely to be cost-effective and increased the expected consequences of uncertainty from 34,930 to 38,521 QALYs (*Figure 12*). In this case a simple weighted average of expected consequences based on the probability assigned to each scenario is, in general, lower than the correct estimate of expected consequences based on the output from PSA.

Elicitation

The single RCT of EECP showed evidence of improvements in quality of life at 12 months;[121] however, the degree to which these are sustained in the long run is uncertain. Rather than make alternative assumptions and present extreme scenarios, formal elicitation of the judgement of clinical experts about the likelihood of QALY gains in subsequent years was undertaken.[s] The uncertainty in these elicited values is included in the estimates of the expected consequences of
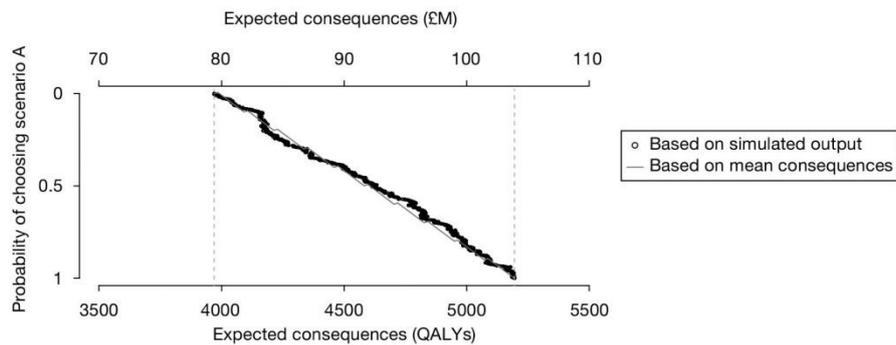
FIGURE 11 Expected consequences of uncertainty with alternative scenarios: CLOP case study.
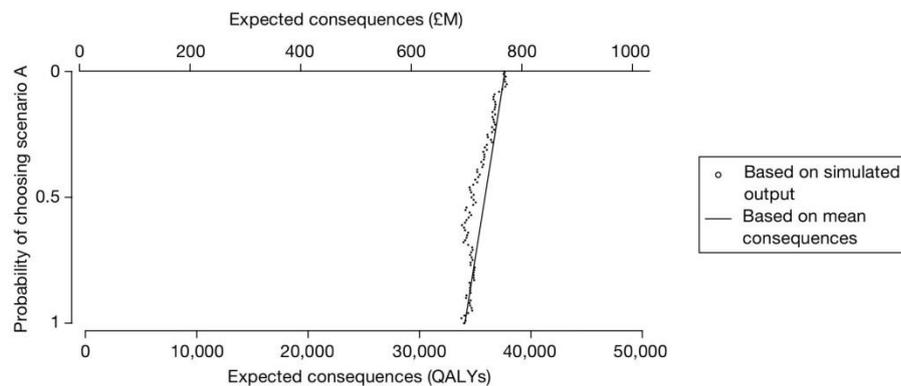
FIGURE 12 Expected consequences of uncertainty with alternative scenarios: PsA case study.

uncertainty reported in *Table 18*, which might otherwise have been represented by alternative scenarios, for example no QALY benefits beyond 12 months could be assumed for scenario A, benefits sustained for a patient's lifetime for scenario B and benefits sustained for 4 years for scenario C. The results of elicitation implied probabilities of 0.243, 0.353 and 0.404 associated with each of these scenarios, respectively. A simple weighted average of the expected consequences within each scenario using these probabilities (1442 QALYs) significantly underestimates both the estimate of expected consequences based on all of the information from elicitation (9287 QAYs) and the estimate based on weighting scenarios using the simulated output rather than the mean estimates (13,081 QALYs). This illustrates that (1) a simple weighted average of expected consequences may be misleading and (2) elicitation may provide a richer characterisation of uncertainty as well the probabilities associated with alternative assumptions (see *Appendix 11, Why averaging scenarios may be misleading* for further discussion).

Point 4: Is research possible with approval?

The fourth point on the checklist requires an assessment of what type of evidence is needed and a judgement of whether the research required to generate it can be conducted while the technology is approved, that is, at the following point in the algorithm:
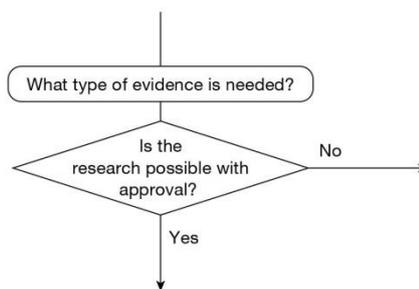
Although the decision at this point does not lead directly to guidance, it does determine whether AWR or OIR is a possibility. This judgement will depend, in part, on whether the type of evidence that is needed will require an experimental research design. For example, more precise estimates of relative treatment effect are likely to require a RCT if the dangers of selection bias are to be avoided; however, further RCTs for this particular indication and patient group are unlikely to be possible once a technology is approved for widespread NHS use.

This requires judgements about (1) how important particular types of parameters (inputs to the economic model) are to estimates of cost and QALYs, (2) what values these parameters would have to take to change a decision based on expected cost-effectiveness, (3) how likely is it that parameters might take such values and (4) what would be the consequences if they did, that is, what might be gained in terms of population NHE if the uncertainty in the values of these parameters could be immediately resolved (see *Appendix 6, Is the research possible if the intervention is approved?*).

Assessing the importance of parameters

The type of economic model used to estimate expected cost-effectiveness in NICE appraisal specifies the relationship between the inputs (the parameters) and the outputs (costs and QALYs). A simple summary of the direction and strength of these relationships can be provided by calculating elasticities for each, that is, the proportionate change in the NHE of each alternative, and differences in NHE, owing to a 1% change in the value of the parameter, for example those parameters with high elasticities (especially with respect to differences in NHE) might be regarded as more 'important'. These elasticities are presented for the CLOP case study in *Table 21*.

They give some indication of (1) the relative importance of parameters for certain comparisons (e.g. RR_death seems particularly important for all comparisons), (2) those parameters that are of no or very limited importance (e.g. parameters 1–6 in the comparison of 12 and 6 months' treatment) and (3) the direction of the relationship (e.g. the elasticity for C_Well is negative indicating that if the costs of NHS care in the well state are greater, then 12 months' treatment will be less cost-effective than 6 months' treatment or current NHS care).

Although these measures of importance are more instructive than a series of arbitrary one-way sensitivity analyses, they do not directly help the assessment of what values parameters must take to change decisions and how likely such values might be. A simple summary of the values that particular parameters must take to make each of the alternatives cost-effective can also be provided. These 'threshold values' for parameters are presented for the CLOP case study in *Table 22*. This provides additional information to the elasticities in *Table 21*, for example there are

TABLE 21 Elasticities associated with parameters: CLOP case study

Parameter			Elasticity over the NHEs (QALYs) of					Elasticity over the incremental NHEs (QALYs) of		
			CLOP12	CLOP6	CLOP3	CLOP1	NHS	CLOP12 vs NHS	CLOP12 vs CLOP6	CLOP12 vs all[a]
Natural history	1	P_die_0.1	−0.208	−0.207	−0.207	−0.207	−0.222	0.014	–	0.003
	2	P_NFMI_0.1	−0.012	−0.012	−0.011	−0.011	−0.015	0.004	–	–
	3	P_die_1.3	−0.137	−0.137	−0.137	−0.147	−0.145	0.008	–	0.004
	4	P_NFMI_1.3	−0.002	−0.002	−0.002	−0.002	−0.002	0.001	–	–
	5	P_die_3.6	−0.146	−0.146	−0.157	−0.157	−0.154	0.008	–	0.007
	6	P_NFMI_3.6	−0.005	−0.005	−0.007	−0.007	−0.007	0.002	–	0.001
	7	P_die_6.12	−0.148	−0.159	−0.158	−0.157	−0.155	0.007	0.011	0.010
	8	P_NFMI_6.12	−0.005	−0.007	−0.007	−0.007	−0.007	0.002	0.002	0.002
	9	TP_AC	−0.121	−0.120	−0.120	−0.120	−0.118	−0.003	−0.001	−0.002
	10	TP_AD	−3.637	−3.622	−3.604	−3.594	−3.541	−0.096	−0.016	−0.047
	11	TP_CD	−0.233	−0.235	−0.239	−0.240	−0.253	0.020	0.002	0.009
	12	TP_BD	−0.586	−0.593	−0.602	−0.605	−0.641	0.055	0.007	0.024
Utilities	13	U_Well	0.746	0.745	0.743	0.742	0.737	0.009	0.001	0.004
	14	U_Well1	6.090	6.064	6.034	6.017	5.929	0.160	0.026	0.079
	15	U_NFMI	0.133	0.134	0.136	0.136	0.144	−0.011	−0.001	−0.005
	16	U_PostMI	1.138	1.150	1.165	1.171	1.236	−0.099	−0.012	−0.043
Relative effect	17	RR_death	−0.639	−0.491	−0.344	−0.207	–	−0.641	−0.150	−0.380
	18	RR_NFMI	−0.024	−0.018	−0.013	−0.011	–	−0.025	−0.006	−0.014
Costs	19	C_Well	−0.740	−0.737	−0.733	−0.731	−0.720	−0.019	−0.003	−0.009
	20	C_MI_LT	−0.051	−0.052	−0.053	−0.053	−0.056	0.004	0.001	0.002
	21	C_PostMI	−0.142	−0.143	−0.145	−0.146	−0.154	0.012	0.002	0.005
	22	TC_Well_Dead	−0.027	−0.027	−0.027	−0.027	−0.027	–	–	–
	23	C_t1	−0.045	–	–	–	–	−0.045	−0.045	−0.045
	24	C_t2	–	−0.033	–	–	–	–	0.033	0.008
	25	C_t3	–	–	−0.026	–	–	–	–	0.007
	26	C_t4	–	–	–	−0.022	–	–	–	0.005
	27	C_t5	–	–	–	–	−0.016	0.016	–	0.004

C, cost; NFMI, non-fatal myocardial infarction; P, probability of events in the short-term decision tree; RR, relative risk; TP, annual transition probabilities between states [well (A), post MI (B), MI (C), dead (D)] of a Markov model characterising long-term progression of disease; U, utility.
a   A weighted average of the NHEs of the remaining treatments was used. The weights reflect the probabilities of each treatment being cost-effective.

TABLE 22 Thresholds associated with parameters: CLOP case study

Parameter			Mean value	CLOP12	CLOP6	CLOP3	CLOP1	NHS
Natural history	1	P_die_0.1	0.032	0 to 0.10	0.11 to 0.54	0.54 to 0.63	0.63 to 1	–
	2	P_NFMI_0.1	0.040	0 to 0.14	0.14 to 0.71	0.71 to 0.82	0.82 to 1	–
	3	P_die_1.3	0.022	0 to 0.10	0.10 to 0.55	0.55 to 1	–	–
	4	P_NFMI_1.3	0.004	0 to 0.10	0.10 to 0.7	0.7 to 1	–	–
	5	P_die_3.6	0.023	0.01 to 0.10	0.10 to 1	0 to 0.01	–	–
	6	P_NFMI_3.6	0.011	0 to 0.11	0.11 to 1	–	–	–
	7	P_die_6.12	0.024	0.02 to 1	0 to 0.02	–	–	–
	8	P_NFMI_6.12	0.009	0.005 to 1	0 to 0.005	–	–	–
	9	TP_AC	0.018	0 to 0.06	0.06 to 1	–	–	–
	10	TP_AD	0.072	0 to 0.08	0.08 to 0.10	–	–	0.10 to 1
	11	TP_CD	0.188	0.12 to 1	0 to 0.12	–	–	–
	12	TP_BD	0.07	0.06 to 1	0.04 to 0.06	–	–	0 to 0.04
Utilities	13	U_Well	0.798	0.29 to 1	0 to 0.29	–	–	–
	14	U_Well1	0.930	0.90 to 1	0.74 to 0.90	–	–	0 to 0.74
	15	U_NFMI	0.801	0 to 1	–	–	–	–
	16	U_PostMI	0.931	0 to 1	–	–	–	–
Relative effect	17	RR_death	0.931	0 to 0.93	0.94 to 0.97	0.97 to 0.98	0.98 to 0.99	1.00 to max.[a]
	18	RR_NFMI	0.710	0 to 0.82	0.83 to 1.55	1.56 to 1.83	–	1.84 to max.[a]
Costs	19	C_Well	2061.5	0 to 2690	2690 to 5611	–	–	5611 to max.[a]
	20	C_MI_LT	6050	0 to max.[a]	–	–	–	–
	21	C_PostMI	2309.7	870 to max.[a]	0 to 870	–	–	–
	22	TC_Well_Dead	871.5	0 to 20,474	20,474 to max.[a]	–	–	–
	23	C_t1	895.1	0 to 910	910 to max.[a]	–	–	–
	24	C_t2	651.6	630 to max.[a]	0 to 630	–	–	–
	25	C_t3	524.2	370 to max.[a]	–	0 to 370	–	–
	26	C_t4	434.8	150 to max.[a]	–	–	0 to 150	–
	27	C_t5	329.8	0 to max.	–	–	–	–

C, cost; max., maximum; NFMI, non-fatal myocardial infarction; P, probability of events in the short-term decision tree; RR, relative risk; TP, annual transition probabilities between states [well (A), post MI (B), MI (C), dead (D)] of a Markov model characterising long-term progression of disease; U, utility.

a   An upper bound was used in evaluating unbounded parameters: relative risk parameters were evaluated to a maximum value of 50 and cost parameters were evaluated to a maximum value of £100,000.

only six parameters that could possibly take values that would lead to current NHS care (without CLOP) generating higher NHEs than 12 months of treatment with CLOP. However, although instructive, such 'threshold values' do not indicate how likely it is that the threshold will be crossed or the combined effect of groups of related parameters.

Assessment of uncertainty

The judgement about how likely it is that parameters might take values that will change the technology expected to be cost-effective can be informed by the results of PSA. This is because the distributions assigned to parameters in PSA describe how uncertain the parameter estimates are, such that they ought to reflect the amount and quality of existing evidence. The probabilities that parameters might take values that would lead to each of the alternatives being cost-effective are reported for the CLOP case study in *Table 23*. This essentially decomposes the overall probabilities reported in *Table 19* into the contribution that each parameter makes.[t] Interestingly, it indicates that it is uncertainty in the estimate of relative effect (RR_death) that contributes most to the probability of error associated with 12 months of treatment. It is the only parameter

TABLE 23 Probabilities associated with parameter values: CLOP case study

Parameter			CLOP12	CLOP6	CLOP3	CLOP1	NHS
Natural history	1	P_die_0.1	1	–	–	–	–
	2	P_NFMI_0.1	1	–	–	–	–
	3	P_die_1.3	1	–	–	–	–
	4	P_NFMI_1.3	1	–	–	–	–
	5	P_die_3.6	1	–	–	–	–
	6	P_NFMI_3.6	1	–	–	–	–
	7	P_die_6.12	0.65	0.35	–	–	–
	8	P_NFMI_6.12	0.91	0.09	–	–	–
	9	TP_AC	1	–	–	–	–
	10	TP_AD	0.83	0.17	–	–	–
	11	TP_CD	1	–	–	–	–
	12	TP_BD	0.85	0.15	–	–	–
Utilities	13	U_Well	1	–	–	–	–
	14	U_Well1	0.94	0.06	–	–	–
	15	U_NFMI	1	–	–	–	–
	16	U_PostMI	1	–	–	–	–
Relative effect	17	RR_death	0.55	0.18	0.01	0.10	0.16
	18	RR_NFMI	0.97	0.03	–	–	–
Costs	19	C_Well	0.78	0.19	–	–	0.03
	20	C_MI_LT	1	–	–	–	–
	21	C_PostMI	0.89	0.11	–	–	–
	22	TC_Well_Dead	1	–	–	–	–
	23	C_t1	0.95	0.05	–	–	–
	24	C_t2	0.99	0.01	–	–	–
	25	C_t3	1	–	–	–	–
	26	C_t4	1	–	–	–	–
	27	C_t5	1	–	–	–	–

C, cost; P, probability of events in the short-term decision tree; RR, relative risk; TP, annual transition probabilities between states [well (A), post MI (B), MI (C), dead (D)] of a Markov model characterising long-term progression of disease; U, utility.

that (alone) might take values that could make any of the other alternatives cost-effective. It is also worth noting that there is a very small chance that cost in the 'well state' (C_Well) might be sufficiently high that standard NHS care would be cost-effective (i.e. a 3% chance that the cost of this state exceeds the threshold value of £5611 required for standard NHS care to have the highest NHE). In other words, if the NHS costs associated with the 'well state' are higher than the mean value then any cost savings associated with moving more patients more quickly to the well state [and thus avoiding the costs of additional events such as non-fatal myocardial infarction (NFMI)] will tend to be lower.

What type of evidence is needed?

Although an understanding of uncertainty and the importance of parameters separately is helpful, an assessment of the likely consequences of this uncertainty, and therefore what might be potentially gained in terms of population NHEs if uncertainty could be immediately resolved, is required. This assessment can directly inform the judgement of what evidence is needed and whether or not the type of research required to generate it will be possible with approval. As in the previous section on whether or not more research seems worthwhile, the results of PSA can inform this judgement as estimates of the expected consequences of uncertainty associated with

each parameter combine both uncertainty in its potential values and its importance in terms of changing decisions and the resulting differences in NHEs (see *Appendix 6, What type of evidence is needed?*). The expected consequences of uncertainty associated with each parameter in the CLOP case study are reported in *Table 24*. This decomposes the overall expected consequences reported in *Table 19* into the contribution that each parameter makes. It also identifies which of the alternatives to 12-month treatment might offer a higher NHE.[u] *Table 24* confirms that it is uncertainty in the estimate of relative effect (RR_death) that contributes most and for which there is potentially the most to be gained by resolving this uncertainty through additional research (4433 QALYs or £88.7M). Since more precise estimates of relative effects are likely to require a RCT, a judgement that the type of research needed will not be possible if a 12-month treatment duration is approved may be reasonable. However, the potential benefits of resolving the uncertainty associated with other groups of parameters, for example costs (547 QALYs or £10.9M) and the natural history (369 QALYs or £7.4M), might mean that other types of

TABLE 24 Consequences of uncertainty associated with parameter values: CLOP case study

Parameter			Expected consequences (QALYs)					
			Decomposed by treatment choice					
			CLOP12	CLOP6	CLOP3	CLOP1	NHS	Overall
Natural history[a]	1	P_die_0.1	0	–	–	–	–	0
	2	P_NFMI_0.1	0	–	–	–	–	0
	3	P_die_1.3	0	–	–	–	–	0
	4	P_NFMI_1.3	0	–	–	–	–	0
	5	P_die_3.6	0	–	–	–	–	0
	6	P_NFMI_3.6	0	–	–	–	–	0
	7	P_die_6.12	0	250	–	–	–	250
	8	P_NFMI_6.12	0	9	–	–	–	9
	9	TP_AC	0	–	–	–	–	0
	10	TP_AD	0	47	–	–	–	47
	11	TP_CD	0	–	–	–	–	0
	12	TP_BD	0	35	–	–	–	35
Utilities[a]	13	U_Well	0	–	–	–	–	0
	14	U_Well1	0	10	–	–	–	10
	15	U_NFMI	0	–	–	–	–	0
	16	U_PostMI	0	–	–	–	–	0
Relative effect[a]	17	RR_death	0	284	16	518	3614	4433
	18	RR_NFMI	0	3	–	–	–	3
Costs[a]	19	C_Well	0	153	–	–	321	474
	20	C_MI_LT	0	–	–	–	–	0
	21	C_PostMI	0	8	–	–	–	8
	22	TC_Well_Dead	0	–	–	–	–	0
	23	C_t1	0	8	–	–	–	8
	24	C_t2	0	0	–	–	–	0
	25	C_t3	0	–	–	–	–	0
	26	C_t4	0	–	–	–	–	0
	27	C_t5	0	–	–	–	–	0

C, cost; P, probability of events in the short-term decision tree; RR, relative risk; TP, annual transition probabilities between states [well (A), post MI (B), MI (C), dead (D)] of a Markov model characterising long-term progression of disease; U, utility.
a Expected consequences for groups of parameters are (1) natural history: 369 QALYs (£7.4M); (2) utilities: 15 QALYs (£0.3M); (3) relative effect: 4504 QALYs (£90.1M); (4) costs: 547 QALYs (£10.9M). These are not equal to the sum of expected consequences for component parameters for the reasons explained in *Chapter 5*, notes p and q.

cheaper, non-experimental research could be worthwhile as well or might be conducted before commissioning potential expensive experimental research that may take some time to complete and report.[131,v]

The full results tables for the other case studies are provided in *Appendices 7*, *9* and *10*. In summary, the EECP case study provides a similar pattern of results with the most significant consequences of uncertainty associated with parameters related to relative treatment effect, suggesting that the research needed might not be possible following approval of EECP. Interestingly, although the probability of sustaining the QALY benefits of EECP in the long run is very uncertain, the greater potential value is in more precise estimates of QALY gains in the first 12 months (8511 QALYs or £170M, respectively).

With regard to the PsA case study, on the other hand, the greater potential value is associated with uncertainty over the natural history of Health Assessment Questionnaire (HAQ) progression (8697 QALYs or £17.4M) rather than relative treatment effect (1201 QALYs or £2.4M). Although this might suggest that AWR, which recommended research on HAQ progression, is possible and worthwhile, the combined potential benefit of resolving uncertainty associated with natural history [both in HAQ and Psoriatic Arthritis Response Criteria (PsARC)] and treatment effect together is much greater than the 'sum of its parts'.[w] This suggests that both types of research could be conducted while etanercept continues to be approved but infliximab and adalimumab are not, that is, a possible OIR rather than reject for infliximab and adalimumab but AWR for etanercept.

Implications of between-scenario uncertainty

In *Point 3: Does more research seem worthwhile?*, *Alternative scenarios* the contribution that alternative scenarios might make to the overall expected consequences of uncertainty and therefore the potential gains from further evidence was considered and discussed. In situations in which more than one scenario might be regarded as credible there will be uncertainty *between* as well as *within* each of the scenarios. It was demonstrated that an assessment of the combined consequences of both sources of uncertainty requires 'weights' (probabilities) to be assigned to represent their credibility, which can then be applied directly to the simulated output from PSA (see *Appendix 11*, *Why averaging scenarios may be misleading*). However, the same analysis can also be used to identify the expected consequences of uncertainty associated with the alternative scenarios themselves, that is, what might be gained if evidence could immediately distinguish which scenario was 'true'. This can help to inform the assessment of what type of evidence might be needed and whether the research required to generate it is likely to be possible once a technology is approved for widespread NHS use.

For example, in the CLOP analysis, scenarios A and B (treatment effect was constant or differed by treatment duration, respectively) were associated with expected consequences of uncertainty of 5195 and 3969 QALYs, respectively. If both scenarios were regarded as equally likely, the overall expected consequences of uncertainty (combining consequences within and between scenarios) would be 4667 QALYs. However, the expected consequences of uncertainty associated with the two alternative scenarios themselves and what might be potentially gained if the uncertainty between them could be immediately resolved is relatively modest at 85 QALYs, that is, most of what might be gained from further evidence is associated with the parameters in *Table 24* rather than the alternative scenarios. This suggests that more evidence about overall relative effect on mortality is more important than resolving uncertainty about whether or not such an effect differs by treatment duration.

In the EECP case study, formal elicitation of the judgement of clinical experts about whether or not observed QALY gains at 12 months are likely to be sustained in subsequent years was undertaken. Because the uncertainty in these elicited values was incorporated into the analysis in the same way as for other parameters, the use of alternative scenarios was not necessary. However, in *Point 3: Does more research seem worthwhile?*, scenarios were used to illustrate the type of analysis that, without elicitation, might otherwise have been required. The scenarios included A, no QALY benefits beyond 12 months; B, benefits sustained for a patient's lifetime; and C, benefits sustained for 4 years. The results of elicitation implied probabilities of 0.243, 0.353 and 0.404 associated with each of these scenarios, respectively. Based on these 'weights' for each scenario the overall expected consequences of uncertainty (combining the consequences within and between scenarios) would be 14,146 QALYs. In this case the expected consequences of uncertainty between the scenarios (13,202 QALYs) are much greater than what might be potentially gained from resolving the uncertainty within each scenario (1765 QALYs). Therefore, unlike in the CLOP case study, most of what might be gained from further evidence about EECP (in the absence of formal elicitation) would be evidence that could help distinguish between the scenarios rather than the parameters associated with each.

Do the benefits of research exceed the costs?

The judgements made at points 5 and 6 of the checklist are critical because if the benefits of research are not judged to exceed the costs then no further assessment are required (unless there are significant irrecoverable costs; see *Table 32* in *Appendix 4*). If they are judged to exceed the costs and research can be conducted with approval then AWR would be appropriate. However, other sources of uncertainty need to be assessed first because they will influence the potential benefits of research and, even when research is not conducted, they will also influence the appropriate category of guidance when there are significant irrecoverable costs.

Point 5: Will other sources of uncertainty resolve over time?

The fifth point on the checklist requires an assessment of whether or not changes are likely to occur in the future that will influence the cost-effectiveness of the alternative technologies and the potential benefits of research, that is, at the following point in the algorithm:
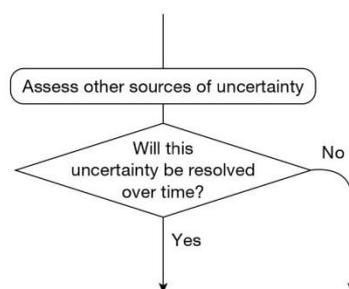
The judgement made at this point will influence the potential benefits of research and therefore subsequent decisions that lead directly to a particular category of guidance (see point 6 below). Even when research was not considered worthwhile (at point 3) the presence of other sources of uncertainty will determine whether or not significant irrecoverable costs are likely to influence the category of guidance. In some circumstances it can lead directly to guidance, that is, if there are no other sources of uncertainty even significant irrecoverable costs will have no influence and a technology that is expected to be cost-effective can be approved:

Assessment	1	2	3	4	5	6	7	Guidance
Pathway number { 29	Yes	Yes	No	–	No	–	–	Approve[12]

This assessment requires information about (1) changes in prices of the technology and its comparators, (2) the emergence of new technologies that might make existing ones obsolete or change their cost-effectiveness and (3) other relevant research reporting (see *Appendix 6, Impact of other sources of uncertainty*). A number of potential sources of information and evidence were examined to inform this assessment for each case study (see *Appendix 4* for the full details of the sources and searches conducted). However, many potentially useful sources were proprietary or public access was restricted, making it surprisingly difficult to inform these assessments with publically available information. When information and estimates were available they were often not directly relevant to a UK context.

Changes in the prices of the technology and its comparators

Changes in prices influence not only expected cost-effectiveness but also uncertainty and the potential benefits of research to future patients, for example if the price of a technology expected to be cost-effective is likely to fall significantly just before research reports the potential benefits will not be realised because approval of the technology will be less uncertain and there may be much less or little to gain from the results of the research. This assessment requires information about when major changes in prices are likely and some evidence about the likely extent of the changes. A major event in the life cycle of a pharmaceutical technology is the date at which the patent expires and cheaper generic versions of the brand become available. Although the date of patent expiry is, of course, known, it is surprisingly difficult to obtain the relevant date for particular products in the UK from publically available sources. Evidence of the extent to which the prices of generic versions are below the original brand price is also difficult to obtain and likely to differ by health-care system, type of technology, indication and time since patent expiry. Therefore, the estimate reported by the OFT that, on average, generic prices tend to be 25% of the original price was used in the subsequent analysis.[12]

At the time of TA80[122] the patent for CLOP was expected to expire 7 years later and subsequent analysis assumes that at that time equivalent generic prices will be 25% of the original price of CLOP at the time of TA80.[x] Although it was possible in the PsA case study to find patent expiry dates for etanercept (Enbrel®, Wyeth), infliximab (Remicade®, Schering-Plough) and adalimumab (Humira®, Abbott) in the USA (2012, 2014 and 2017, respectively), they were not available for the UK on the national patent database (Intellectual Property Office). It is even more difficult to locate patent information relevant to devices such as EECP because a device may only have a CE mark, which, unlike a patent, does not offer protection and can be renewed every 10 years. Any patent is likely to relate to some aspect of the device rather than the device itself. Although prices may change over time they can also be relatively stable but with incremental innovation of the original device. Again, this is likely to differ by health-care system, technology and indication. For these reasons future changes in prices are only quantitatively explored in the subsequent analysis of the CLOP case study in *Point 6: Are the benefits of research greater than the costs?*. There is a need to consider how access to the type of information required during NICE appraisal can be provided and how estimates of likely changes in prices relevant to the UK can be made readily available if these assessments are to be routinely made.

Entry of new technologies

The entry of a new technology will tend to change the relative cost-effectiveness of the alternatives, influencing how uncertain a decision to approve the original technology will be for future patients and the potential gains from research. For example, the entry of a new technology may make the existing technology that is expected to be cost-effective obsolete (no

longer the most cost-effective alternative) or may make it more cost-effective and decisions less uncertain when used in combination with the new technology. A number of potential sources of information were examined to identify new technologies relevant to the indications that were likely to become available. These included a variety of sources related to NICE topic selection, information about licence applications and clinical research in phases I, II and III as well as evidence of the probability that earlier phase research leads to entry (probability of successful licence) and the likely time of entry (time to launch from initiating phases I, II and III research). Again, this information and evidence is fragmented, and in some cases restricted, for example NIHR Horizon Scanning Centre. Nevertheless, the information that was available indicated that one new technology relevant to CLOP and one relevant to PsA might have been expected to enter. Information about these technologies was limited and so scenarios are used to explore the implication for CLOP and PsA in *Point 6: Are the benefits of research greater than the costs?*.

Other research reporting of the technology and its comparators

Research that is already under way, commissioned or likely to be undertaken, whether in the UK or elsewhere, is relevant for two reasons: first, if it is research based in the UK then guidance might impact on recruitment and the successful completion of this research (see *Point 7: Are the benefits of approval greater than the costs?*); second, when this research reports that there is a chance that it will change the estimates of cost-effectiveness and resolve some of the current uncertainties. In other words, there is little to be gained by recommending OIR or AWR if the uncertainty is likely to be resolved in the near future when other research reports. A number of potential sources of information were examined to identify clinical research under way at the time of the relevant appraisal, including national and international trial registries and other databases that report NHS-funded research and not just clinical trials (e.g. National Research Register and UK Clinical Research Network). Despite an assiduous search no records relevant to the case studies were identified. This may suggest that no other research was ongoing or expected for these comparators in these indications, or it may indicate that currently available sources are fragmented, incomplete and/or difficult to access.

Point 6: Are the benefits of research greater than the costs?

The sixth point on the checklist requires a reassessment of the potential benefits of conducting further research that were initially considered at point 3, and a judgement of whether the benefits of research are likely to exceed the costs, that is, at the following point in the algorithm:
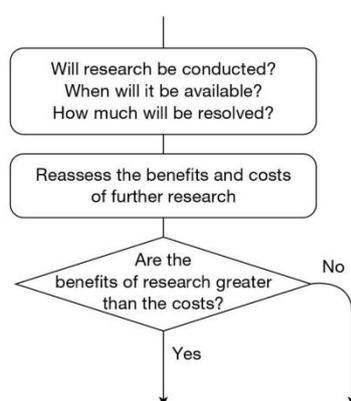
A judgement about whether or not the potential benefits of research identified earlier (see *Is further research required?*) will be realised requires an assessment of (1) whether or not the type of research that is required is likely to be conducted, (2) if conducted, when the results are likely

to be available, (3) how much uncertainty is likely to be resolved and (4) the likely impact of the other sources of uncertainty identified in *Point 5: Will other sources of uncertainty resolve over time?* on the longer-term benefits of research (see *Will the research be conducted?, How long until the research reports?, How much of the uncertainty will the research resolve?* and *How do other sources of uncertainty impact on the need for evidence?, Appendix 6*, respectively).

The decision at this point may not necessarily lead directly to guidance, for example when the benefits of research exceed the costs but research is not possible with approval or there are significant irrecoverable costs. Which category of guidance will ultimately be appropriate will depend on whether or not the benefits of approval are judged to exceed the costs, that is, point 7 of the checklist (see next section). However, in many other circumstances the decision at this point will lead directly to a particular category of guidance. These circumstances or pathways through the algorithm are detailed below (extracted from *Table 32* in *Appendix 4*):

Assessment	1	2	3	4	5	6	7	Guidance
1	Yes	No	Yes	Yes	Yes/no	Yes	–	AWR[1]
2	Yes	No	Yes	Yes	Yes/no	No	–	Approve[1]
5	Yes	No	Yes	No	Yes/no	No	–	Approve[3]
7	No	No	Yes	Yes	Yes/no	Yes	–	OIR[2]
8	No	No	Yes	Yes	Yes/no	No	–	Reject[1]
11	No	No	Yes	No	Yes/no	No	–	Reject[3]
19	Yes	Yes	Yes	Yes	No	No	–	Approve[6]
26	Yes	Yes	Yes	No	No	No	–	Approve[10]
30	No	Yes	Yes	Yes	Yes/no	Yes	–	OIR[7]
31	No	Yes	Yes	Yes	Yes/no	No	–	Reject[8]
34	No	Yes	Yes	No	Yes/no	No	–	Reject[10]

(Pathway number is indicated to the left, bracketing assessments 2 through 34.)

The expected consequences of uncertainty reported in *Point 3: Does more research seem worthwhile?* represented the NHE that could be gained over the lifetime of the technology if the uncertainty surrounding the decision based on expected cost-effectiveness could be immediately and completely resolved. This represents an upper bound on the potential benefits of research for a number of reasons: (1) research, although recommended, might not be commissioned and/or recruit and report, (2) any research will take some time to complete before results are available and (3) not all of the uncertainty is likely to be resolved. In addition, future events (identified in *Point 5: Will other sources of uncertainty resolve over time?*) might change the NHE expected to be gained by future patient populations. Finally, the expected benefits of research once properly reassessed must be compared with the likely costs.

Will the research be conducted?

Even if research is recommended in OIR or AWR it might not be undertaken by manufacturers or commissioned by research funders. Even if undertaken or commissioned there is no guarantee that research will be able to recruit or it may not complete for other reasons (see *Appendix 6, Will the research be conducted?*). The expected consequences of uncertainty for CLOP and EECP reported in *Point 3: Does more research seem worthwhile?* are illustrated in *Figures 13* and *14*, respectively, for a range of probabilities that research will be successfully undertaken. This indicates that the potential gains depend on a judgement of whether the research recommended as part of OIR or AWR will be successfully completed. This also illustrates that the cost of research (in this case considered to be either £1.5M or £10M) can be compared directly with the potential benefits by either (1) expressing the potential gains in population NHE as the equivalent NHS resources (i.e. the resources that would be required to generate the same NHE) or (2) expressing the cost of research in terms of the QALYs that could be gained elsewhere in the NHS by using the same resources to provide access to health care.
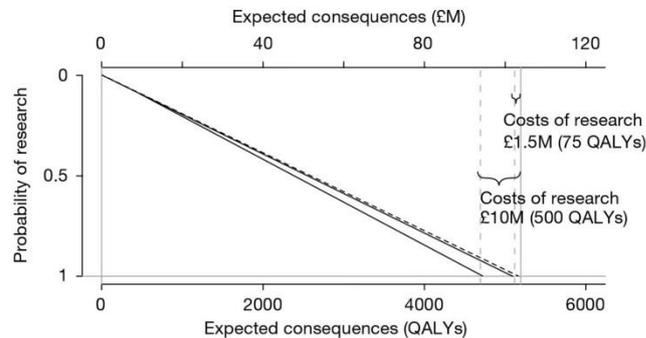
FIGURE 13 Expected potential benefits of research: CLOP case study.
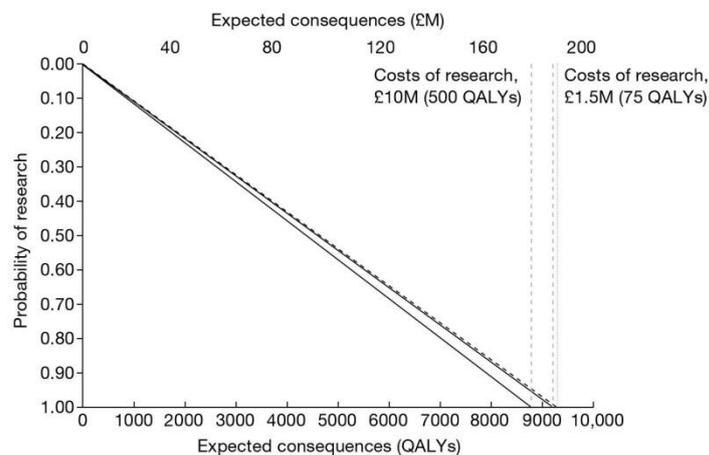
FIGURE 14 Expected potential benefits of research: EECP case study.

When will it be available?

Research, even if commissioned and successfully completed, will take time to commission, complete and report. Therefore, any assessment of the potential benefits should account for the fact that patient populations will not benefit from the results of research until they are available and change clinical practice.[y] Whether or not those patients who are prevalent while research is under way will be able to benefit from the results will depend on whether treatment decisions for presenting patients are irreversible or not (see *Point 2: Are there significant irrecoverable costs?*). If treatment decisions are irreversible, for example in the CLOP case study, it is only those patients incident after the research reports who will realise any of the potential benefits. In contrast, treatment decisions in EECP are not irreversible (it is a chronic condition) and so, although patients prevalent while research is undertaken will not benefit immediately, those who survive can benefit from the results once the research is completed. How long research might take to report will depend in part on the design (follow-up, sample size and end points), recruitment rate and size of the eligible patient population, as well as on how efficient the organisation and data collection might be. The potential value of research in CLOP and EECP over a range of possible time horizons is reported in *Figures 15* and *16*, respectively. In both cases the potential value of further research declines with increase in time to research reporting. This relationship gives some indication of the value of improving the timeliness of research through, for example,
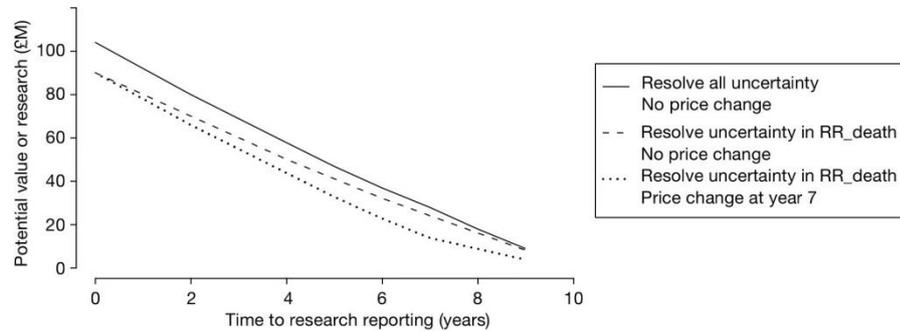
FIGURE 15 Potential value of research by time to report: CLOP case study.
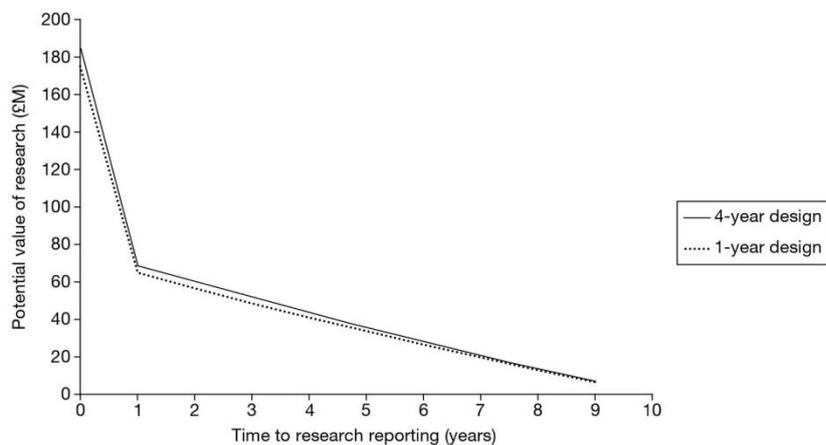
FIGURE 16 Potential benefits of research by time to report: EECP case study.

investment in research infrastructure or adopting a research design that, although offering less potential benefits, can be conducted more quickly. It also indicates the value of efforts to ensure that any revised NICE guidance following OIR or AWR is more quickly and more completely implemented.

How much will be resolved?

Most research will not inform all of the parameters that determine expected cost and QALYs but usually a subset of them. Therefore, the potential benefits of research that might be conducted will not be the total expected cost of uncertainty surrounding expected cost-effectiveness but some part of it (see *Appendix 6, What type of evidence is needed?*). In *Assesstment of uncertainty* the potential benefits of different types of evidence were assessed. In the CLOP case study it was additional evidence about relative treatment effects that was most valuable and therefore experimental research may be required to provide a more precise estimate of RR_death. The potential value of research that resolved only uncertainty about this relative treatment effect over a range of times to report is also represented in *Figure 15* (denoted by the legend 'resolve uncertainty in RR_death'). Although the potential value of research is lower at every time point, unless research is likely to take more than 8 years the potential value is still likely to exceed the costs. In the EECP case study there was most benefit to be gained by resolving the uncertainty

in the improvement in quality of life at 12 months, which in common with the CLOP case study is likely to require an experimental research design. *Figure 16* represents the potential benefits of alternative trial designs with either 1 or 4 years of follow-up (1-, 2-, 3- and 4-year follow-up designs were evaluated; see *Appendices 7–10*). Although longer follow-up offers greater potential benefits they are relatively small compared with the loss of potential value if longer follow-up delays the time until research findings are available, that is, a 4-year design will require a minimum of 4 years to complete. Again, as long as research reports before 8 years the potential benefits are likely to exceed the costs.

What is the impact of other sources of uncertainty?

In *Point 5: Will other sources of uncertainty resolve over time?* the information that was publically available identified that the patent for CLOP was due to expire 7 years after the appraisal. Based on the OFT estimate that generic prices tend to be 25% of the original brand price, this other source of uncertainty can be integrated quantitatively when estimating the potential value of research over the lifetime of the technology. In this case a significant fall in price in year 7 will substantially reduce the uncertainty surrounding 12 months of treatment with CLOP. Therefore, after year 7 there is less to be gained from resolving uncertainty and so the potential and value of research findings for patients incident after year 7 are reduced. The effect of a price change on research that could potentially resolve uncertainty in cost, natural history and utilities as well as relative effect is also illustrated in *Figure 15*. The potential value of the research is lower whenever the research reports because it includes the value to future as well as current patient populations. Nevertheless, even if research did not report until 7 years the potential value is likely to exceed the costs. The expected price reduction reduces the potential value of research at each time point, for example, from 174,519 QALYs when research is immediately available (see *Figure 17*) to 165,701 with a price change at year 7.[z]

There was some evidence of possible entry of a new technology (comparator) in the indication described in the CLOP case study. However, there was limited information on its characteristics. Therefore, two alternative but somewhat extreme scenarios are illustrated in *Figure 17*. In scenario A the new technology enters at year 5 and makes CLOP entirely obsolete, that is, not cost-effective and not uncertain (equivalent to a shorter technology time horizon of 5 years). At this point there is no value in the evidence generated by research about CLOP.[aa] In these circumstances research is likely to be worthwhile only if it reports quickly. In scenario B the new technology has a similar NHE to 12 months of treatment with CLOP[ab] and the uncertainty surrounding its expected cost-effectiveness is also similar. Now research about CLOP has more potential value in the future because it will also help resolve some of the uncertainty in the choice between CLOP and the new technology for patients who become incident after that time. Although there was no evidence of new technologies emerging in EECP, the same scenarios are explored as the development and launch of new devices are more difficult to identify in advance. The impact on the potential value of research is illustrated in *Figure 18*, demonstrating similar qualitative effects as for CLOP. In scenario A (EECP becomes obsolete) the potential benefits of further research about EECP are likely to exceed the costs only if the research reports quickly. Nevertheless, even in this extreme scenario the benefits of research with only 1 year of follow-up are likely to exceed the costs as long as it reports before 4 years.

The potential value of research presented in these figures, even after accounting for the type of evidence, follow-up and time until research reports, should still be regarded as an upper bound to the value that is likely to be realised by actual research for two reasons: (1) even well-designed research with large sample sizes will not fully resolve the uncertainty in the value that a parameter might take, especially in specific target populations and in particular (future) contexts and (2) insofar as implementation of NICE guidance is not 'perfect' and all clinical practice might not immediately respond to the results of research, the full benefits will be realised only over time or
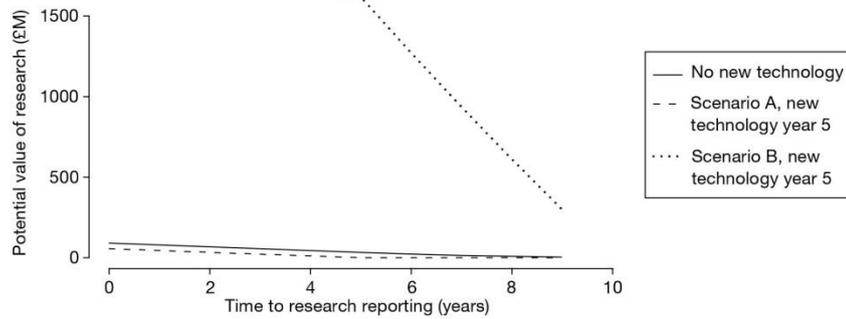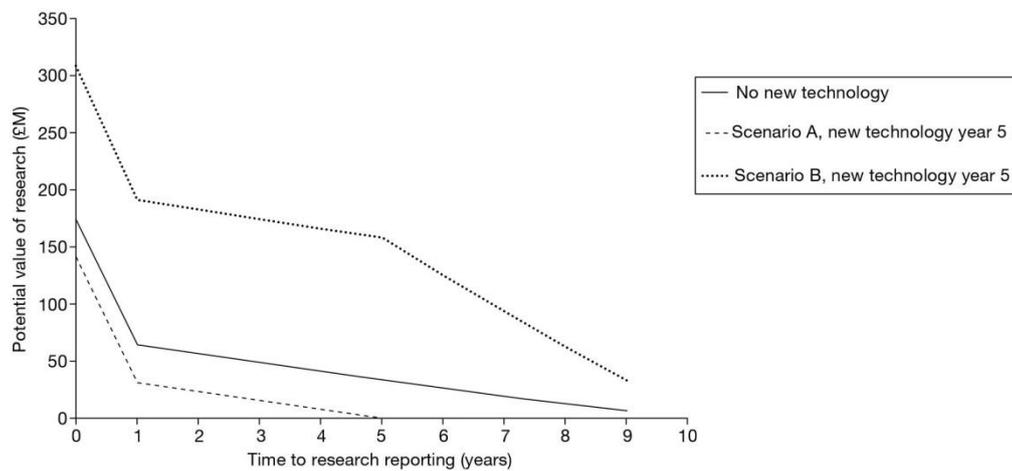
FIGURE 17 Potential value of research by time to report and other sources of uncertainty: CLOP case study.

FIGURE 18 Potential value of research by time to report and other sources of uncertainty: EECP case study. These potential values are based on a 1-year follow-up design.

with additional implementation efforts. For these reasons a judgement of whether or not benefits of research are likely to exceed the costs might be made conservatively, requiring evidence that even in pessimistic scenarios the research would still be worthwhile.

Point 7: Are the benefits of approval greater than the costs?

The seventh and final point on the checklist requires an assessment and comparison of the benefits and costs of early approval. The costs of approval are not financial ones but opportunity costs and will include the potential value of any research that may be forgone as a consequence, for example if the research needed cannot be conducted once the technology is approved for use. They will also include any costs that are irrecoverably committed by approval. As well as the capital costs of equipment and facilities (or training and learning), they will also include the irrecoverable opportunity costs of initially negative NHEs (if treatment decisions are not irreversible – see the discussion in *Chapter 3, Technologies with significant irrecoverable costs* and *Chapter 5, Point 1: Is it expected to be cost-effective?* and *Point 2: Are there significant irrecoverable costs?*). A judgement of whether or not the benefits of approval and early access for current

patients are likely to exceed the opportunity costs for future patients is required, that is, at the following point in the algorithm:
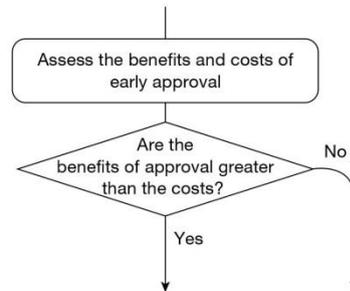
The decision at this point always leads directly to guidance, with all remaining possible pathways allocated to a particular type and category of guidance. These remaining (20) pathways through the algorithm are detailed below (extracted from *Table 32* in *Appendix 4*):

Assessment	1	2	3	4	5	6	7	Guidance
3	Yes	No	Yes	No	Yes/no	Yes	Yes	Approve[2]
4	Yes	No	Yes	No	Yes/no	Yes	No	OIR[1]
9	No	No	Yes	No	Yes/no	Yes	Yes	AWR[2]
10	No	No	Yes	No	Yes/no	Yes	No	Reject[2]
13	Yes	Yes	Yes	Yes	Yes	Yes	Yes	AWR[3]
14	Yes	Yes	Yes	Yes	Yes	Yes	No	OIR[3]
15	Yes	Yes	Yes	Yes	Yes	No	Yes	Approve[5]
16	Yes	Yes	Yes	Yes	Yes	No	No	Reject[5]
17	Yes	Yes	Yes	Yes	No	Yes	Yes	AWR[4]
18	Yes	Yes	Yes	Yes	No	Yes	No	OIR[4]
20	Yes	Yes	Yes	No	Yes	Yes	Yes	Approve[7]
21	Yes	Yes	Yes	No	Yes	Yes	No	OIR[5]
22	Yes	Yes	Yes	No	Yes	No	Yes	Approve[8]
23	Yes	Yes	Yes	No	Yes	No	No	Reject[6]
24	Yes	Yes	Yes	No	No	Yes	Yes	Approve[9]
25	Yes	Yes	Yes	No	No	Yes	No	OIR[6]
27	Yes	Yes	No	–	Yes	–	Yes	Approve[11]
28	Yes	Yes	No	–	Yes	–	No	Reject[7]
32	No	Yes	Yes	No	Yes/No	Yes	Yes	AWR[5]
33	No	Yes	Yes	No	Yes/No	Yes	No	Reject[9]

Pathway number

Significant irrecoverable costs

Technologies without significant irrecoverable costs

Only 4 of the 20 possible pathways illustrated above are associated with technologies without significant irrecoverable costs. In these four pathways either (1) research was not considered possible *with* approval for those expected to be cost-effective (i.e. 'Approve[2]' or 'OIR[1]') or (2) research was not possible *without* approval for those not expected to be cost-effective (i.e. 'AWR[2]' or 'Reject[2]'). The CLOP case study provides an example of the former. It is research that would provide more precise estimates of the relative effect of CLOP and of shorter treatment durations that is potentially valuable (see *Point 4: Is research possible without approval?*). As a consequence, the type of experimental design that is likely to be needed is unlikely to be possible if 12 months of treatment with CLOP is already approved for widespread NHS use. Although treatment with CLOP does commit initially negative NHEs that are irrecoverable, these should not be regarded

as significant as the treatment decision for a presenting patient is irreversible in relevant time frames (see *Point 2: Are there significant irrecoverable costs?*). Therefore, AWR may not be possible and so the benefits of early access to 12 months of treatment with CLOP (approval) must be compared with the potential value of OIR.

'Only in research' is more likely to offer greater expected NHEs than approve if the research can be conducted quickly and report sooner, as fewer patients forgo their access to CLOP and more can have their treatment choice informed by the research findings. This is illustrated in *Figure 19*, which reports the difference between approve and OIR in population NHEs over a range of times for when the research recommended in OIR might report. This takes account of both the expected changes in price at year 7 and research costs of £10M. It shows that OIR will be appropriate only if the research reports within 3 years of appraisal (T* = 3) because beyond this time the NHEs forgone by withholding access to CLOP will exceed the potential gains to future patients.

The trade-off between NHEs for current patients and NHEs for future patients that lies behind *Figure 19* is illustrated in *Figure 20* using undiscounted values for ease of exposition. It illustrates the (per-period) population NHEs of approval and OIR if the research recommended as part of OIR reports at year 3. At this point the initial losses of NHEs caused by restricting access to CLOP (area A) start to be offset by the potential gains from the research findings (area B). The price change at year 7 increases the NHEs of approval (i.e. CLOP is more cost-effective) but on balance reduces the NHEs of OIR. In other words, because CLOP is more cost-effective and offers greater NHEs the evidence generated by the research is less valuable because the choice of treatment and duration is less uncertain (see *Point 5: Will other sources of uncertainty resolve over time?* and *Point 6: Are the benefits of research greater than the costs?*). With research reporting at 3 years the initial losses of OIR (area A) are just offset by the later gains (area B) so T* = 3. If research reported earlier than 3 years (area A < area B) OIR would be appropriate but if it reported later than 3 years (area A > area B) approve would be more appropriate.

However, there is no guarantee that the research recommended as part of OIR guidance will be conducted by manufacturers or commissioned by research funders. Even if it is, it is not certain that it will be successfully completed (see discussion in *Point 6: Are the benefits of research greater than the costs?*). Therefore, the probability that research will report at a particular time also needs to be considered. The implications of considering whether the recommended research will be conducted and when it might report are illustrated in *Figure 21*, which presents a boundary for
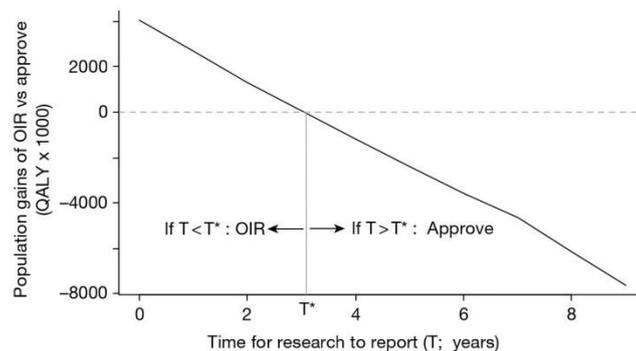
FIGURE 19 Population NHEs of approve and OIR for time to research reporting: CLOP case study. T* is the time to research reporting at which the NHEs of OIR and approve are equal. That is, for a time to report of less than T* OIR offers higher NHEs than approve and for a time to report of more than T* approve offers higher NHEs than OIR.
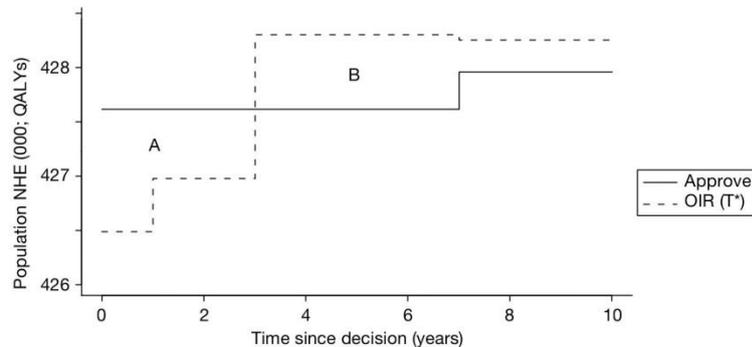
FIGURE 20 Population NHEs of approve and OIR at T*: CLOP case study.
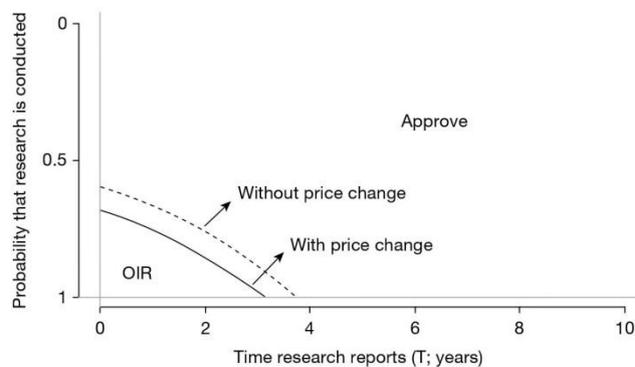
FIGURE 21 An OIR or approve boundary: CLOP case study.

when OIR might be appropriate or when approval should be granted. For example, if research is certain to report but will take 4 years, or when it will take only 1 year but has only a 50% chance of reporting, OIR would not be appropriate and 12 months of treatment with CLOP should be approved, that is, points that fall to the north-east of the boundary. Points to the south-west of the boundary indicate that OIR might be appropriate (see *Appendix 6, When might it be commissioned and report*).

However, the estimates of the potential value of additional evidence on which these boundaries are based are still likely to overestimate the value that will be realised by research (see discussion in *Point 6: Are the benefits of research greater than the costs?*); they represent a necessary condition for OIR. Therefore, OIR guidance should require a conservative judgement that the point is almost certain to be below the boundary rather than on balance close to it. For the same reason, points anywhere above the boundary represent a sufficient condition for approval. The boundary when the change in price is included is to the south-west, reflecting the lower potential value of research, and OIR guidance, once CLOP becomes more cost-effective. In this case it seems unlikely that the type of research required could report quickly enough and with sufficient confidence that OIR would be appropriate. Therefore, these assessments would support a judgement that the benefits of approval are likely to exceed the opportunity costs, and 'Approve[2]' (pathway 3 for CLOP in *Table 32, Appendix 4*) would seem more appropriate.

The assessments that have been undertaken for CLOP can be brought together to consider (1) what would be the value of being able to conduct research while CLOP is approved (value of AWR) and (2) what would be the value of making evidence that is needed by the NHS available at launch. These questions can be informed by the results already presented elsewhere but these results are also reported together in *Table 25*.

The difference in NHEs between AWR (if it was possible) and the next best feasible policy, which is OIR when T < T* and approve when T > T*, is £30M and £4M, respectively. This difference represents the value to the NHS of being able to conduct research while CLOP is approved for use, for example to inform whether investment in better data collection, registries or information systems is worthwhile that might make this possible.[ac] The difference in population NHEs if all uncertainty had been resolved before appraisal (at launch) and the next best available policy, OIR for T < T* and approve for T > T*, is £54M and £80M, respectively. This difference represents the value to the NHS of having access to the evidence needed at launch. This can also directly inform policies that might make better and more relevant evidence available.

It is also possible to consider the commercial as well as NHS value in each of the cells of this table. The value of early evidence at launch can then be considered from the perspective of the manufacturer (the expected revenue streams), taking account of prices (see discussion of price in *Chapter 3*, *Changes in effective prices* and *Incentives for evaluative research*) and expected volumes over the remaining patent life and technology time horizon. Together with estimates of the costs of conducting research by manufacturers or through public funding, this assessment might inform when manufacturers might be expected to conduct the research needed (e.g. high commercial value that exceeds the cost to manufacturers) or when the NHS might be expected to undertake it (e.g. low commercial value but high potential value to the NHS that exceeds the costs to the NHS). When the research is worthwhile from both a commercial and a NHS perspective the question of who should conduct, or pay for, the research might be informed by which sector has a comparative advantage, that is, which can gain the most net social value. Of course, the value to the NHS and to manufacturers will depend, to a large extent, on what type of flexible pricing arrangements and value-based pricing schemes might be in place. The question will also turn on how any agreements can be made and incentive consistent contracts written and enforced.

Technologies with significant irrecoverable costs

Most of the possible pathways illustrated above are associated with technologies with significant irrecoverable costs (16 out of the remaining 20). This is because, even when research is possible with approval (or even when not needed), the impact of committing irrecoverable costs

TABLE 25 Population NHEs over the technology time horizon for different policies: CLOP case study

NHE for time to research reporting	Approve	OIR	AWR*	Reject	Value of AWR	Uncertainty resolved at launch	Value of evidence at launch
NHEs in QALYs							
T < T* (T = 2)	3,680,187	3,681,480	3,682,995	3,671,660	1515	3,684,181	2701
T > T* (T = 7)	3,680,187	3,675,487	3,680,362	3,671,660	175	3,684,181	3994
NHEs in £M							
T < T* (T = 2)	73,604	73,630	73,660	73,433	30	73,684	54
T > T* (T = 7)	73,604	73,510	73,607	73,433	4	73,684	80

AWR*, the expected population NHEs if AWR was to become a possibility; T, time to research reporting; T*, time to research reporting at which the NHEs of OIR and approve are equal.

through AWR (or approval) must be considered and so OIR (or reject) remains a possibility (see *Appendix 6*, *What is the impact of irrecoverable costs?*). The EECP case study provides an example of this, in which research that would provide more precise estimates of the effect of treatment on quality of life accounts for all of the potential value (see *Point 4: Is research possible with approval?*). EECP does commit both capital costs associated with long-lived equipment and initially negative per-patient NHEs. Unlike in the CLOP case study these irrecoverable opportunity costs at a patient level are significant because treatment choice for a presenting patient is not irreversible over relevant time frames (see *Point 2: Are there significant irrecoverable costs?*). As a consequence, even if research is possible with approval it is not clear that AWR would be appropriate, because OIR avoids the commitment of irrecoverable costs until research finding are available and a more informed decision can be made.[ad]

Research is possible with approval

Even when research is possible with approval OIR offers greater expected NHEs than AWR as long as research reports before 9 years (*Figure 22*). This is because the consequences (losses of population NHE) of committing both aspects of irrecoverable costs through AWR are greater than the NHEs forgone by restricting access to EECP through OIR. The costs of research have not been included because they are incurred with both AWR and OIR guidance.[ae]

As previously for the CLOP case study, there is no guarantee that the research recommended as part of OIR or AWR guidance will be conducted and research report. A boundary for when OIR rather than AWR might be appropriate is illustrated in *Figure 23* for four research designs with differing times of follow-up. A 2-year follow-up will generate evidence with the lowest potential value (so the boundary is to the south-west) but it is likely to report sooner. Therefore, OIR might be appropriate even if the probability that the research will be conducted and report is relatively low. In this case it seems likely that the type of research required could report quickly enough and with sufficient confidence that OIR would be appropriate even though the research could be conducted while EECP is approved. Therefore, these assessments would support a judgement that the benefits of approval (through AWR) are unlikely to exceed the opportunity costs (the NHEs of OIR) and so 'OIR⁴' (pathway 18 in *Table 32*, *Appendix 4*) rather than 'AWR⁴' (pathway 17) would seem more appropriate.
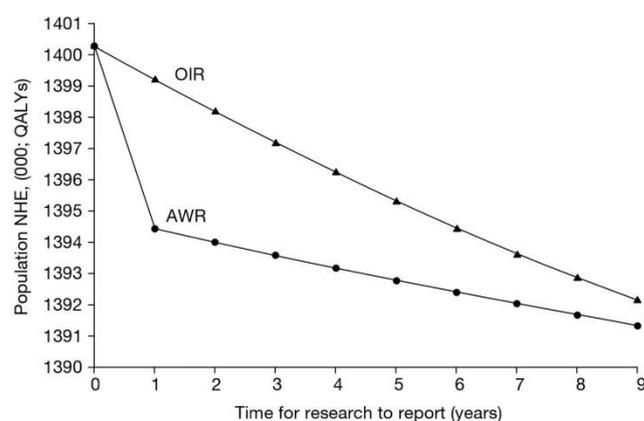
FIGURE 22 Population NHEs of AWR and OIR for time to research reporting: EECP case study.
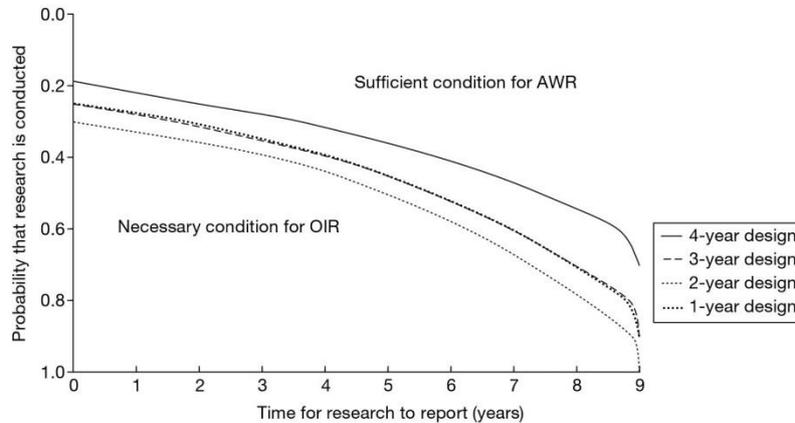
FIGURE 23 An OIR or AWR boundary: EECP case study.

Research is not possible with approval

For the general reasons discussed in *Chapter 3* and those specific to EECP discussed in *Point 4: Is research possible with approval?*, the type of experimental research required to robustly estimate the effect of EECP on quality of life is unlikely to be possible once it is approved and in widespread use. Now approval (now through approve rather than AWR) not only commits the type of irrecoverable costs discussed above but also means that the potential value of evidence to future patients must be forgone. This is reflected in *Figure 24* in which the difference between OIR and approve is always greater than the difference between OIR and AWR in *Figure 22*. It suggests that as long as the cost of the research exceeds the difference between OIR and approve, when it is expected to report, OIR rather than approve would be appropriate. This is also reflected in the boundaries for OIR and approve reported in *Figure 25*. These boundaries are always to the north-east of the OIR/AWR boundaries reported in *Figure 23*, again reflecting the fact that approval not only commits irrecoverable costs but also forgoes the potential value of evidence that might have been generated through an OIR recommendation. These assessments would support a judgement that the benefits of approval are unlikely to exceed the opportunity costs (the NHEs of OIR) and so 'OIR[6]' (pathway 25 in *Table 32*, *Appendix 4*) rather than 'Approve[9]' (pathway 24) would be more appropriate.

As with the CLOP case study, the assessments that have been undertaken for EECP are bought together in *Table 26* and can help inform the same policy questions: (1) what would be the value of being able to conduct research while EECP is approved? and (2) what would be the value of making the evidence that is needed by the NHS available at launch? In this case, because of the irrecoverable costs associated with EECP, there is no value to the NHS of being able to conduct research while EECP is approved for use. In fact, these figures are negative, indicating that even if AWR was possible it would not be appropriate. However, as in the CLOP case study, there is value to the NHS of having the evidence needed before appraisal. This value, expressed in the equivalent NHS resources, depends on how long it would otherwise have taken for an OIR recommendation to deliver the same evidence, for example £62M if 3 years and £134M if 7 years. As with the CLOP case study, these assessments can also inform policies that might make better and more relevant evidence available and answer the questions of how to provide the evidence needed at the right time and who might contribute most to providing this evidence.
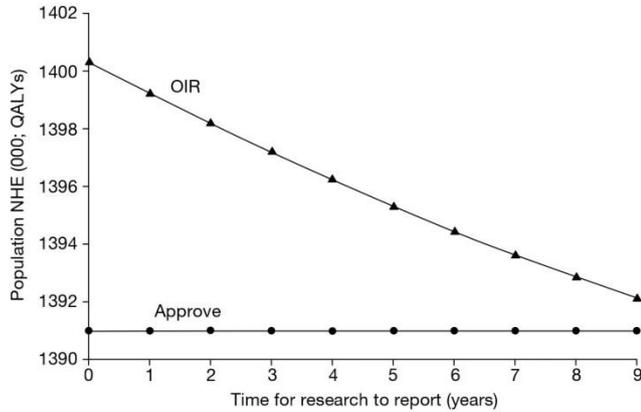
FIGURE 24 Population NHEs of approve and OIR for time to research reporting: EECP case study.
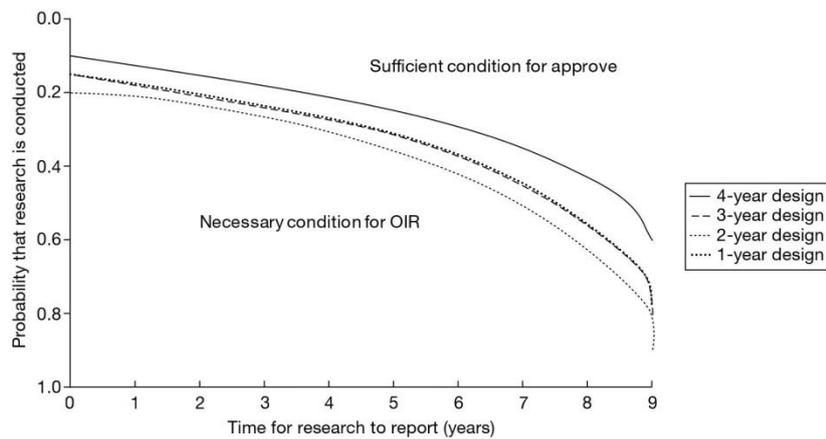
FIGURE 25 An OIR or approve boundary: EECP case study.

TABLE 26 Population NHEs over the technology time horizon for different policies: EECP case study

NHE for time to research reporting	Approve	OIR	AWR*	Reject	Value of AWR	Uncertainty resolved at launch	Value of evidence at launch
NHEs in QALYs							
T = 3	1,391,001	1,397,192	1,393,578	1,389,596	−3614	1,400,288	3096
T = 7	1,391,001	1,393,608	1,392,030	1,389,596	−1578	1,400,288	6680
NHEs in £M							
T = 3	27,820	27,944	27,872	27,792	−72	28,006	62
T = 7	27,820	27,872	27,841	27,792	−32	28,006	134

AWR*, the expected population NHEs if AWR was to become a possibility; T, time to research reporting.

Chapter 6

Implications for policy, process and methods

This final chapter of the main report draws together some of the implications that a more explicit assessment of OIR and AWR might have for policy (e.g. NICE guidance and drug pricing), the process of appraisal (e.g. greater involvement of research commissioners) and methods of appraisal (e.g. should additional information, evidence and analysis be required). Within each of the broad topics discussed below, we try to distinguish between policy issues directly relevant to the NICE remit (at a 'higher level' than the process and methods of technology appraisal); those that might be most relevant to other bodies (e.g. Department of Health, Department for Business, Innovation & Skills) and stakeholders; and specific issues for the process and the methods of technology appraisal. It constitutes a brief discussion of possibilities that NICE and others might choose to take forward. How a consideration of uncertainty and the need for evidence might influence value-based pricing and the impact of patient access schemes on OIR and AWR guidance is also considered. This discussion draws on the feedback provided at both workshops, which included a wide range of relevant stakeholders, including members of NICE and its Advisory Committees (including lay members and other NICE programmes), patient representatives, manufacturers, and research and NHS commissioners, as well as relevant academics. The feedback from both workshops has been summarised and is available online (see www.york.ac.uk/che/research/teehta/workshops/only-in-research-workshop//).

Key principles and assessments needed

The key principles and assessments needed fall into four broad areas:

1. expected cost-effectiveness and population NHEs (including benefits, harms and NHS costs)
2. the need for evidence and whether the type of research required can be conducted once a technology is approved for widespread use
3. whether or not there are sources of uncertainty that cannot be resolved by research but only over time
4. whether or not there are significant (opportunity) costs that will be committed and cannot be recovered once the technology is approved.

Guidance will depend on the combined effect of all of these assessments because they influence whether the benefits of research are likely to exceed the costs and whether any benefits of early approval are greater than the costs of withholding approval until additional research is conducted or other sources of uncertainty are resolved. There was general consensus that the key principles represented by a sequence of assessments and judgements (see *Chapter 3* and the algorithm in *Appendix 2*) were the relevant ones, complete and in an appropriate order. In response to feedback from the first workshop this sequence of assessment and judgement was summarised as a simple 7-point checklist (see *Chapter 5, A checklist of assessment*) that could be considered by the TAR groups and Appraisal Committees as well as manufacturers during appraisal.

Other NICE programmes

The principles outlined in *Chapter 3* and the checklist do not presuppose how the assessments required might be informed and judgements made. Distinguishing principles from methods of

analysis in this way meant that there was a general consensus that the principles and the checklist might be useful in other NICE programmes, whilst recognising that how the assessment might be made is likely to differ. It was recognised that some amendments might be required when cost-effectiveness is not a consideration (e.g. in interventional procedures) or when only technologies that are expected to be cost saving to the NHS are considered (e.g. the Medical Technologies Evaluation Programme). Similarly, the complexity of multiple outcomes in public health and the greater scope and complexity of decision problems in clinical guidelines offer greater challenges for quantitative assessment but do not change the key principles and considerations. The relative paucity of evidence and the speed of innovation in devices are likely to influence how the assessment might be made. On the other hand, it was considered by some workshop participants that the more iterative process of appraisal in some other NICE programmes (e.g. clinical guidelines) might make it easier to integrate the assessments required in the checklist. Further research, using case studies from these other NICE programmes, would be required to identify how best the assessments might be informed given the programme-specific challenges, the existing methods of appraisal and the resource and time constraints of their current processes.

The checklist was subsequently applied to a series of four case studies (three pharmaceuticals from the Technology Appraisal Programme and one device from the HTA programme; see *Chapter 5, Introduction to the case studies*). These applications allow an examination of how each of the assessments might be made based on the type of evidence and analysis currently provided in NICE technology appraisal and how they might be better informed with a range of additional information and/or analyses. The possible implications for process and methods of appraisal are discussed below.

The scope of 'only in research' and 'approval with research' recommendations

Categories and type of guidance

Each sequence of assessment and decision leads to different categories and 'types' of guidance for technologies with differing characteristics, indications and target populations, that is, each sequence of 'yes' or 'no' judgements in the checklist defines a single pathway leading to a particular type and category of guidance (see *Table 32, Appendix 4*). The different types of apparently similar guidance illustrate how the same category of guidance (e.g. approve, AWR, OIR or reject) might be arrived at in different ways, helping to identify the particular combinations of considerations that might underpin guidance, contributing to the transparency of the appraisal process. There was a general view that the application of the checklist would improve transparency in communicating the considerations that underpin guidance but that this alone would not be sufficient, especially in situations in which OIR guidance was made for a technology that was, on balance, expected to be cost-effective. Evidence of how, not just what, assessments and judgements were made would be required (see *What additional information and analysis might be required?*).

These principles suggest that the categories of guidance available to NICE have wider application than is reflected in previous guidance (see the review of NICE guidance in *Chapter 4*). For example, there are five different types of OIR guidance that may be appropriate when a technology is expected to be cost-effective (see *Table 1*). Indeed, OIR may be appropriate even when research is possible with approval if there are significant irrecoverable costs. AWR can be considered only when research is possible with approval but reject remains a possibility even for a cost-effective technology if there are irrecoverable costs. Therefore, the full range of categories of guidance (OIR and reject as well as AWR and approve) ought to be considered for technologies that, on the balance of existing evidence and current prices, are expected to be cost-effective. It

is only approval that can be ruled out if a technology is not expected to be cost-effective, that is, cost-effectiveness is necessary but not sufficient for approval but lack of cost-effectiveness is neither necessary nor sufficient for rejection.

Scope of 'only in research' recommendations

Importantly, which category of guidance will be appropriate depends only partly on an assessment of expected cost-effectiveness and hence this assessment should be regarded only as an initial step in formulating guidance. Guidance will depend on a number of other key assessments, which include (1) the need for evidence, (2) whether or not the type of research required is possible with approval, (3) the expected longer-term benefits and costs of the type of research likely to be conducted, (4) the impact of other sources of uncertainty that will resolve over time and (5) the significance of any irrecoverable costs.

In general, it was accepted that as well as AWR for technologies expected to be cost-effective and OIR for those not, there are other important circumstances when OIR should be considered. In particular, for technologies expected to be cost-effective, OIR rather than approve may be appropriate when research *is not possible* with approval and OIR or even reject rather than AWR or approve may be appropriate even if research *is possible* with approval when there are significant irrecoverable costs. The ethical implications of such decisions were set out in *Chapter 3* (see *Social value judgements and ethical principles*) and are discussed below.

Implications of patient access schemes and value-based pricing

'Only in research', 'approval with research' and effective price

Any change in the effective price of the technology, either through patient access schemes (which offer some form of discount that reduces NHS costs) or through direct price changes (possibly negotiated though a future value-based pricing scheme), will affect the key assessments leading to different categories of guidance. For example, provisional OIR guidance for a technology that is expected to be cost-effective might be revised to approve with a sufficient price reduction because the benefits of early approval will be greater and uncertainty about its cost-effectiveness and the value of additional evidence will tend to be lower. Similarly, AWR might be revised to approve if the consequences of uncertainty and what might be gained from additional evidence is sufficiently reduced. Equally, provisional guidance to reject a technology that is not expected to be cost-effective might be revised to OIR if the reduction in price (although not sufficient to make it cost-effective) made a reject decision more uncertain and hence research worthwhile.

Therefore, consideration of the effect of price changes on OIR and AWR is needed when assessing the potential impact of patient access schemes and more direct price negotiation through value-based pricing. However, there are limits to the effects of price reductions as even at a zero price the technology might not be cost-effective and/or further research may still be required if there is a lack of confidence that the technology is effective (harms may not be compensated for by benefits) and/or it imposes other non-acquisition costs on the NHS.

Implications for value-based pricing

The price at which the technology would just be expected to be cost-effective is commonly regarded as the value-based price for the technology. It is the maximum price that the NHS can afford to pay without imposing negative health effects. This describes the threshold price below which approve rather than reject would be appropriate when (1) OIR or AWR guidance is not available to the decision-maker, (2) there is no uncertainty over cost-effectiveness or (3) the research, if needed, can be conducted with approval and there are no irrecoverable costs. In all other circumstances there are a number of other value-based prices, each representing the

threshold price below which guidance would change. Importantly, once uncertainty and the need for evidence as well as the impact of irrecoverable costs are recognised, the threshold price that would lead to approval will always be lower than or equal to a single value-based price based on expected cost-effectiveness alone, that is, disregarding uncertainty over costs and effects.

For example, when research could be conducted without approval but not with it, there will be a price threshold above which reject rather than OIR would be appropriate and a lower price threshold below which approve rather than OIR would be appropriate. This approve/OIR threshold price will always be lower than a value-based price based on expected cost-effectiveness. If a technology also imposes significant irrecoverable costs then this threshold price will be lower still. There may also be up to three threshold prices as all four categories of guidance might become relevant at different prices.

Value-based pricing and irrecoverable costs

Health technologies with patent protection are more likely to be priced close to the point at which the ICER is close to or equal to the threshold, that is, expected incremental NHEs are close to zero. A value-based pricing scheme would formalise these existing incentives so technologies will tend to impose greater irrecoverable opportunity costs (initially negative NHEs). A technology that is only just expected to be cost-effective will not break-even until close to the end of the patient time horizon and not for much longer for the population of patients likely to benefit from its use. Therefore, those technologies already priced close to the threshold, and all new technologies considered in a value-based pricing scheme, will tend to increase the scale of irrecoverable costs committed by approval. If these irrecoverable costs are significant (because treatment decisions are not irreversible) then the price threshold for unrestricted access (i.e. approve or AWR) will be lower than the price at which the technology is just expected to be cost-effective, even if research can be conducted with approval (see *Chapter 3, Issues specific to 'only in research' recommendations*). Further research is required to establish how these considerations could be integrated within a practical value-based pricing scheme. In addition, irrecoverable opportunity costs are commonly associated with many health technologies that offer future benefits following treatment. The significance of these types of irrecoverable costs is not widely recognised and further research to demonstrate their potential impact more generally is needed.

Retaining 'only in research' as part of value-based pricing

Even in circumstances in which price negotiation becomes possible alongside NICE appraisal, it will be important to retain OIR and AWR as available categories of guidance for two reasons. First, there is no guarantee that manufacturers will always agree to the lower price below which approval rather than OIR or AWR would be appropriate. For example, manufacturers will be unwilling to agree prices that are below marginal production costs. They will also be unwilling to lower UK prices if they can be referenced by other countries and have an impact on global revenues. Second, and possibly more importantly, there may be many circumstances when there is no effective price reduction that would make approval appropriate. For example, reject or OIR guidance may still be appropriate even if the effective price of a technology is zero if there is substantial uncertainty about its effectiveness (and/or potential for harms) or if the technology imposes other non-acquisition costs on the NHS.

Incentives for evaluative research

Flexible pricing agreements

Consideration of OIR and AWR recommendations provides a link between uncertainty, evidence and price which might appropriately align incentives for manufacturers conducting the type of evaluative research that would be most valuable for the NHS. The current PPRS provides for

flexible pricing agreements in which price is revised once the research reports and the results are known, increasing prices if the evidence suggests that benefits were originally underestimated or reducing prices if benefits were overestimated. This means that manufacturers retain an incentive to conduct further evaluative research if they believe that there are additional benefits that could not be evidenced at launch. Publically funded evaluative research, however, will still be required when these incentives are insufficient and especially in those cases in which the original evidence is likely to have overestimated the benefits or underestimated the potential for harm. However, it should be noted that linking effective prices to the results of publically funded research means that the NHS will benefit (realise the value of evidence) only if the results lead to a lower price or more restrictive guidance because the technology is found not to be cost-effective (thus avoiding the losses associated with negative NHEs). Manufacturers will, however, be able to appropriate the value of evidence through higher prices when it suggests that NHEs were originally underestimated. Even under current arrangements this value can be appropriated when the technology is reappraised by NICE. For example, patient access schemes, which offer an effective discount to the NHS, can be withdrawn or approval might be made less restrictive when a technology is reappraised following further research.

Incentives and early advice

It is important that policy provides (or at least does not undermine) appropriate incentives for manufacturers to conduct the type of research needed to support NICE guidance at launch. The use of OIR and AWR guidance, and its link to price, provides a clear signal and an incentive for manufacturers to ensure that the type of evidence that cannot be acquired following approval is available and sufficient at launch (e.g. relative effectiveness and subtle but important differences in side effect profiles). Therefore, a predictable OIR and AWR policy signals what type of evidence is likely to be most important at an early stage. It offers manufacturers a choice, to (1) accept OIR guidance at a higher price but restricted volume, (2) reduce the effective price to achieve approval, or AWR when that is possible, or (3) conduct the evaluative research at an earlier stage so that additional evidence is not required at launch.

Other things being equal, those new technologies that are supported by more, better-quality and relevant evidence will be more likely to be approved and at higher prices than those that are not. Therefore, greater consideration of OIR and AWR guidance will tend to reward those manufacturers who have invested in good-quality and relevant evidence with earlier approval of their technology. It was also suggested that the checklist of assessments could be used within the NICE Scientific Advice Programme, providing advice to manufacturers at an early stage before technology appraisal.

Who should conduct 'approval with research' and 'only in research' research?

Consideration of how the NHS and manufacturers are likely to share the value of evidence might inform whether manufacturers should be expected to conduct the research specified in AWR or OIR guidance. Alternatively, manufacturers might be expected to make some contribution to the costs of publically funded research that may ultimately benefit their product. As well as an assessment of how the value of additional evidence is likely to be shared between the NHS and the manufacturers, it was widely recognised that two other issues need to be considered. First, the resource constraints on publically funded research may mean that other research priorities (often without commercial interest) may be more valuable to the NHS (see *How should the assessments be undertaken?*). Second, the success of AWR recommendations when manufacturers are asked to conduct the research will depend on whether or not NICE and/or the Department of Health are able to establish incentive-consistent contractual arrangements as part of an AWR recommendation, that is, arrangements that can be monitored and enforced with credible penalties to ensure that agreed research is conducted and in the way intended. It was widely recognised that, at present, NICE does not have a credible mechanism to ensure that the type

of research recommended in AWR guidance would actually be undertaken by manufacturers. Removing approval of a technology simply because recommended research has not been conducted was not considered an ethical or a credible threat. Further research could consider what type of feasible contractual arrangements would offer credible threats that would make AWR more likely to be successful. This work could include more detailed examination of linking an AWR agreement to price penalties within a value-based pricing scheme.

Although OIR provides a greater incentive to undertake research, it was recognised that there may be circumstances when manufacturers would nonetheless choose not to undertake it, that is, accepting an effective reject. If the research is also not a sufficient priority to secure public funding, then approval rather than OIR (an effective reject in these circumstances) would be appropriate only at the lower approve/OIR price threshold. Many workshop participants took the view that it would be better if all AWR and OIR research was publically funded rather than undertaken by manufacturers to ensure wide availability of research findings and provide confidence that commercial interests do not influence its design, conduct and reporting. It was suggested that the costs of publically funded research might be recovered directly from manufacturers or indirectly through other price discounts. Such arrangements might be mutually beneficial in some circumstances (e.g. if the costs of publically funded research are lower), while allowing wider access to the data generated and more transparency and accountability in the conduct of the research.

Further research could examine the relative costs of conducting this type of evaluative research and explore feasible mechanisms for recovering the costs of publically funded research from manufacturers. It should also include an examination of how the value of research is likely to be shared between manufacturers and the NHS in different circumstances and under different price renegotiation rules. The pay-offs for the different policies reported in *Tables 25* and *26* are from an NHS perspective only. It would be interesting to estimate the commercial pay-offs associated with each, which could start to inform when manufacturers or public bodies might be expected to conduct or pay for the research recommended as part of AWR or OIR guidance.

Value of 'approval with research' and evidence at launch

The assessments that need to be made (especially in *Chapter 5, Point 6: Are the benefits of research greater than the costs?* and *Point 7: Are the benefits of approval greater than the costs?*) can be used to consider (1) what would be the value of being able to conduct research while a technology is approved (value of AWR), (2) what would be the value of making evidence that is needed by the NHS available at launch and (3) what is the value of being able to acquire evidence more quickly.

The difference in population NHEs between AWR and the next best feasible policy represents the value to the NHS of being able to conduct research while a technology is approved for use. This can inform investment decisions in better data collection, registries or information systems that might make AWR possible. Importantly, this value will differ by technology and will depend on the scale and significance of irrecoverable costs (if they are sufficiently high there will be no value in AWR – see *Chapter 5, Technologies with significant irrecoverable costs*). The difference in population NHEs between resolving all uncertainty before appraisal and the next best available policy represents the value to the NHS of having access to the evidence needed at launch. This might inform a range of policies, such as early advice, public investment in transitional and evaluative research earlier in the development process and other incentives for research and development. Understanding the relationship between the time taken for research to report and the value of the evidence to future populations can help to inform (1) investments that might make research findings more quickly available, (2) the trade-off implicit in the choice of alternative research designs (i.e. greater precision or timeliness) and (3) research prioritisation

(identifying those areas in which if research is to be undertaken there must be confidence that it can report quickly).

The value of early evidence at launch and AWR can also be considered from the perspective of the manufacturer (the expected revenue streams), taking account of prices and expected volumes over the remaining patent life and technology time horizon. Together with estimates of the costs of conducting research by manufacturers or through public funding, this assessment might inform when manufacturers might be expected to conduct the research needed (e.g. high commercial value that exceeds the costs to manufacturers) or when the NHS might be expected to undertake it (e.g. low commercial value but high potential value to the NHS that exceeds the costs to the NHS). In some circumstances, the value to manufacturers and to the NHS will exceed their respective costs. The question of who should conduct, or pay for, the research might then be informed by which sector has the comparative advantage, that is, which sector can offer greater gains in net social value. The value to the NHS and to manufacturers will depend, to a large extent, on the type of pricing arrangements that might be in place (see *Implications of patient access schemes and value-based pricing*) and how any agreements that might be made can be enforced with incentive-consistent contracts. As well as understanding the pay-offs from an NHS perspective, these policy questions could be better informed with further research to estimate the likely commercial pay-offs and possible contractual arrangements that might align incentives and make AWR possible.

Social value judgements and ethical principles

Although OIR and AWR recommendations pose important ethical issues, the types of judgements required were considered ethically permissible and consistent with the social values and principles that underpin existing NICE appraisal. In general, no significant concerns about ethical issues or the types of social value judgements required when applying the checklist were raised during the workshops. Nonetheless, the analysis of the ethical issues did identify some that might require further consideration and investigation. In particular, it was recognised that an OIR recommendation may be made when clinicians and researchers would not be in equipoise about the effectiveness (as opposed to the cost-effectiveness) of an intervention. Therefore, RECs may disallow trials required by OIR guidance unless they are informed of, and agree with, the rationale for contravening the traditional principle of equipoise. This ought to be reflected in an assessment of the likelihood of research being commissioned and reporting. However, more research on how a useful notion of equipoise might be informed by the types of social decisions that bodies like NICE have to make may be valuable. The social values that underpin NICE appraisal evolve. For example, the government's plans for a value-based pricing scheme include an assessment of severity, unmet need and wider social benefits as well as QALY gains. Irrespective of how these different aspects of outcome might be included, some assessment of uncertainty and the need for evidence will still be required. Therefore, the principles of what assessments are required appear robust to such changes. Insofar as social values can be expressed quantitatively as 'weights', the assessments required can still be informed in the ways described in *Chapter 5* but based on a boarder measure of net social (rather than health) effects.

Some potential for inequity was recognised when OIR guidance might mean that patients are less likely to be included in the research. Criteria for access to the treatment trial need to be established, mindful of how unfairness might be minimised. Whether or not an AWR recommendation will be undertaken by manufacturers (see discussion in *Incentives for evaluative research*) also has an ethical dimension. One option is to incentivise research by threatening to withdraw approval if the research fails to report. However, to carry out the threat would be unfair

on future patients because the evidence base would not have changed. This potential for inequity means that NICE has no credible threat to ensure that the research recommended as part of AWR guidance will be undertaken in the way intended. It strengthens the case for establishing enforceable, incentive-consistent contractual arrangements or acknowledging that AWR may not be possible and so OIR is the only effective policy option.

How should the assessments be undertaken?

Order and sequence of assessments

The order of the assessments in the checklist relates to the sequence of decision nodes that fully describe the algorithm in *Appendix 2*. This order of considerations means that all seven assessments do not necessarily need to be made when an earlier judgement (e.g. at points 3, 5 and 6) can lead directly to guidance. In addition, the early assessment of the scale and potential significance of irrecoverable costs (at point 2) can avoid further and more complex assessment later.

There was a general consensus that the order was appropriate and that all of the assessments do not need to be routinely made for every appraisal, especially at points 6 and 7, which would need to be made only when the Appraisal Committee was considering OIR or AWR as a possibility. However, the assessment of whether, in principle, further research might be worthwhile and what type of evidence might be required would need to be undertaken routinely (additional information or analysis that might inform these assessments is discussed in *What additional information and analysis might be required?*). For example, at point 3 a judgement that more research does not seem worthwhile will lead directly to guidance (see *Point 3: Does more research seem worthwhile?*). However, if research may be worthwhile, some indication of the type of evidence needed would also be useful for those making an assessment of the prospects of research (see next section) and whether or not the type of research required to generate the evidence would be possible with approval. Therefore, routine assessment up to point 4 of the checklist (see *Point 4: Is research possible with approval?*) would seem appropriate before others with expertise and responsibility for research design and commissioning consider the prospects of the type of research needed. In fact, some assessment of other sources of uncertainty, such as future price changes, entry of new technologies and research reporting, which influence the future value of research, probably ought to be undertaken routinely as well to inform these deliberations. A number of participants suggested that this type of information, required to make an assessment at point 5 of the checklist, ought to be collected at a much earlier stage in the appraisal process – ideally at scoping.

One model for an efficient order of assessment would be to consider points 1–5 routinely (possibly with information required for point 5 collected at scoping). The Appraisal Committee would then be in a position to either rule out OIR or AWR and issue guidance in the usual way or indicate in the ACD that OIR or AWR was provisionally recommended subject to advice from a Research Advisory Committee and subsequent analysis to support an assessment of points 6 and 7 of the checklist at FAD. This model would avoid unnecessary analysis and assessment and incorporate the judgements of the research community without necessarily adding to the time that an appraisal might take. However, it does pose some questions about who will conduct the additional work between ACD and FAD and whether this is consistent with current contractual arrangements with the TAR groups or whether this work might be undertaken by the Decision Support Unit. This is a matter that NICE and NIHR are best placed to resolve.

Assessing the prospects of research

When considering OIR or AWR guidance there must be some assessment of (1) the type of research needed to address the key uncertainties, (2) whether or not this will be regarded as ethical and can be undertaken while the technology is approved for use, (3) whether or not it is likely to be a priority for public funding and be commissioned and (4) when it is likely to report.

Although the NICE appraisal process may be well suited to identifying the *need for evidence* when assessing cost-effectiveness, these other critical assessments (*the type of research and its priority*) are not necessarily ones for which NICE and its Advisory Committees, as currently constituted, have particular expertise, not least because they reflect the decisions of those responsible for research design, prioritisation and commissioning. Without sufficient co-ordination between these communities there is a danger that OIR or AWR guidance could be issued when the type of research required would either not be regarded as ethical or feasible or not of sufficient priority compared with other competing research needs to be commissioned. Because publically funded research is also budget constrained, it is perfectly possible that research that might be valuable from a wider NHS perspective might nevertheless not be a priority if other more valuable research might be displaced. This might be a particular concern if there is a possibility that the research could be undertaken by the manufacturer rather than displacing other research without a commercial interest. Therefore, a decision of whether OIR or AWR research should be undertaken by the manufacturer or through publically funded research is one that NICE cannot properly take alone (see *Incentives for evaluative research*).

Some judgement about how the research community might respond to OIR or AWR recommendations when NICE is formulating guidance is clearly possible, but more informed judgements and better decisions are likely to be possible through greater involvement of the research community. For example, a Research Advisory Committee could be constituted that could consider provisional OIR or AWR guidance (at ACD), making recommendations about the type of research needed and its ethics, feasibility and likely priority during the consultation period before final appraisal and guidance. It might also make recommendations about whether research should be publically funded or undertaken by the manufacturer with appropriate contractual arrangements (which may require the involvement of the Department of Health at some stage). Such recommendations might be informed by the type of analysis described in *Chapter 5, Point 6: Are the benefits of research greater than the costs?* and *Point 7: Are the benefits of approval greater than the costs?* and discussed in *Incentives for evaluative research* above. There are of course many different ways in which greater co-ordination might be achieved. However, as some of the assessments that must be made in formulating OIR or AWR guidance are, in fact, research decisions that fall outside the NICE remit, it would seem sensible to draw on the expertise of those involved in, and responsible for, these types of research decisions to help make these assessments, for example translating the need for particular types of evidence into a need for particular types of research, its cost, relative priority, likelihood of success and when it is likely to report. There was a general consensus among workshop participants that greater involvement of research commissioners when NICE is considering AWR or OIR recommendations would be useful. However, the finer detail of how the process of appraisal might need to be amended to accommodate this is for NICE to consider and may require an update to the methods and process of technology appraisal.

Reappraisal

A number of participants pointed out that the triggers for reappraisal need to be linked to OIR and AWR guidance in two respects: (1) to ensure that guidance is reconsidered once research findings are available and (2) to reconsider guidance if the type of research anticipated as part of AWR or OIR guidance is not undertaken and there is little prospect that it will be, for example when there is no agreement with a manufacturer or publically funded research was

not prioritised or commissioned. In addition, other changes (prices, entry of new technologies and new evidence) that might mean that research is no longer required ought to be included as criteria for reappraisal, to be considered by NICE when prioritising reappraisals. This suggests that the criteria for reappraisal and the process by which they are judged by NICE might need to be amended. Some consideration also needs to be given to the unintended consequences of reconsidering OIR before sufficient time has elapsed to encourage manufacturers to undertake research or publically funded research to be prioritised and commissioned.

What additional information and analysis might be required?

Population net health effects

In current NICE appraisal cost-effectiveness is most commonly summarised in terms of ICERs. However, such summary measures give little indication of the scale of the expected NHEs for current and future patient populations, which is a key assessment when considering OIR and AWR. For this reason cost-effectiveness was presented in terms of NHEs per patient treated and for the population of patients over time, which provides information in a way that is directly relevant to the assessments that need to be made (especially at points 2 and 7 for the checklist – see *Chapter 5*, *Point 1: Is it expected to be cost-effective?*). All of the information required to express expected cost-effectiveness in these ways is already available during appraisal and they are entirely equivalent to the more familiar ICER.

Irrecoverable costs and time horizons

Amending how irrecoverable capital costs are incorporated into the estimates of expected costs poses few technical difficulties. Some of the judgements required are already made in current methods of appraisal, for example the patient time horizon is specified in existing economic models and the expected lifetime of equipment is assumed when annuitising capital costs. The technology time horizon is not currently explicitly considered and some judgement by the Appraisal Committee supported by sensitivity analysis as well as expert clinical views is likely to be required. Considering the significance of the irrecoverable opportunity costs of initially negative NHEs of a technology (the investment profile) is more nuanced (see *Chapter 3*, *Issues specific to OIR recommendations*) and their precise effect will depend on assessments made at points 5 and 6, which can be supported by additional analysis using the type of economic models developed during appraisal. However, some early indication of their potential importance can be based on their scale relative to expected NHEs, the point at which initial losses are expected to be compensated by later gains, whether or not treatment decisions are reversible and what opportunities to improve health might be forgone by a delay to initiating treatment.

Assessing the consequences of uncertainty

The judgement at point 3 of the checklist requires some assessment of (1) how uncertain a decision based on expected cost-effectiveness might be and (2) what the consequences, in terms of population NHEs, are likely to be if an incorrect decision is made. The methods of analysis presented in *Chapter 5*, *Point 3: Does more research seem worthwhile?* attempt to decompose this assessment into a series of steps, each presenting what is already available within current methods of appraisal but in ways that can more directly inform the assessment required. Unfortunately, a simple proxy measure of the expected consequences based on expected costs and effects is not useful because it provides neither an upper or a lower bound, that is, it may seriously under- or overestimate what might be gained through further research. However, the information required to estimate the expected consequences of uncertainty and decompose them (see *Figure 10*) is already available in existing probabilistic analysis but is generally unused. Such estimates do require some judgement (implicit in all research decisions) about the period over which research findings will continue to be useful. Again, such judgements can be supported by

sensitivity analysis as well as expert views, for example from a Research Advisory Committee. The implications of scenarios and other sources of uncertainty are discussed below.

What type of evidence is required?

This assessment at point 4 of the checklist requires judgements about (1) how important particular types of parameters (inputs to the economic model) are to estimates of cost and QALYs, (2) what values these parameters would have to take to change a decision based on expected cost-effectiveness, (3) how likely is it that parameters might take such values and (4) what would be the consequences if they did, that is, what might be gained in terms of population NHEs if the uncertainty in the values of these parameters could be immediately resolved. The methods of analysis presented in *Chapter 5, Point 4: Is research possible without approval?* decompose this assessment into this series of steps, presenting in turn what is already available within current methods of appraisal but in ways that more directly inform the assessment required. It is only when assessing the consequences of uncertainty associated with particular parameters that additional analysis (using the results of existing probabilistic analysis) is required to provide quantitative estimates. There are circumstances when this additional computation is a significant burden although, increasingly, suitable simplifications and approximations are available.

Uncertainty between scenarios

It was recognised by participants that uncertainty in the parameters included in probabilistic analysis generally do not represent all sources of uncertainty. Commonly there is also uncertainty between alternative assumptions or judgements that might be made, often represented by alternative scenarios. How the consequences of uncertainty between as well as within scenarios can be presented was explored in *Chapter 5, Point 3: Does more research seem worthwhile?* and *Point 4: Is research possible without approval?*. Properly accounting for this source of uncertainty requires 'weights' (probabilities) representing the credibility of each scenario to be applied directly to the simulated output from probabilistic analysis. Although this does not require additional simulation and is quick and easy to implement, it does require that either the probabilities are made explicit in advance or estimates are presented for a range of probabilities that might represent the judgement of the Appraisal Committee following deliberation. Rather than make alternative assumptions and present extreme scenarios, formal elicitation of the judgement of clinical experts about the unknown parameters for which assumptions are required is possible. Such elicitation may provide a richer characterisation of uncertainty than weighting alternative scenarios but would require relevant experts to be identified in advance and the Appraisal Committee to accept their judgements. Alternatively, this type of analysis might be possible in real time during Appraisal Committee meetings using recent initiatives such as the Transparent Interactive Decision Interrogator, so the quantitative implications of the judgements of the Appraisal Committee could be represented.[132]

Additional information

The current appraisal process generally already provides the information and much of the analysis required to complete all of the analysis reported in *Chapter 5* and as a consequence requires little additional resource. However, the information required to assess whether or not other sources of uncertainty will resolve over time (point 5 on the checklist) is not commonly sought as part of NICE appraisal. This information includes (1) likely changes in prices of the technology and its comparators (e.g. patent expiry and likely generic prices), (2) the emergence of new technologies that might make existing ones obsolete or change their cost-effectiveness and (3) other relevant research reporting. A number of potential sources of information and evidence were examined to inform this assessment for each case study (see *Chapter 5, Point 5: Will other sources of uncertainty resolve over time?* and *Appendix 4* for full details of the sources and searches conducted). However, many potentially useful sources were proprietary or public access was

restricted, making it surprisingly difficult to inform these assessments from the information that is currently publically available. When information and estimates were available they were often not complete or directly relevant to a UK context. NICE may need to consider how ERGs, TAR groups and manufacturers can be provided with access to this type of information or whether the Institute should extract this type of information itself at an early stage of appraisal, for example at scoping or topic selection, as suggested by a number of participants.

Who should conduct the analysis?

Whether or not the ERG/TAR group or the manufacturers should be primarily responsible for conducting any additional analysis to inform these assessments was a question raised by a number of participants. Some were concerned that this analysis should be undertaken independently and others that ERGs might not have sufficient resources and time to properly interrogate submissions that include such analysis. Because the STA process bases appraisal primarily on manufacturers' submissions, any additional analysis would need to be included in the submissions and be reviewed by the ERG. Although the time and resource implications of the additional analysis are limited, even if all of the quantitative analysis in *Chapter 5* was undertaken (most is already required but is sometimes presented in different forms), more explicit consideration of OIR and AWR and the link to price would make the critique of how uncertainty and its consequences have been characterised more important. An assessment of whether or not the point estimate of cost-effectiveness is reasonable is inevitably a more limited task than also assessing whether or not the uncertainty surrounding that assessment is credible. Any additional burden on TAR groups, ERGs and manufacturers might be eased with clear guidance on the details of how analysis should be conducted and presented and what common assumptions are deemed reasonable (e.g. time horizons) and the provision of additional information by NICE. Others also suggested that greater opportunities for iteration between ERGs and manufacturers might be helpful as well as considering only points 6 and 7 on the checklist when NICE has recommended AWR or OIR at ACD and after advice from a Research Advisory Committee (see *Implications of patient access schemes and value-based pricing*).

Implementing changes informing assessments

Although it was recognised that most of the analysis presented in *Chapter 5* was appropriate and by and large simply used existing analysis to more directly inform the assessment required (consequently with limited resource implications), a number of participants felt that this might not be self-evident to Appraisal Committee members, ERGs/TAR groups and manufacturers. They suggested that some 'training' activities might be considered to ensure that appropriate skills were in place and that any additional analysis or change in the way that cost-effectiveness is presented is properly conducted and interpreted appropriately.

Acknowledgements

We would like to thank the members of our Advisory Group, Kalipso Chalkidou, Iain Chalmers, Sarah Garner, Chris Henshall, John Hutton, Catherine Law, Jonathan Michaels, Bhash Naidoo, Jon Nicholl, Hans Severens, Ken Stein, Sean Tunis and Tom Walley, who have provided invaluable feedback and comment throughout this research project. We would especially like to thank Steve Holland and Tony Hope, as ethics advisors to the project, and Iain Chalmers, as a member of the Advisory Group, for their overall contribution but in particular for their contribution to the section of the report that discusses social value judgements and ethical principles. We would like to acknowledge that they have substantially revised the original version of this discussion presented in the briefing documents to the first workshop. It is their revised version that is presented in *Chapter 3* of this report. We would also like to thank all those who participated in the two workshops, whose engagement and feedback were invaluable. We would especially like to thank Peter Littlejohns for suggesting this topic to the MRC/NIHR Methodology Research Programme and his support and encouragement throughout, including chairing both workshops. Of course, the views expressed in this report, as well as any errors or omissions, are the sole responsibility of the authors.

Contribution of authors

KC, SP and LL secured the funding for this research. All authors contributed to the concepts and analysis and participated in drafting this report. KC drafted *Chapters 1*, *3*, *5* and *6*. SP and LB drafted *Chapter 2*. LL and JY drafted *Chapter 4*. MS, CM and LB were responsible for the case study analyses. SG and ES drafted *Appendix 6*. *Appendices 7–11* were drafted by LB, SG, CM, MS and ES.

Notes

Chapter 3

a. In *Key principles and assesstments needed* and *Changes in prices and evidence*, OIR was defined as when a technology is approved only for NHS use (i.e. with the current funding directive) within the context of a suitable research study. AWR refers to approval while research is being conducted; in principle, it is when approval is conditional on research being undertaken, which in the context of NICE guidance is when the research recommendation is part of the guidance in *Key principles and assessments needed*. Whether or not such research will be undertaken and successfully reported is examined in later sections.

b. When considering OIR rather than approve in these circumstances full account must be taken of the irreversible effects of withholding access to the technology for the population of patients who could have benefited from it, for example some may not survive to benefit from the results of the research or disease may have progressed so that the expected health benefits are lower.

c. If the primary source of uncertainty is whether or not the technology is effective (i.e. whether or not there any health benefits compared with its comparators and, if so, do they compensate for any potential harms), then reductions in price will have a more limited impact on uncertainty and the need for evidence than when there might be confidence of improved overall effectiveness but there is uncertainty about if the magnitude of benefit is sufficient to justify the NHS costs. In all cases a reduction in price will increase the expected benefits of early approval for current patients so it will change the final assessment (where it is relevant) of whether or not the benefits of approval are greater than the opportunity costs (see point 7 in the checklist in *Chapter 5, A checklist of assessment*).

d. See note c.

e. Any reduction in price will make a cost-ineffective technology less so (the NHEs, even if remaining negative, will be greater) and a decision to reject will be more uncertain. However, there are limits to the effects of price reductions because even at a zero price the technology might not be cost-effective and/or further research might still may be required, for example there is no confidence that it is effective (harms may not be compensated by benefits) and/or it imposes other non-acquisition costs on the NHS.

f. The price reduction required for these different types of approval will generally be greater if research is not possible with approval. 'Approve[2]' would require the greatest reduction in price and 'Approve[4]' the lowest. However, any price reduction (price greater than zero) may not make approval appropriate in these circumstances.

g. See notes a, c and e.

h. We would like to thank Steve Holland and Tony Hope, as ethics advisors to the project, and Iain Chalmers, as a member of the Advisory Group, for their overall contribution, but especially to this section of the report. We would like to acknowledge that they have substantially revised the original version of this discussion of ethical principles presented in the briefing documents to the first workshop (www.york.ac.uk/che/research/teehta/workshops/only-in-research-workshop/). It is their revised version that is presented here.

Chapter 5

a. TA104[126] included an AWR recommendation in the ACD but this was removed in the FAD. The recommended research was to enter patients onto the BSR Biologics Register on the

grounds of the possibility of severe side effects and little information on the use of these agents beyond the duration of RCTs.

b. In each case study the estimates of expected costs and QALYs reported and used throughout are the mean costs and QALYs derived from probabilistic analysis using Monte Carlo simulation. The costs and QALYs from a deterministic analysis will be incorrect unless the model is multilinear with independent parameters (see *Appendix 6, Is the intervention cost-effective?*).

c. This is the time horizon over which costs and benefits are likely to differ for an individual patient (commonly termed the model time horizon). In some circumstances (e.g. when there is a mortality effect) this will be the lifetime of the patient. Expected costs and QALYs in each period are the expectations (means) from the results of probabilistic analysis. All future costs and QALYs (per patient or population) are discounted at 3.5% throughout (see *Appendix 6, Is the intervention cost-effective?, Cost-effectiveness at the patient level*).

d. The expected per-patient NHE for each intervention (*i*) is the difference between the expected health (QALYs) with the intervention (h_i) and the health likely to be forgone elsewhere as a consequence of the costs of the intervention (c_i), which requires an estimate of the cost-effectiveness threshold (*k*). Therefore, the per-patient expected NHE of each intervention ($NHE_i = h_i - c_i/k$) can be expressed using the same information required to present the more familiar ICERs. It can also be expressed in terms of the NHS resources required to generate the NHE_i ($k \times h_i - c_i$) The intervention that is expected to be cost-effective is the one with the highest expected NHE. This is entirely equivalent to drawing conclusions about cost-effectiveness based on ICERs but has many advantages once an assessment of uncertainty and its consequences is required. It is also needed when considering the impact of irrecoverable costs and is especially important when decisions require a trade-off to be made between benefits to current and future patients (see *Appendix 6, Is the intervention cost-effective?, Cost-effectiveness at the patient level*).

e. All analysis has been conducted at the upper and lower bound for the range that NICE has adopted for the threshold; however, unless otherwise stated results in the text relate to a threshold of £20,000 per QALY.

f. The time at which initially negative NHEs are expected to be offset by cumulating positive NHEs, or the break-even point, is only an indicator or proxy for the scale of irrecoverable opportunity costs as, for example, the scale of initial loses also matters. The presence of even very large irrecoverable opportunity costs does not necessarily mean that they are significant and will influence guidance. That will depend on whether or not treatment decisions are reversible and the impact of withholding treatment for patients who might receive it (see *Chapter 3, Incentives for evaluative research* and this chapter, *Point 2: Are there significant irrecoverable costs?*), as well as whether uncertainty is likely to resolve (research reporting or other sources of uncertainty resolving; see *Appendix 6, Are there significant irrecoverable costs?*).

g. The time horizon for the technology might be longer or shorter than the patient time horizon. Technology time horizons might be based on historical evidence of the obsolescence of health technologies but any estimate will be a proxy for a complex and uncertain process of future changes in new technologies, prices and evidence. Therefore, the impacts of different technology time horizons have been explored in each case study (see *Appendix 6, Is the intervention cost-effective?, Cost-effectiveness at the population level*).

h. The annual payments required each year over the lifetime of the equipment (discounted at 3.5%), which would be equivalent to the capital cost at the start of the first year.

i. The commitment of irrecoverable capital costs will depend on the rate at which any resale value of purchased equipment depreciates over its lifetime. Often rental agreements are possible that reduce commitment but commonly at a higher total cost. These issues are explored in the *Appendices 7–10*.

j. For OMAL, the profile of NHEs at a patient level did not exhibit significant irrecoverable opportunity costs. Assessment at a population level and for longer technology time horizons simply increases the magnitude of the expected negative NHE; therefore, there are no irrecoverable costs in this case study.

k. In *Figure 7* the technology time horizon happens to coincide with the lifetime of the equipment but it need not.

l. It is also important to consider the risk profile of the health technologies and activities likely to be displaced. Insofar as additional NHS costs do not just displace new technologies the net effect will still tend to increase 'risk' (*see Appendix 11, Accounting for the investment profile of displaced interventions*).

m. In mathematics and economics this is referred to as expected opportunity loss. In decision theory and its applications, including economic evaluation, it is referred to as the expected value of perfect information (EVPI). It is also directly related to option value in financial economics (see *Appendix 11, Why averaging scenarios may be misleading*).

n. The alternative treatment, which is expected to be cost-effective, may not have the highest probability of being cost-effective if, as in this case, the distribution of NHEs are skewed, that is, when the NHEs of EECP are greater than those of standard care they are much greater but when standard care offers higher NHEs they are only a little higher than those of EECP.

o. The time horizon over which evidence generated by research about a technology might be valuable may be longer (or shorter) than the period over which the technology is used. Therefore, there is a distinction between the technology time horizon and the time horizon for the benefits of research. To simplify the exposition in this summary of the case studies they are assumed to be equal but other credible assumptions are explored more fully in *Appendices 7–10*.

p. If an assessment of expected consequences based on mean NHEs was always an underestimate this would be a useful, simple assessment of a lower bound to the potential benefits of research. However, such estimates can also overestimate expected consequences, for example in the analysis of EECP at a threshold of £30,000. Unfortunately such circumstances cannot be specified in advance without conducting a proper analysis of the expected consequences anyway (see *Appendix 11, Why the consequences of uncertainty differ from mean incremental effects*).

q. This is substantially greater than the estimate of 800 QALYs based on mean incremental population NHEs, demonstrating that such simple estimates may be misleading (see *Figure 8*).

r. It should be noted that these estimates of QALYs that might be gained are for the population over the time horizon for the benefit of research (in this case equal to the technology time horizon) if all sources of uncertainty could be immediately resolved. They include both improvements in health outcomes for this population but also NHS resource savings that could be made and used to generate QALYs elsewhere.

s. Five experts with experience and knowledge of EECP in the UK independently completed an Microsoft Excel-based exercise (Microsoft Corporation, Redmond, WA, USA). The uncertainty associated with any judgement is critical, so a frequency chart format, in which experts place 20 crosses on a frequency chart to represent a distribution, was adopted. The results from each expert were linearly pooled, with equal weight, providing the probability of continuing to respond to treatment in each subsequent year. The uncertainty associated with these pooled estimates was characterised by fitting beta distributions to pooled responses.

t. The probability of error associated with 12 months of treatment reported in *Table 19* will not, in general, equal the sum of the probabilities of error across the parameters because the overall probability from PSA takes account of the joint effect of uncertainty in all parameters simultaneously. Even if parameters are independent they will be related to differences in NHE in different ways (indicated by the sign and magnitude of the elasticities – see

Table 21), so sometimes the effect of uncertainty in one may, to some extent, 'substitute' or 'complement' the effect of uncertainty in others.

u. For a similar reason to that in note t, the overall expected consequences of uncertainty reported in *Table 19* (5194 QALYs) will not, in general, equal the sum of the expected consequences for each of the parameters separately (5432 QALYs).

v. The sequence in which research might be conducted can also be considered. This is discussed at greater length and more formally in *Appendix 6*, that is, OIR for treatment effects followed by AWR for natural history if that research continues to be necessary once research recommended in OIR reports. If it is feasible to withdraw approval of a technology that is expected to be cost-effective to allow research to be conducted then in this case AWR followed by OIR if necessary is likely to offer greater expected population NHEs (see *Appendix 6, Does the sequence of research have implications for AWR and OIR?*).

w. These values represents an upper bound on what might be gained by resolving each alone. The potential value of resolving these uncertainties together is much greater for the reasons given in note t.

x. This assumes that either prescribing will switch from the brand to the equivalent generic (brands tend to maintain or even increase premium prices in some health-care systems after patent expiry) or that any new branded technologies will be appraised and/or priced using generic versions of the old brand as a comparator.

y. The value of the research will be realised only when clinical practice changes because of either revised NICE guidance or diffusion of research findings.[73,74] Some assessment of the speed of implementation of research findings and revised NICE guidance ought to be considered at this point (see *Appendix 6, How long until the research reports?*).

z. These are much higher values of immediate research than the 4495 QALYs or £89.9M in *Figure 15* without the entry of a new technology but with a similar price change.

aa. There may continue to be value if evidence about CLOP remains an important link in mixed or indirect treatment comparisons required to evaluate the new technology.

ab. This is likely to be an increasingly common scenario if value-based pricing effectively makes all branded technologies equally cost-effective.

ac. Even with such investment AWR might not be possible if there is insufficient variation in treatment assignment and no robust way of controlling for unobserved characteristics through selection models, for example use of instrumental variables.

ad. Research as part of OIR might also commit irrecoverable costs although the scale of this commitment will be more limited than with approve or AWR, that is, OIR will avoid at least some of the commitment.

ae. Any difference in costs of research under AWR or OIR guidance can easily be integrated into these assessments.

Appendix 5

a. The term 'health technology' is used here as a generic descriptor for all types of interventions examined in NICE's Technology Appraisal Workstream.

b. Department of Health. *Faster access to modern treatment: how NICE appraisal will work.* London: Department of Health; 1999.

c. http://www.nice.org.uk/media/A1A/E6/NICEAHRQWorkshopReportFINAL.pdf.

Appendix 6

a. For interventions with non-marginal impacts the extent of displacement is large enough to alter *k*.

b. Throughout we assume that an unbiased estimate of the expectation is derived from a probabilistic analysis.

c. In principle, when applying this framework it would be possible that a new intervention that offered lower health benefits to patients than current treatments $(h(2,\theta) < h(1,\theta))$ could offer an improvement in NHEs if the costs were sufficiently lower $(c(2,\theta) < c(1,\theta))$ because the health gains achievable by investing the released funds in other activities outweigh the health losses to patients receiving the less effective intervention

$$[c(1,\theta) - c(2,\theta)]/k > h(1,\theta) - h(2,\theta).$$

Some might have concerns that this could lead to patients being provided with harmful interventions, but in most jurisdictions there exists a mechanism for preventing this outcome in the form of social values, laws and regulatory agencies that prevent interventions that are harmful compared with no treatment from being developed or coming to market.

d. We assume that the investment profile of the displaced activities matches that of the new intervention.

e. This irrecoverable cost was incorporated in *Tables 36* and *37* using a time horizon of $T_C = 10$ for annuitisation and a discount rate of zero and assuming that it provided a joint resource which was utilised by the whole patient population during that time period.

f. As yet we have not considered how guidance might change and so cannot estimate the benefits that might accrue after that point, but this is considered later (see *How long until the research reports?*).

g. The value of further research is not contingent on the adoption decision. In *Table 39*, regardless of whether or not the new technology is the cost-effective alternative there is an additional benefit to resolving θ. If $NHE(1,\theta)$ is the net benefit of the new treatment, *Table 39* shows that it is not cost-effective to approve the new treatment but there is still a benefit to further research rather than simply rejecting it and continuing to use intervention 2.

h. When alternative modelling assumptions would suggest that different interventions would be expected to be cost-effective the NHEs of choosing the best intervention when integrating both parameter and modelling uncertainty cannot be found by averaging the NHEs of choosing the best intervention within each scenario. This is because

$$\max_j E_\theta E_s NHE(j,\theta,s) \neq E_s\left[\max_j E_\theta NHE(j,\theta,s)\right].$$

i. The pay-offs from approve and reject can be expanded by splitting at $\tau = 6$ to facilitate a more easy comparison of the pay-offs pre and post research reporting, e.g.

$$\Pi_{Approve} = \max_j E_\theta NHE(j,\theta) . \sum_{t=1}^{\tau} \frac{I_t}{(1+d)^t} + \max_j E_\theta NHE(j,\theta) . \sum_{t=\tau}^{T} \frac{I_t}{(1+d)^t}.$$

Appendix 7

a. A pilot exercise was initially conducted with one expert to ensure that the questions were clear and interpreted correctly. For the final exercise, seven experts were identified on the basis of their experience and knowledge of EECP in the UK. Five of them completed the exercise.

b. All analysis has been conducted at the upper and lower bounds for the range that NICE has adopted for the threshold; however, unless otherwise stated results in the text and figures relate to a threshold of £20,000 per QALY.

c. In *Figure 45* the technology time horizon happens to coincide with the lifetime of the equipment but it need not.

d. In mathematics and economics this is referred to as expected opportunity loss. In decision theory and its applications, including economic evaluation, it is referred to as the EVPI. It is also directly related to option value in financial economics.

e. The alternative, which is expected to be cost-effective, may not have the highest probability of being cost-effective if, as in this case, the distribution of NHEs are skewed, that is, when the NHEs of EECP are greater than those of standard care they are much greater but when standard care offers higher NHEs they are only a little higher than those of EECP.

f. Any difference in costs of research under AWR or OIR guidance can easily be integrated into these assessments.

Appendix 8

a. Estimates of expected costs and QALYs reported and used in subsequent analysis are the mean costs and QALYs derived from probabilistic analysis using Monte Carlo simulation. The costs and QALYs from a deterministic analysis will be incorrect unless the model is linear or multilinear with independent parameters. Within the model, CLOP was assumed to not impact directly on the long-term outcomes of patients and this is important in evaluating model linearity, specifically in analyses over treatment effect parameters. Treatment effects were applied in a multilinear formulation of the decision model (the decision tree); because of this we do not expect significant non-linearity when analysing the value of further research associated with these parameters.

b. The following table depicts the expected cost-effectiveness of CLOP per patient treated estimated using a deterministic analysis. These results indicate no evidence of non-linearity of the model.

Treatment	Cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY NHE, QALYs (£)	Cost-effectiveness threshold £30,000 per QALY NHE, QALYs (£)
CLOP12	20,094	8.117	18,723	7.112 (142,246)	7.447 (223,416)
CLOP6	19,826	8.093	10,514	7.102 (142,034)	7.432 (222,964)
CLOP3	19,678	8.078	9425	7.094 (141,882)	7.422 (222,662)
CLOP1	19,564	8.066	4999	7.088 (141,756)	7.414 (222,416)
NHS	19,467	8.047	–	7.074 (141,473)	7.398 (221,943)

c. All analysis has been conducted at the upper and lower bounds for the range that NICE has adopted for the threshold; however, unless otherwise stated results in the text relate to a threshold of £20,000 per QALY.

d. The time horizon for the technology might be longer or shorter than the patient time horizon. Technology time horizons might be based on historical evidence of the obsolescence of health technologies but any estimate will be a proxy for a complex and uncertain process of future changes in new technologies, prices and evidence. Therefore, the impacts of different technology time horizons have been explored.

e. In mathematics and economics this is referred to as expected opportunity loss. In decision theory and its applications, including economic evaluation, it is referred to as the EVPI. It is also directly related to option value in financial economics.

f. The consequences of uncertainty based on means can be estimated from $(1-p)\times(NHE_{CLOP12}-NHE_{NHS})$, where p represents the probability that the decision to adopt CLOP is correct.

g. It should be noted that these estimates of QALYs that might be gained are for the population over the time horizon for the benefits of research (in this case equal to the technology time horizon) if all sources of uncertainty could be immediately resolved. They include both improvements in health outcomes for this population but also NHS resource savings that could be made and used to generate QALYs elsewhere.

h. The probability of error associated with 12 months of treatment reported in *Table 93* will, in general, not equal the sum of the probabilities of error across the parameters because the overall probability from PSA takes account of the joint effect of uncertainty in all parameters simultaneously. Even if parameters are independent they will be related to differences in NHEs in different ways (indicated by the sign and magnitude of the elasticities), so sometimes the effect of uncertainty in one may, to some extent, 'substitute' or 'complement' the effect of uncertainty in others.

i. The overall expected consequences of uncertainty reported in *Table 93* (5194 QALYs) will not, in general, equal the sum of the expected consequences for each of the parameters separately (5432 QALYs).

j. The sequence in which research might be conducted can also be considered. This is discussed at greater length and more formally in *Appendix 6*, that is, OIR for treatment effects followed by AWR for natural history if that research continues to be necessary once research recommended in OIR reports. If it is feasible to withdraw approval of a technology that is expected to be cost-effective to allow research to be conducted then in this case AWR followed by OIR if necessary is likely to offer greater expected population NHEs.

k. This assumes that either prescribing will switch from the brand to the equivalent generic (brands tend to maintain, or even increase, premium prices in some health-care systems after patent expiry) or that any new branded technologies will be appraised using generic versions of the old brand as a comparator.

l. There may continue to be value if evidence about CLOP remains an important link in mixed or indirect treatment comparisons required to evaluate the new technology.

m. This is likely to be an increasingly common scenario if value-based pricing effectively makes all branded technologies equally cost-effective.

n. These are much higher values of immediate research than the 4495 QALYs or £89.9M in *Tables 103* and *104* without the entry of a new technology but with a similar price change.

o. Under OIR access to CLOP is not granted until research reports. In calculations we assumed that, before research reports, patients accrue an average NHE estimated using treatments other than CLOP. In averaging weights derived from the PSA results were used.

p. If patients waiting for research to report under OIR use the next best treatment (6 months of treatment with CLOP), instead of an average NHE, T* is 7.5 years.

Appendix 9

a. Estimates of expected costs and QALYs reported and used in subsequent analysis are the mean costs and QALYs derived from probabilistic analysis using Monte Carlo simulation. The costs and QALYs from a deterministic analysis will be incorrect unless the model is linear or multilinear with independent parameters.

b. All analysis has been conducted at the upper and lower bounds for the range that NICE has adopted for the threshold; however, unless otherwise stated results in the text and figures relate to a threshold of £20,000 per QALY.

c. All analysis has been conducted using an annual incidence of 307 patients unless otherwise stated.

Appendix 10

a. TA104[126] included an AWR recommendation in the ACD but this was removed in the FAD. The recommended research was to enter patients into the BSR register on the grounds of the possibility of severe side effects and little information on the use of these agents beyond the duration of RCTs.

b. Throughout the case study, estimates of expected costs and QALYs reported and used in subsequent analysis are the mean costs and QALYs derived from probabilistic analysis using

Monte Carlo simulation. The costs and QALYs from a deterministic analysis will be incorrect unless the model is multilinear with independent parameters.

c. This is the time horizon over which costs and benefits are likely to differ for an individual patient (commonly termed the model time horizon). In some circumstances (e.g. when there is a mortality effect) this will be the lifetime of the patient. Expected costs and QALYs each period are the expectations (means) from the results of probabilistic analysis. All future costs and QALYs (per patient or population) are discounted at 3.5% throughout.

d. Because of the non-linear nature of the PsA model, the deterministic results are somewhat different from the means of the probabilistic analysis. The deterministic results show that adalimumab is the most cost-effective treatment, with an ICER of around £18,000.

e. The time horizon for the technology might be longer or shorter than the patient time horizon. Technology time horizons might be based on historical evidence of the obsolescence of health technologies but any estimate will be a proxy for a complex and uncertain process of future changes in new technologies, prices and evidence. Therefore, the impacts of different technology time horizons have been explored.

Appendix 11

a. For a more comprehensive introduction to the evaluation of the value of further research associated with specific parameters please see Brennan *et al.*[238] The authors also highlight briefly the issues of non-linearity and correlation in this context.

b. To estimate the value of resolving all uncertainty (upper bound for the value of further research) in a specific parameter (or subset of parameters) of interest, θ^I, when uncertainty in other parameters will not be resolved (θ^I such that $\theta = \{\theta_I, \theta^C\}$), the following expression needs to be evaluated

$$E_{\theta^I} \max_j E_{\theta^C|\theta^I} NHE_j(\theta) - \max_j E_\theta NHE_j(\theta).$$

We need to know not only the expected NHEs of each intervention but also the expected maximum NHEs. The latter involves a nested expectation that, when evaluated using probabilistic analysis, involves two nested simulation procedures to be used, one for each expectation.

c. Following the notation used in note b, the expected NHEs conditional on θ^I assuming a specific value x can be represented by

$$E_{\theta^C|\theta^I=x} NHE_j(\theta).$$

The evaluation of conditional NHEs is integral to the evaluation of the value of resolving all uncertainty in specific parameters (see note b).

d. If the distribution of $\theta^C | \theta^I$ is known, we can evaluate directly the expected conditional NHE:

$$E_{\theta^C|\theta^I=x} NHE_j\left(\theta^I = x, \theta^C | \theta^I = x\right).$$

e. In using simulated Markov chain Monte Carlo results informing multiple model parameters after Bayesian inference, the Markov chain Monte Carlo simulations on the set of parameters must be kept paired to maintain correlation within the group of parameters.

f. Brennan *et al.*[238] has demonstrated that the presence of small correlations may not impact significantly on estimates of the value of resolving uncertainty when assuming linearity.

g. Also, in multilinear models the relation $E\left[NHE_{\theta_1|\theta_2=x}\right] \approx E\left[NHE_{E(\theta_1)|\theta_2=x}\right]$ should sustain in the absence of meaningful correlation.

h. For the purpose of this example we assume that the decision-maker allocates the budget over 10-year periods. More complexity is added if the budget must be balanced in a time frame less than the lifetime of the intervention. This is because of the difference between when the costs for the new treatment are accrued and when the costs from the displaced treatment are released.

References

1. NICE. *Guide to the methods of technology appraisal*. London: NICE; 2008.

2. Tunis S, Pearson S. Coverage options for promising technologies: Medicare's 'coverage with evidence development'. *Health Aff* 2006;**25**:1218–30.

3. Chalkidou K, Hoy A, Littlejohns P. Making a decision to wait for more evidence: when the National Institute for Health and Clinical Excellence recommends a technology only in the context of research. *J R Soc Med* 2007;**100**:453–60.

4. Hutton J, Trueman P, Henshall C. Coverage with evidence development: an examination of conceptual and policy issues. *Int J Technol Assess Health Care* 2007;**23**:425–35.

5. Tunis S, Chalkidou K. Coverage with evidence development: a very good beginning, but much to be done. Commentary to Hutton *et al. Int J Technol Assess Health Care* 2007;**23**:432–5.

6. Eckermann S, Willan AR. The option value of delay in health technology assessment. *Med Decis Making* 2008;**28**:300–5.

7. Palmer S, Smith P. Incorporating option values into the economic evaluation of health care technologies. *J Health Econ* 2000;**19**:755–66.

8. Stafinski T, McCabe CJ, Menon D. Funding the unfundable: mechanisms for managing uncertainty in decisions on the introduction of new and innovative technologies into healthcare systems. *Pharmacoeconomics* 2010;**28**:113–42.

9. Cooksey D. *A review of UK health research funding*. London: HM Treasury; 2006.

10. House of Commons Health Committee. *The influence of the pharmaceutical industry: fourth report of session 2004–2005*. London: The Stationary Office; 2005.

11. NICE Citizens Council. *'Only in research'*. London: NICE; 2007.

12. Office of Fair Trading (OFT). *The pharmaceutical price regulation scheme. An OFT market study*. London: OFT; 2007.

13. Brezis M, Lehmann LS. Medicare requirement for research participation. *JAMA* 2006;**296**:2924.

14. Gafni A, Birch S. Building bridges between academic research and policy formulation: when costing less means costing more. *Pharmacoeconomics* 2007;**25**:523–8; author reply 528–32.

15. Goeree R, O'Reilly D, Tarride JE, Bowen JM, Blackhouse G, Campbell K, *et al.* Being led down the wrong garden PATH: the importance of knowledge and facts for the crossroads. *Pharmacoeconomics* 2007;**25**:528–32.

16. Groeneveld PW. Medicare requirement for research participation. *JAMA* 2006;**296**:2923–4.

17. Bowen JM, Patterson LL, O'Reilly D, Hopkins RB, Blackhouse G, Burke N, *et al.* Conditionally funded field evaluations and practical trial design within a health technology assessment framework. *J Am Coll Radiol* 2009;**6**:324–31.

18. Briggs A, Ritchie K, Fenwick E, Chalkidou K, Littlejohns P. Access with evidence development in the UK: past experience, current initiatives and future potential. *Pharmacoeconomics* 2010;**28**:163–70.

19. Carbonneil C, Quentin F, Lee-Robin SH, European network for Health Technology Assessment (EUnetHTA). A common policy framework for evidence generation on promising health technologies. *Int J Technol Assess Health Care* 2009;**25**(Suppl. 2):56–67.

20. Goeree R, Levin L. Building bridges between academic research and policy formulation: the PRUFE framework – an integral part of Ontario's evidence-based HTPA process. *Pharmacoeconomics* 2006;**24**:1143–56.

21. McCabe CJ, Stafinski T, Edlin R, Menon D. Access with evidence development schemes: a framework for description and evaluation. *Pharmacoeconomics* 2010;**28**:143–52.

22. Miller W. *Value-based coverage policy in the United States and the United Kingdom: different paths to a common goal.* Washington, DC: National Health Policy Forum; 2006.

23. Towse A, Garrison LP, Jr. Can't get no satisfaction? Will pay for performance help?: toward an economic framework for understanding performance-based risk-sharing agreements for innovative medical products. *Pharmacoeconomics* 2010;**28**:93–102.

24. Trueman P, Grainger DL, Downs KE. Coverage with evidence development: applications and issues. *Int J Technol Assess Health Care* 2010;**26**:79–85.

25. Carlson JJ, Sullivan SD, Garrison LP, Neumann PJ, Veenstra DL. Linking payment to health outcomes: a taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. *Health Policy* 2010;**96**:179–90.

26. Chalmers I. Addressing uncertainties about the effects of treatments offered to NHS patients: whose responsibility? *J R Soc Med* 2007;**100**:440–1.

27. Ramsey SD, Sullivan SD. Evidence, economics, and emphysema: Medicare's long journey with lung volume reduction surgery. *Health Aff* 2005;**24**:55–66.

28. Wallner P, Konski A. A changing paradigm in the study and adoption of emerging health care technologies: coverage with evidence development. *J Am Coll Radiol* 2008;**5**:1125–9.

29. Chalkidou K, Lord J, Fischer A, Littlejohns P. Evidence-based decision making: when should we wait for more information? *Health Aff* 2008;**27**:1642–53.

30. Tunis S, Whicher D. The National Oncologic PET Registry: lessons learned for coverage with evidence development. *J Am Coll Radiol* 2009;**6**:360–5.

31. Medicare's coverage with evidence development: a policy-making tool in evolution. *J Oncol Pract* 2007;**3**:296–301.

32. Kamerow D. Paying for promising but unproven technologies: is the policy on 'coverage with evidence development' a helpful way forward? *BMJ* 2007;**335**:965.

33. Mohr PE, Tunis SR. Access with evidence development: the US experience. *Pharmacoeconomics* 2010;**28**:153–62.

34. Carino T, Williams RD II, Colbert AM, Bridger P. Medicare's coverage of colorectal cancer drugs: a case study in evidence development and policy. *Health Aff* 2006;**25**:1231–9.

35. Brice J. Imaging research efforts place focus on clinical effectiveness. Research to Practice. 2008. URL: www.diagnosticimaging.com/display/article/113619/1180672.

36. Breckenridge A, Walley T. Risk sharing and payment by results. *Clin Pharmacol Ther* 2008;**83**:666–7.

37. Carino T, Sheingold S, Tunis S. Using clinical trials as a condition of coverage: lessons from the National Emphysema Treatment Trial. *Clinical Trials* 2004;**1**:108–21.

38. Dhalla IA, Garner S, Chalkidou K, Littlejohns P. Perspectives on the National Institute for Health and Clinical Excellence's recommendations to use health technologies only in research. *Int J Technol Assess Health Care* 2009;**25**:272–80.

39. Garber A. Is having more preapproval data the best way to assure drug safety? *Health Aff* 2008;**5**:371–3.

40. Krich TB. CMS new coverage with evidence policy. *J BioLaw Bus* 2007;**10**:43–4.

41. Lindsay MJ, Siegel BA, Tunis SR, Hillner BE, Shields AF, Carey BP, *et al.* The National Oncologic PET Registry: expanded Medicare coverage for PET under coverage with evidence development. *AJR Am J Roentgenol* 2007;**188**:1109–13.

42. Macdonald F. *Coverage with evidence development: a personal, industry view*. Presentation slides. 19 September 2008. URL: www.docstoc.com/docs/105704904/ Coverage-With-Evidence-Development.

43. Menon D, McCabe CJ, Stafinski T, Edlin R, signatories to the Consensus Statement. Principles of design of access with evidence development approaches: a consensus statement from the Banff Summit. *Pharmacoeconomics* 2010;**28**:109–11.

44. Neumann PJ, Kamae MS, Palmer JA. Medicare's national coverage decisions for technologies, 1999–2007. *Health Aff* 2008;**27**:1620–31.

45. Niezen M, de Bont A, Stolk E, Eyek A, Niessen L, Stoevalaar H. Conditional reimbursement with the Dutch drug policy. *Health Policy* 2007;**84**:39–50.

46. Pearson SD, Miller FG, Emanuel EJ. Medicare's requirement for research participation as a condition of coverage: is it ethical? *JAMA* 2006;**296**:988–91.

47. Reed S, Shea A, Schulman K. Economic implications of potential changes to regulatory and reimbursement policies for medical devices. *J Gen Intern Med* 2008;**23**:50–6.

48. Straube BM. How changes in the Medicare coverage process have facilitated the spread of new technologies. *Health Aff* 2005;(Suppl. Web Exclusives):W5-314–W5-316.

49. Thornton S. Drug price reform in the UK: debunking the myths. *Health Econ* 2007;**16**:981–92.

50. Whalley GA, Gamble GD, Walsh HJ, Wright SP, Agewall S, Sharpe N, *et al.* Effect of tissue harmonic imaging and contrast upon between observer and test–retest reproducibility of left ventricular ejection fraction measurement in patients with heart failure. *Eur J Heart Fail* 2004;**6**:85–93.

51. NICE and AHRQ. *Managing uncertainty in healthcare*. Report of a meeting organised by NICE and AHRQ. 2008. URL: www.nice.org.uk/aboutnice/howwework/ researchanddevelopment/managinguncertaintyinhealthcare.jsp.

52. Department of Health. *Faster access to modern treatment: how NICE appraisal will work*. London: Department of Health; 1999.

53. NHS Quality Improvement Scotland. *Coverage with evidence development in NHS Scotland*. Discussion paper. Edinburgh: NHS Quality Improvement Scotland; 2008.

54. Office for Life Sciences, HM Government. *Life sciences blueprint*. A statement from the Office for Life Sciences. July 2009. London: Department for Business, Innovation and Skills; 2009.

55. Centers for Medicare & Medicaid Services. Medicare and Medicaid programs; conditions for coverage for organ procurement organizations (OPOs). Final rule. *Fed Regist* 2006;**71**:30981–1054.

56. Mechanic R, Zinner D. *Optimizing information under coverage with evidence development*. Health Industry Forum Policy Brief. 24 October 2007. Waltham, MA: The Health Industry Forum; 2007.

57. Slutsky J. *CED for promising technologies: USA*. Paper presented at HTAi 2008, Montreal, Canada, July 2008.

58. Sansom L. *Medicines use: what's next for the PBS*. Presentation slides.

59. Chalkidou K. *'Only in research': a polite 'no' or a valuable policy option?* Birmingham: NICE; 2006.

60. Reeves B. *Verteporfin photodynamic therapy (PDT) cohort study for the UK 'coverage with evidence development'*. Paper presented at HTAi 2008, Montreal, Canada, July 2008.

61. Garrison L, Sullivan S, Carlson J, Bresnahan B, Carlson R, Neumann P, *et al*. *Performance-based risk-sharing reimbursement agreements for new medical products*. Washington, DC: University of Washington; 2008.

62. Culyer A, McCabe C, Briggs A, Claxton K, Buxton M, Akehurst R, *et al*. Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. *J Health Serv Res Policy* 2007;**12**:56–8.

63. Phelps CE, Mushlin AI. On the (near) equivalence of cost-effectiveness and cost–benefit analyses. *Int J Technol Assess Health Care* 1991;**7**:12–21.

64. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998;**18**:S68–80.

65. Claxton K, Briggs A, Buxton M, Culyer A, McCabe C, Walker S, *et al*. Value based pricing for NHS drugs: an opportunity not to be missed? *BMJ* 2008;**336**:251–4.

66. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics* 2008;**26**:733–44.

67. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;**18**:342–64.

68. Claxton K, Cohen J, Neumann P. When is evidence sufficient? *Health Aff* 2005;**24**:93–101.

69. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence. *Lancet* 2002;**360**:711–15.

70. Griffin SC, Claxton KP, Palmer SJ, Sculpher MJ. Dangerous omissions: the consequences of ignoring decision uncertainty. *Health Econ* 2011;**20**:212–24.

71. Philips Z, Claxton K, Palmer S. The half-life of truth: what are appropriate time horizons for research decisions? *Med Decis Making* 2008;**28**:287–99.

72. Eckermann S, Willan A. Expected value of information and decision making in HTA. *Health Econ* 2006;**16**:195–209.

73. Fenwick E, Claxton K, Sculpher M. The value of implementation and the value of information: combined and uneven development. *Med Decis Making* 2008;**28**:21–32.

74. Hoomans T, Fenwick EA, Palmer S, Claxton K. Value of information and value of implementation: application of an analytic framework to inform resource allocation decisions in metastatic hormone-refractory prostate cancer. *Value Health* 2009;**12**:315–24.

75. McKenna C, Claxton K. Addressing adoption and research design decisions simultaneously: the role of value of sample information analysis. *Med Decis Making* 2011;**31**:853–65.

76. Basu A. Economics of individualization in comparative effectiveness research and a basis for a patient-centered health care. *J Health Econ* 2011;**30**:549–59.

77. Claxton K, Sculpher M, Carroll S. *Value-based pricing for pharmaceuticals: its role, specification and prospects in a newly devolved NHS.* York: Centre for Health Economics, University of York; 2011.

78. Department of Health. *A new value-based approach to the pricing of branded medicines: a consultation.* London: Department of Health; 2010.

79. Department of Health. *A new value-based approach to the pricing of branded medicines: government response to consultation.* London: Department of Health; 2010.

80. Claxton K. OFT, VBP: QED? *Health Econ* 2007;**16**:545–58.

81. Department of Health. *The pharmaceutical price regulation scheme.* London: Department of Health; 2008.

82. Chalkidou K, Walley T, Culyer A, Littlejohns P, Hoy A. Evidence-informed evidence-making. *J Health Serv Res Policy* 2008;**13**:167–73.

83. NICE. *Social value judgements guideline. Principles for the development of NICE guidance.* London: NICE; 2008.

84. Kumar A, Soares H, Wells R, Clarke M, Hozo I, Bleyer A, *et al.* Are experimental treatments for cancer in children superior to established treatments? Observational study of randomised controlled trials by the Children's Oncology Group. *BMJ* 2005;**331**:1295.

85. Levine RJ. The 'best proven therapeutic method' standard in clinical trials in technologically developing countries. *IRB* 1998;**20**:5–9.

86. NICE. *Process for advising on the feasibility of implementing a patient access scheme.* London: NICE; 2009.

87. Department of Health. *Health service circular 2002/004 – cost effective provision of disease modifying therapies for people with multiple sclerosis.* London: Department of Health; 2002.

88. NICE. *The use of liquid-based cytology for cervical screening.* TA5. London: NICE; 2003.

89. NICE. *Infliximab for acute exacerbations of ulcerative colitis.* TA163. London: NICE; 2008.

90. NICE. *Hearing disability – hearing aids (withdrawn).* TA8. London: NICE; 2000.

91. NICE. *Inhaled insulin for the treatment of type 1 and type 2 diabetes.* TA113. London: NICE; 2006.

92. NICE. *Guidance on the use of etanercept and infliximab for the treatment of rheumatoid arthritis.* TA36. London: NICE; 2002.

93. NICE. *Guidance on the use of patient-education models for diabetes.* TA60. London: NICE; 2003.

94. NICE. *Guidance on the use of photodynamic therapy for age-related muscular degeneration.* TA68. London: NICE; 2003.

95. NICE. *Computerised cognitive behaviour therapy for depression and anxiety: guidance.* TA97. London: NICE; 2006.

96. NICE. *Computerised cognitive behaviour therapy (CCBT) for the treatment of depression and anxiety.* TA51. London: NICE; 2002.

97. NICE. *Rituximab for aggressive non-Hodgkin's lymphoma.* TA65. London: NICE; 2003.

98. NICE. *Guidance on the use of metal on metal hip resurfacing arthroplasty.* TA44. London: NICE; 2002.

99. NICE. *Interferon alfa (pegylated and non-pegylated) and ribavirin for the treatment of chronic hepatitis C.* TA75. London: NICE; 2004.

100. NICE. *Lymphoma (follicular non-Hodgkin's) – rituximab.* TA37. London: NICE; 2002.

101. NICE. *Endovascular stent–grafts for the treatment of abdominal aortic aneurysms.* TA167. London: NICE; 2009.

102. NICE. *The use of autologous chondrocyte implantation for the treatment of cartilage defects in knee joints.* TA89. London: NICE; 2005.

103. NICE. *Bortezomib monotherapy for relapsed multiple myeloma.* TA129. London: NICE; 2007.

104. NICE. *Cochlear implants for severe to profound deafness in children and adults.* TA166. London: NICE; 2009.

105. NICE. *Taxanes for the treatment of breast cancer.* TA6. London: NICE; 2000.

106. NICE. *Guidance on the use of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer.* TA33. London: NICE; 2002.

107. NICE. *Cervical cancer – cervical screening (review).* TA69. London: NICE; 2003.

108. NICE. *Taxanes for the treatment of breast cancer.* TA30. London: NICE; 2001.

109. NICE. *Knee joints (defective) – autologous cartilage transplantation: guidance.* TA16. London: NICE; 2000.

110. NICE. *Guidance on the use of laparoscopic surgery for colorectal cancer.* TA17. London: NICE; 2000.

111. NICE. *Laparoscopic surgery for colorectal cancer.* TA105. London: NICE; 2006.

112. NICE. *Advanced breast cancer: NICE guideline.* CG81. London: NICE; 2009.

113. NICE. *Colorectal cancer (advanced) – irinotecan, oxaliplatin and raltitrexed.* TA93. London: NICE; 2005.

114. NICE. *Rheumatoid arthritis – adalimumab, etanercept and infliximab.* TA93. London: NICE; 2007.

115. NICE. *Rituximab for the treatment of relapsed or refractory stage III or IV follicular non-Hodgkin's lymphoma.* TA137. London: NICE; 2008.

116. NICE. *Rheumatoid arthritis – anakinra.* TA72. London: NICE; 2003.

117. NICE. *Rheumatoid arthritis: the management of rheumatoid arthritis in adults.* CG79. London: NICE; 2009.

118. NICE. *Supplementary guidance to the appraisal committees: appraising life-extending, end of life treatment*; 2009. URL: www.nice.org.uk (accessed 18 February 2011).

119. NICE. *Report to the NICE board: update report on the application of the 'end-of-life' supplementary advice in health technology appraisals*; 2009. URL: www.nice.org.uk (accessed 18 February 2011).

120. NICE. *Research recommendations process and methods guide.* London: NICE; 2011.

121. McKenna C, McDaid C, Suekarran S, Hawkins N, Claxton K, Light K, *et al.* Enhanced external counterpulsation for the treatment of stable angina and heart failure: a systematic review and economic analysis. *Health Technol Assess* 2009;**13**(24).

122. NICE. *Clopidogrel in the treatment of non-ST-segment-elevation acute coronary syndrome.* TA80. London: NICE; 2004.

123. NICE. *Unstable angina and NSTEMI: the early management of unstable angina and non-ST-segment-elevation myocardial infarction.* CG94. London: NICE; 2010.

124. NICE. *Omalizumab for the treatment of severe persistent allergic asthma in children aged 6 to 11 years. NICE technology appraisal guidance 201.* London: NICE; 2010. URL: www.nice.org.uk/nicemedia/live/13256/51345/51345.pdf (accessed 22 July 2011).

125. NICE. *Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis.* TA199. London: NICE; 2010.

126. NICE. *Etanercept and infliximab for the treatment of psoriatic arthritis.* TA104. London: NICE; 2006.

127. Claxton K. Exploring uncertainty in cost-effectiveness results. *Pharmacoeconomics* 2008;**926**:781–98.

128. Claxton K, Sculpher M, McCabe C, Briggs AH, Akehurst R, Buxton M, *et al.* Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ* 2005;**14**:339–47.

129. Claxton K, Sculpher M. Using value of information analysis to prioritise health research: some lessons from recent UK experience. *Pharmacoeconomics* 2006;**24**:1055–68.

130. Price MJ, Welton NJ, Briggs AH, Ades AE. Model averaging in the presence of structural uncertainty about treatment effects: influence on treatment decision and expected value of information. *Value Health* 2011;**14**:205–18.

131. Griffin S, Welton N, Claxton K. Exploring the research decision space: the expected value of information for sequential research designs. *Med Decis Making* 2010;**30**:155–62.

132. Bujkiewicz S, Jones HE, Lai MC, Cooper NJ, Hawkins N, Squires H, *et al.* Development of a transparent interactive decision interrogator to facilitate the decision-making process in health care. *Value Health* 2011;**14**:768–76.

133. Kary W, Bergthold L, Talbott M, Aubry W. *Coverage for evidence development: a conceptual frame.* Baltimore, MD: Center for Medical Technology Policy; January 2009.

134. NICE. *Guidance on the selection of prostheses for primary total hip replacement.* TA2. London: NICE; 2000.

135. NICE. *Guidance on the use of temozolomide for the treatment of recurrent malignant glioma (brain cancer).* TA23. London: NICE; 2001.

136. NICE. *Etanercept for the treatment of juvenile idiopathic arthritis.* TA35. London: NICE; 2002.

137. NICE. *Imatinib for chronic myeloid leukaemia.* TA50. London: NICE; 2002.

138. NICE. *Guidance on the use of imatinib for chronic myeloid leukaemia.* TA70. London: NICE; 2003.

139. NICE. *Imatinib for the treatment of unresectable and/or metastatic gastro-intestinal stromal tumours.* TA86. London: NICE; 2004.

140. NICE. *HealOzone for the treatment of tooth decay (occlusal pit and fissure caries and root caries).* TA92. London: NICE; 2005.

141. NICE. *Immunosuppressive therapy for renal transplantation in children and adolescents.* TA99. London: NICE; 2006.

142. NICE. *Donepezil, galantamine, rivastigmine (review) and memantine for the treatment of Alzheimer's disease*. TA111. London: NICE; 2007.TA121

143. NICE. *Carmustine implants and temozolomide for the treatment of newly diagnosed high grade glioma*. TA121. London: NICE; 2007.

144. NICE. *Pemetrexed disodium for the treatment of mesothelioma*. TA135. London: NICE. TA135. London: NICE; 2008.

145. NICE. *Erythropoetin (alpha and beta) and darbepoetin for the treatment of cancer-treatment induced anaemia*. TA142. London: NICE; 2008.

146. NICE. *Adalimumab, etanercept and infliximab for ankylosing spondylitis*. TA143. London: NICE; 2008.

147. NICE. *Spinal cord stimulation for chronic pain of neuropathic or ischaemic origin*. TA159. London: NICE; 2008.

148. Rogowski W, Burch J, Palmer S, Craigs C, Golder S, Woolacott N. The effect of different treatment durations of clopidogrel in patients with non-ST-segment elevation acute coronary syndromes: a systematic review and value of information analysis. *Health Technol Assess* 2009;**13**(31).

149. Shuchman M. Delaying generic competition – corporate payoffs and the future of Plavix. *N Engl J Med* 2006;**355**:1297–300.

150. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ* 2003;**22**:151–85.

151. Drummond MF, Sculpher MJ, Torrance GW, O'Brien B, Stoddart GL. *Methods for the economic evaluation of health care programmes*. 3rd edn. Oxford: Oxford University Press; 2005.

152. Briggs AH, Sculpher MJ, Claxton K. *Decision modelling for health economic evaluation*. Oxford: Oxford University Press; 2006.

153. Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. *Med Decis Making* 1998;**18**:95–109.

154. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ* 1996;**5**:513–24.

155. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision-analytic models: a review and application of methods. *Value Health* 2009;**12**:739–49.

156. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. *Stat Sci* 1999;**14**:382–417.

157. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. *Med Decis Making* 2004;**24**:207–27.

158. Willan A, Eckermann S. Optimal clinical trial design using value of information methods with imperfect implementation. *Health Econ* 2010;**19**:549–61.

159. Allender S, Peto V, Scarborough P, Boxer A, Rayner M. *Morbidity. Coronary heart disease statistics*. 15th edn. London: British Heart Foundation; 2007.

160. Arora R, Chou T, Jain D, Fleishman B, Crawford L, McKiernan T, *et al.* The multicenter study of enhanced external counterpulsation (MUST-EECP): effect of EECP on exercise-induced myocardial ischemia and anginal episodes. *J Am Coll Cardiol* 1999;**33**:1833–40.

161. Arora R, Chou T, Jain D, Fleishman B, Crawford L, McKiernan T, *et al.* Effects of enhanced external counterpulsation on health-related quality of life continue 12 months after treatment: a substudy of the Multicenter Study of Enhanced External Counterpulsation. *J Investig Med* 2002;**50**:25–32.

162. Lawson WE, Hui JCK, Soroff HS, Zheng ZS, Kayden DS, Sasvary D, *et al.* Efficacy of enhanced external counterpulsation in the treatment of angina pectoris. *Am J Cardiol* 1992;**70**:859–62.

163. Lawson WE, Hui JCK, Zheng ZS, Oster Z, Katz JP, Diggs P, *et al.* Three-year sustained benefit from enhanced external counterpulsation in chronic angina pectoris. *Am J Cardiol* 1995;**75**:840–1.

164. Scottish Intercollegiate Guidelines Network (SIGN). *Management of stable angina: a national clinical guideline*. SIGN guideline no. 96. Edinburgh: SIGN; 2007.

165. Sinvhal RM, Gowda RM, Khan IA. Enhanced external counterpulsation for refractory angina pectoris. *Heart* 2003;**89**:830–3.

166. Bonetti PO, Holmes DR, Jr, Lerman A, Barsness GW. Enhanced external counterpulsation for ischemic heart disease: what's behind the curtain? *J Am Coll Cardiol* 2003;**41**:1918–25.

167. O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, *et al. Uncertain judgements. Eliciting experts' probabilities*. Chichester: Wiley; 2006.

168. van Noortwijk JM, Dekker A, Cooke RM, Mazzuchi TA. Expert judgment in maintenance optimization. *IEEE Trans Reliabil* 1992;**41**:427–32.

169. Cooke RM. *Experts in uncertainty: opinion and subjective probability in science*. New York, NY: Oxford University Press; 1991.

170. Briggs A, Mihaylova B, Sculpher M, Hall A, Wolstenholme J, Simoons M, *et al.* Cost effectiveness of perindopril in reducing cardiovascular events in patients with stable coronary artery disease using data from the EUROPA study. *Heart* 2007;**93**:1081–6.

171. Office for National Statistics. *Interim life tables*. 2007. URL: www.gad.gov.uk/ Demography%20Data/Life%20Tables/Interim_life_tables.html (accessed 23 June 2011).

172. Michaels A, Barsness G, Soran O, Kelsey S, Kennard E, Hui J, *et al.* Frequency and efficacy of repeat enhanced external counterpulsation for stable angina pectoris (from the International EECP Patient Registry). *Am J Cardiol* 2005;**95**:394–7.

173. Vasogenics price list (effective from April 2007).

174. National Institute for Health and Clinical Excellence Centre for Health Technology Evaluation. *Evaluation pathway programme for medical technologies*. London: NICE; 2010. URL: www.nice.org.uk/media/5C8/62/ProductsConsideredByMTAC220211.pdf (accessed 10 December 2010).

175. Main C, Palmer S, Griffin S, Jones L, Orton V, Sculpher M, *et al.* Clopidogrel used in combination with aspirin compared with aspirin alone in the treatment of non-ST-segment-elevation acute coronary syndromes: a systematic review and economic evaluation. *Health Technol Assess* 2004;**8**(40).

176. Peters RJ, Mehta SR, Fox KAA, Zhao F, Lewis BS, Kopecky SL, *et al.* Effects of aspirin dose when used alone or in combination with clopidogrel in patients with acute coronary syndromes: observations from the Clopidogrel in Unstable angina to prevent Recurrent Events (CURE) study. *Circulation* 2003;**108**:1682–7.

177. Morais J. Insights from CURE: using clopidogrel on top of standard therapy. *Cerebrovasc Dis* 2002;**13**(Suppl. 1):17–21.

178. Yusuf S, Mehta SR, Zhao F, Gersh BJ, Commerford PJ, Blumenthal M, *et al.* Early and late effects of clopidogrel in patients with acute coronary syndromes. *Circulation* 2003;**107**:966–72.

179. Budaj A, Yusuf S, Mehta SR, Fox KAA, Tognoni G, Zhao F, *et al.* Benefit of clopidogrel in patients with acute coronary syndromes without ST-segment elevation in various risk groups. *Circulation* 2002;**106**:1622–6.

180. Yusuf S, Zhao F, Mehta SR, Chrolavicius S, Tognoni G, Fox KAA, *et al.* Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N Engl J Med* 2001;**345**:494–502.

181. Keltai M, Tonelli M, Mann JF, Sitkei E, Lewis BS, Hawken S, *et al.* Renal function and outcomes in acute coronary syndrome: impact of clopidogrel. *Eur J Cardiovasc Prev Rehabil* 2007;**14**:312–18.

182. Scottish Intercollegiate Guidelines Network (SIGN). *Acute coronary syndromes: a national clinical guideline.* SIGN guideline no. 93. Edinburgh: SIGN; 2007.

183. Karnon J, Bakhai A, Brennan A, Pandor A, Flather M, Warren E, *et al.* A cost–utility analysis of clopidogrel in patients with non-ST-segment-elevation acute coronary syndromes in the UK. *Int J Cardiol* 2006;**109**:307–16.

184. British Medical Association and Royal Pharmaceutical Society of Great Britain. *British national formulary.* No. 54, September 2007. London: BMA and RPS; 2007.

185. Collinson, J, Flather MD, Fox KAA, Findlay I, Rodrigues E, Dooley P, *et al.* Clinical outcomes, risk stratification and practice patterns of unstable angina and myocardial infarction without ST elevation: Prospective Registry of Acute Ischaemic Syndromes in the UK (PRAIS-UK). *Eur Heart J* 2000;**21**:1450–7.

186. Lanier B, Bridges T, Kulus M, Taylor AF, Berhane I, Vidaurre CF. Omalizumab for the treatment of exacerbations in children with inadequately controlled allergic (IgE-mediated) asthma. *J Allergy Clin Immunol* 2009;**124**:1210–16.

187. Asthma UK. *For journalists: key facts & statistics.* 2010. URL: www.asthma.org.uk/news_media/media_resources/for_journalists_key.html (accessed 3 June 2011).

188. Novartis Pharmaceuticals UK Ltd. *Xolair (omalizumab) for severe persistent allergic asthma in children aged 6–<12 years. Manufacturer submission of evidence.* Single technology appraisal. London: NICE; 2010.

189. Scottish Intercollegiate Guidelines Network (SIGN). *British guideline on the management of asthma; a national clinical guideline.* Edinburgh: SIGN; 2009.

190. Walker S, Burch J, McKenna C, Wright K, Griffin S, Woolacott N. *Omalizumab for the treatment of severe persistent allergic asthma in children aged 6 to 11 years.* Evidence review group report. London: NICE; 2010.

191. Dewilde S, Turk F, Tambour M, Sandstrom T. The economic value of anti-IgE in severe persistent, IgE-mediated (allergic) asthma patients: adaptation of INNOVATE to Sweden. *Curr Med Res Opin* 2006;**22**:1765–76.

192. Humbert M, Beasley R, Ayres J, Slavin R, Hebert J, Bousquet J, *et al.* Benefits of omalizumab as add-on therapy in patients with severe persistent asthma who are inadequately controlled despite best available therapy (GINA 2002 step 4 treatment): INNOVATE. *Allergy* 2005;**60**:309–16.

193. Watson L, Turk F, James P, Holgate ST. Factors associated with mortality after an asthma admission: a national United Kingdom database analysis. *Respir Med* 2007;**101**:1659–64.

194. Lloyd A, Price D, Brown R. The impact of asthma exacerbations on health-related quality of life in moderate to severe asthma patients in the UK. *Prim Care Respir J* 2007;**16**:22–7.

195. Kyle S, Chandler D, Griffiths CE, Helliwell P, Lewis J, McInnes I, *et al.* Guideline for anti-TNF-alpha therapy in psoriatic arthritis. *Rheumatology* 2005;**44**:390–7. [Erratum published in *Rheumatology* 2005;**44**:569.]

196. Symmons DP, Barrett EM, Bankhead CR, Scott DG, Silman AJ. The incidence of rheumatoid arthritis in the United Kingdom: results from the Norfolk Arthritis Register. *Rheumatology* 1994;**33**:735–9.

197. Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. *Lancet* 2000;**356**:385–90.

198. Mease PJ, Kivitz AJ, Burch FX, Siegel EL, Cohen SB, Ory P, *et al.* Etanercept treatment of psoriatic arthritis: safety, efficacy, and effect on disease progression. *Arthritis Rheum* 2004;**50**:2264–72.

199. Mease PJ, Kivitz AJ, Burch FX, Siegel EL, Cohen SB, Ory P, *et al.* Continued inhibition of radiographic progression in patients with psoriatic arthritis following 2 years of treatment with etanercept. *J Rheumatol* 2006;**33**:712–21.

200. Fleischmann R, Baumgartner SW, Weisman MH, Liu T, White B, Peloso P. Long term safety of etanercept in elderly subjects with rheumatic diseases. *Ann Rheum Dis* 2006;**65**:379–84.

201. Lebwohl M, Gottlieb A, Goffe BS, Jahreis A. Etanercept in psoriatic arthritis: sustained improvement in skin and joint disease and inhibition of radiographic progression at 2 years. *J Am Acad Dermatol* 2005;**52**(Suppl. 2):AB8.

202. Wanke LA, Mease PJ, Gottlieb AB. Sustained improvement in functional status and vitality of psoriatic arthritis patients treated with etanercept. *J Invest Dermatol* 2004;**122**:A64.

203. Mease PJ, Woolley JM, Singh A, Tsuji W, Dunn M, Chiou CF. Patient-reported outcomes in a randomized comparison of etanercept and placebo for treatment of psoriatic arthritis. *Ann Rheum Dis* 2006;**65**(Suppl. 2):535.

204. Centocor. *A multicenter placebo-controlled, double-blind, randomised study of anti-TNF chimeric monoclonal antibody (cA2, infliximab) in patients with active psoriatic arthritis (IMPACT): protocol no. P02114* [industry submission]. Malvern, PA: Centocor; 2003.

205. Schering-Plough Ltd. *Remicade in the treatment of psoriatic arthritis in the United Kingdom: a submission to the National Institute for Clinical Excellence* [industry submission]. Kenilworth, NJ: Schering-Plough Ltd; 2004.

206. Antoni CE, Kavanaugh A, Kirkham B, Tutuncu Z, Burmester GR, Schneider U, *et al.* Sustained benefits of infliximab therapy for dermatologic and articular manifestations of psoriatic arthritis: results from the infliximab multinational psoriatic arthritis controlled trial (IMPACT). *Arthritis Rheum* 2005;**52**:1227–36.

207. Antoni CE, Kavanaugh A, van der Heijde D, Beutler A, Keenan G, Zhou B, *et al.* Two-year efficacy and safety of infliximab treatment in patients with active psoriatic arthritis: findings of the Infliximab Multinational Psoriatic Arthritis Controlled Trial (IMPACT). *J Rheumatol* 2008;**35**:869–76.

208. Kavanaugh A, Antoni CE, Gladman D, Wassenberg S, Zhou B, Beutler A, *et al.* The Infliximab Multinational Psoriatic Arthritis Controlled Trial (IMPACT): results of radiographic analyses after 1 year. *Ann Rheum Dis* 2006;**65**:1038–43.

209. Kavanaugh A, Kruegar GG, Birbara C, Halter D, de Vlam K, Zhou B, *et al.* Treatment with infliximab is associated with 'major clinical response' in psoriatic arthritis patients treated with infliximab: analysis of two double-blind placebo controlled trials. *Arthritis Rheum* 2005;**52**(Suppl. 12):4065.

210. Kavanaugh A, Kruegar GG, Birbara C, Halter D, Geusens P, de Vlam K, *et al.* Treatment with infliximab is associated with 'major clinical response' in psoriatic arthritis patients treated with infliximab: analysis of two double-blind placebo controlled trials. *J Invest Dermatol* 2007;**127**:1824.

211. Antoni CE, Kavanaugh A, Kirkham B, Tutuncu ZN, Burmester G, Schneider U, *et al.* 2-year data: infliximab maintains clinical response in psoriatic arthritis patients – data from the Infliximab Multinational Psoriatic Arthritis Controlled Trial (IMPACT). *Arthritis Rheum* 2005;**52**(Suppl. S):S209.

212. Antoni CE, Kavanaugh A, Gladman D, Wassenberg S, Zhou B, Beutler A, *et al.* The infliximab multinational psoriatic arthritis controlled trial (IMPACT): results of radiographic analyses after 1 year. *Ann Rheum Dis* 2005;**64**(Suppl. 3):107.

213. Antoni CE, Kavanaugh A, Kirkham B, Burmester G, Manger B, Tutuncu Z, *et al.* The infliximab multinational psoriatic arthritis controlled trial (IMPACT): substantial efficacy on synovitis and psoriatic lesions with or without concomitant DMARD therapy. *Ann Rheum Dis* 2003;**62**(Suppl. 1):90.

214. Antoni CE, Kavanaugh A, Kirkham B, Burmester G, Manger B, Schneider U, *et al.* The infliximab multinational psoriatic arthritis controlled trial (IMPACT1): the concomitant use of DMARDs does not influence the efficacy and safety of infliximab over a one year period. *Ann Rheum Dis* 2004;**63**(Suppl. 1):411–12.

215. Antoni C, Krueger G, de Vlam K, Birbara C, Beutler A, Guzzo C, *et al.* Infliximab improves signs and symptoms of psoriatic arthritis: results of the IMPACT 2 trial. *Ann Rheum Dis* 2005;**64**:1150–7.

216. van der Heijde D, Kavanaugh A, Gladman DD, Antoni C, Krueger GG, Guzzo C, *et al.* Infliximab inhibits progression of radiographic damage in patients with active psoriatic arthritis through one year of treatment: results from the induction and maintenance psoriatic arthritis clinical trial 2. *Arthritis Rheum* 2007;**56**:2698–707.

217. Kavanaugh A, Krueger GG, Beutler A, Guzzo C, Zhou B, Dooley LT, *et al.* Infliximab maintains a high degree of clinical response in patients with active psoriatic arthritis through 1 year of treatment: results from the IMPACT 2 trial. *Ann Rheum Dis* 2007;**66**:498–505.

218. Kavanaugh A, Antoni C, Mease P, Gladman D, Yan S, Bala M, *et al.* Effect of infliximab therapy on employment, time lost from work, and productivity in patients with psoriatic arthritis. *J Rheumatol* 2006;**33**:2254–9.

219. Kavanaugh A, Antoni C, Krueger GG, Yan S, Bala M, Dooley LT, *et al.* Infliximab improves health related quality of life and physical function in patients with psoriatic arthritis. *Ann Rheum Dis* 2006;**65**:471–7.

220. Papp K, Menter A, Antoni C, Kavanaugh A, Gottlieb AB, Paulin Y, *et al.* Rapid and persistent improvement of psoriasis in patients with active psoriatic arthritis treated with infliximab: 24-week results of the impact 2 trial. *J Eur Acad Dermatol Venereol* 2004;**18**:783.

221. Kavanauagh A, Krueger GG, De Vlam K, Birbara C, Beutler A Guzzo C, *et al.* Infliximab significantly improves joint and skin involvement in psoriatic arthritis to a substantial extent and irrespective of baseline joint involvement or MTX use: analysis of clinical response from the IMPACT2 trial. *Arthritis Rheum* 2004;**50**(Suppl. S):S617.

222. Mease PJ, Gladman DD, Ritchlin CT, Ruderman EM, Steinfeld SD, Choy EH, *et al.* Adalimumab for the treatment of patients with moderately to severely active psoriatic arthritis: results of a double-blind, randomized, placebo-controlled trial. *Arthritis Rheum* 2005;**52**:3279–89.

223. Mease PJ, Ory P, Sharp JT, Ritchlin CT, Van den Bosch F, Wellborne F, *et al.* Adalimumab for long-term treatment of psoriatic arthritis: 2-year data from the Adalimumab Effectiveness in Psoriatic Arthritis Trial (ADEPT). *Ann Rheum Dis* 2009;**68**:702–9.

224. Gladman DD, Mease PJ, Ritchlin CT, Choy EH, Sharp JT, Ory PA, *et al.* Adalimumab for long-term treatment of psoriatic arthritis: forty-eight week data from the adalimumab effectiveness in psoriatic arthritis trial. *Arthritis Rheum* 2007;**56**:476–88.

225. Gladman DD, Mease PJ, Cifaldi MA, Perdok RJ, Sasso E, Medich J. Adalimumab improves joint-related and skin-related functional impairment in patients with psoriatic arthritis: patient-reported outcomes of the Adalimumab Effectiveness in Psoriatic Arthritis Trial. *Ann Rheum Dis* 2007;**66**:163–8.

226. Choy E, Gladman D, Sasso E. Efficacy of adalimumab by disease duration in psoriatic arthritis: subanalysis of ADEPT. *J Am Acad Dermatol* 2007;**56**:AB186.

227. Gladman D, Mease P, Ritchlin C, Okun M. PASI-100 is associated with better dermatology-specific patient-reported outcomes compared with PASI-75e99 in adalimumab-treated patients with psoriatic arthritis: subanalysis of ADEPT. *J Am Acad Dermatol* 2007;**56**:AB185.

228. Gladman D, Mease P, Kavanaugh A, Weinberg M. Adalimumab is efficacious in treating skin disease in psoriatic arthritis: subanalysis of moderate versus severe psoriasis in the ADEPT trial. *J Am Acad Dermatol* 2006;**54**:AB10–11.

229. Mease P, Gladman D, Ritchlin C, Weinberg M. Clinical efficacy, safety and inhibition of joint destruction of adalimumab in the treatment of moderate to severe psoriatic arthritis: 48-week results of the ADEPT trial. *J Am Acad Dermatol* 2006;**54**:AB214.

230. Mease P, Sharp J, Ory P, Gladman D, Ritchlin C, Choy E, *et al.* Adalimumab treatment effects on radiographic progression of joint disease in patients with psoriatic arthritis: results from ADEPT. *J Eur Acad Dermatol Venereol* 2005;**19**(Suppl. 2):163.

231. Genovese MC, Mease PJ, Thomson GT, Kivitz AJ, Perdok RJ, Weinberg MA, *et al.* Safety and efficacy of adalimumab in treatment of patients with psoriatic arthritis who had failed disease modifying antirheumatic drug therapy. *J Rheumatol* 2007;**34**:1040–50.

232. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics* 2008;**26**:753–67.

233. Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, *et al.* Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. *Health Technol Assess* 2011;**15**(10).

234. Kobelt G, Jönsson L, Lindgren P, Young A, Eberhardt K. Modelling the progression of rheumatoid arthritis: a two-country model to estimate costs to estimate costs and consequences of rheumatoid arthritis. *Arthritis Rheum* 2002;**46**:2310–19.

235. Department of Health. *Reference costs 2007–2008*. London: Department of Health; 2009. URL:www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyandGuidance/DH_09895.

236. Symmons DP, Silman AJ. The Norfolk Arthritis Register (NOAR). *Clin Exp Rheumtol* 2003;**21**:S94–9.

237. Yang H, Epstein D, Bojke L, Craig D, Light K, Bruce I, *et al. Evidence review group's report: golimumab for the treatment of psoriatic arthritis.* London: NIHR HTA; 2010.

238. Brennan A, Kharroubi S, O'Hagan A, Chilcott J. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Med Decis Making* 2007;**27**:448–70

Appendix 1

Literature review, searches and results

Initial set of key papers

Policy papers

1. Cooksey D. *A review of UK health research funding*. London: The Stationery Office; 2006.

2. Centers for Medicare & Medicaid Services. *National coverage determinations with data collection as a condition of coverage: coverage with evidence development*. 2006. URL: www.cms.gov/Medicare/Coverage/Coverage-with-Evidence-Development/index.html.

3. House of Commons Health Committee, National Institute for Health and Clinical Excellence. Report no. HC 27–1. London: The Stationary Office; 2008. Chapter 6.

4. Mechanic R, Zinner D. *Optimizing information under coverage with evidence development*. Health Industry Forum Policy Brief. 24 October 2007. URL: http://healthforum.brandeis.edu/meetings/materials/2007-24-Oct./Policy-Brief.pdf.

5. NICE Citizens Council. *Only in research*. London: NICE; 2007.

6. NHS Quality Improvement Scotland. *Coverage with evidence development in NHS Scotland*. Discussion paper. 2008. URL: www.Healthcareimprovementscotland.org/previous_resources/hta_report/evidence_development.aspx.

7. Office of Fair Trading. *The pharmaceutical price regulation scheme. An OFT market study*. London: OFT; 2007. Chapters 4 and 5.

8. Office for Life Sciences, HM Government. *Life sciences blueprint*. A statement from the Office for Life Sciences. July 2009. URL: www.medilink.co.uk/Libraries/connect/Life-sciences-2009-blueprint.sflb.ashx.

Discussion and application papers

9. Medicare's coverage with evidence development: a policy-making tool in evolution. *J Oncol Pract* 2007;**3**:296–301.

10. Briggs A, Ritchie K, Fenwick E, Chalkidou K, Littlejohns P. Access with evidence development in the UK: past experience, current initiatives and future potential. *Pharmacoeconomics* 2010;**28**:163–70.

11. Carino T, Sheingold S, Tunis S. Using clinical trials as a condition of coverage: lessons from the National Emphysema Treatment Trial. *Clin Trials* 2004;**1**:108–21.

12. Chalmers R. Addressing uncertainties about the effects of treatments offered to NHS patients: whose responsibility? *J R Soc Med* 2007;**100**:440–1

13. Chalkidou K. *'Only in research': a polite 'no' or a valuable policy option?* Birmingham: NICE; 2006.

14. Chalkidou K, Hoy A, Littlejohns P. Making a decision to wait for more evidence: when the National Institute for Health and Clinical Excellence recommends a technology only in the context of research. *J R Soc Med* 2007;**100**:453–60.

15. Chalkidou K, Lord J, Fischer A, Littlejohns P. Evidence-based decision making: when should we wait for more information? *Health Aff* 2008;**27**:1642–53.

16. Dhalla IF, Garner S. Perspectives on the National Institute for Health and Clinical Excellence's recommendations to use technologies only in research. *Int J Technol Assess Health Care* 2009;**25**:272–80.

17. Garrison L, Sullivan S, Carlson J, Bresnahan B, Carlson R, Neumann P, *et al. Performance-based risk-sharing reimbursement agreements for new medical products*. Washington, DC: University of Washington; 2008.

18. Kary W, Bergthold L, Talbott M, Aubry W. Coverage for evidence development: a conceptual frame. Baltimore, MD: Center for Medical Technology Policy; 2009.

19. Macdonald F. Coverage with evidence development: a personal, industry view. Presentation slides. 19 September 2008. URL: www.docstoc.com/docs/105704904/ Coverage-With-Evidence-Development.

20. McCabe CJ, Stafinski T, Edlin R, Menon D. Access with evidence development schemes: a framework for description and evaluation. *Pharmacoeconomics* 2010;**28**:143–52.

21. Menon D, McCabe CJ, Stafinski T, Edlin R, signatories to the Consensus Statement. Principles of design of access with evidence development approaches: a consensus statement from the Banff Summit. *Pharmacoeconomics* 2010;**28**:109–11.

22. Miller FG, Pearson SD. Coverage with evidence development: ethical issues and policy implications. *Med Care* 2008;**46**:746–51.

23. Mohr P, Tunis SR. Access with evidence development: the US experience. *Pharmacoeconomics* 2010;**28**:153–62.

24. Hutton J, Trueman P, Henshall C. Coverage with evidence development: an examination of conceptual and policy issues. *Int J Technol Assess Health Care* 2007;**23**:425–35.

25. Pearson SD, Miller FG, Emanuel EJ. Medicare's requirement for research participation as a condition of coverage: is it ethical? *JAMA* 2006;**296**:988–91.

26. Stafinski T, McCabe C, Menon D. Funding the unfundable: mechanisms for managing uncertainty in decisions on the introduction of new and innovative technologies into healthcare systems. *Pharmacoeconomics* 2010;**28**:113–42.

27. Towse A, Garrison LP. Can't get no satisfaction? Will pay for performance help?: toward an economic framework for understanding performance-based risk-sharing agreements for innovative medical products. *Pharmacoeconomics* 2010;**28**:93–102.

28. Trueman P. Coverage with evidence development: applications and issues. *Int J Technol Assess Health Care* 2010;**26**:79–85.

29. Tunis SR, Pearson SD. Coverage options for promising technologies: Medicare's 'coverage with evidence development'. *Health Aff* 2006;**25**:1218–30.

30. Tunis SR, Chalkidou K. Coverage with evidence development: a very good beginning, but much to be done. Commentary to Hutton *et al. Int J Technol Assess Health Care* 2007;**23**:432–5.

31. Tunis S, Whicher D. The National Oncologic PET Registry: lessons learned for coverage with evidence development. *J Am Coll Radiol* 2009;**6**:360–5.

Search strategies

Traditional literature searches
Scoping search
Ovid MEDLINE(R) –1996 to week 2 February 2010
Searched via Ovid, 18 February 2010.

Medical subject heading (MeSH) terms:

1. Health Policy/ (24,246)
2. Decision Making/ (32,178)
3. exp policy making/ (10,169)
4. Insurance Coverage/ (5950)
5. Insurance, Health/ (8791)
6. Insurance, Health, Reimbursement/ (3860)
7. Insurance Benefits/ (1246)
8. Evidence-Based Medicine/ (35,935)
9. uncertainty/ (2644)
10. Medicare/ (14,219)
11. Reimbursement Mechanisms/ (4658)
12. Technology Assessment, Biomedical/ (4437)
13. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 (132,518)

Key terms:

14. ((access$ or cover$ or approv$ or reimburs$ or fund$ or licens$ or licenc$ or list$) adj (with or in) adj (evidence or condition$)).ti,ab. (30)
15. "only in research".ti,ab. (26)
16. "only with research".ti,ab. (2)
17. "in the context of research".ti,ab. (125)
18. "condition$ of coverage$".ti,ab. (57)
19. "condition$ of reimburse$".ti,ab. (13)
20. 14 or 15 or 16 or 17 or 18 or 19 (251)

Words near to words terms:

21. ((interim or condition$ or restrict$ or evidence) adj2 (fund$ or reimburse$ or cover$ or approv$ or list$ or access$ or licens$ or licenc$)).ti,ab. (3602)
22. 13 and 20 (82)
23. limit 22 to yr = "1999 - 2010" (77)
24. limit 20 to yr = "1999 - 2010" (223)
25. 13 and 21 (443)
26. limit 25 to yr = "1999 - 2010" (404)
27. limit 21 to yr = "1999 - 2010" (3082)

Results were then deduplicated so that only additional records not found by the preceding search were added to the library.

Search	Results	Results after deduplication
Key terms AND MeSH	77	77
Key terms only	223	146
near to terms AND MeSH	404	341
near to terms only	3082	2648

Results were reviewed and the final strategy adopted was to use only the 'Key terms' searches, with several additions. The 'near to terms' retrieved few relevant records and additional key terms were added to ensure that these were captured. The relatively low number of results meant that combining 'Key terms' with MeSH terms to reduce the results set was not necessary.

Final search strategies
Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) –1950 to present
Searched via Ovid, 24 February 2010.

1. ((access$ or cover$ or approv$ or reimburs$ or fund$ or licens$ or licenc$ or list$) adj (with or in) adj (evidence or condition$)).ti,ab. (45)
2. "only in research".ti,ab. (43)
3. "only with research".ti,ab. (2)
4. "in the context of research".ti,ab. (158)
5. "condition$ of coverage$".ti,ab. (82)
6. "condition$ of reimburse$".ti,ab. (19)
7. (condition$ adj fund$).ti,ab. (21)
8. 1 or 2 or 3 or 4 or 5 or 6 or 7 (365)
9. limit 8 to yr = "1999 - 2010" (258)

EMBASE –1980 to week 7 2010
Searched via Ovid, 24 February 2010.

1. ((access$ or cover$ or approv$ or reimburs$ or fund$ or licens$ or licenc$ or list$) adj (with or in) adj (evidence or condition$)).ti,ab. (36)
2. "only in research".ti,ab. (38)
3. "only with research".ti,ab. (1)
4. "in the context of research".ti,ab. (108)
5. "condition$ of coverage$".ti,ab. (39)
6. "condition$ of reimburse$".ti,ab. (13)
7. (condition$ adj fund$).ti,ab. (20)
8. 1 or 2 or 3 or 4 or 5 or 6 or 7 (252)
9. limit 8 to yr = "1999 - 2010" (182)

The Cochrane Library [Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE), HTA database, NHS Economic Evaluation Database (NHS EED)]
Searched via www3.interscience.wiley.com/tools/citex, 24 February 2010.

#1 ((access* or cover* or approv* or reimburs* or fund* or licens* or licenc* or list*) NEXT (with or in) NEXT (evidence or condition$)) (1)

#2 "only in research" (1)
#3 "only with research" (0)
#4 "in the context of research" (4)
#5 "condition* of coverage*" (0)
#6 "condition* of reimburse*" (0)
#7 condition* NEXT fund* (22)
#8 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7), from 1999 to 2010 (22)

EconLit –1969 to February 2010

Searched via Ovid, 24 February 2010.

1. ((access$ or cover$ or approv$ or reimburs$ or fund$ or licens$ or licenc$ or list$) adj (with or in) adj (evidence or condition$)).ti,ab. (0)
2. "only in research".ti,ab. (0)
3. "only with research".ti,ab. (0)
4. "in the context of research".ti,ab. (15)
5. "condition$ of coverage$".ti,ab. (12)
6. "condition$ of reimburse$".ti,ab. (4)
7. (condition$ adj fund$).ti,ab. (7)
8. 1 or 2 or 3 or 4 or 5 or 6 or 7 (38)
9. limit 8 to yr = "1999 - 2010" (37)

Cumulative Index to Nursing and Allied Health Literature

Searched via EBSCO, 24 February 2010.
Results limited to 1999 onwards.

1. TX "access* with evidence" or TX "coverage* with evidence" or TX "approv* with evidence" or TX "reimburs* with evidence" or TX "fund* with evidence" or TX "licenc* with evidence" or TX "licens* with evidence" and TX "list* with evidence"
2. TX "access* in evidence" or TX "coverage* in evidence" or TX "approv* in evidence" or TX "reimburs* in evidence" or TX "fund* in evidence" or TX "licenc* in evidence" or TX "licens* in evidence" and TX "list* in evidence"
3. TX "access* with condition*" or TX "coverage* with condition*" or TX "approv* with condition*" or TX "reimburs* with condition*" or TX "fund* with condition*" or TX "licenc* with condition*" or TX "licens* with condition*" and TX "list* with condition*"
4. TX "access* in condition*" or TX "coverage* in condition*" or TX "approv* in condition*" or TX "reimburs* in condition*" or TX "fund* in condition*" or TX "licenc* in condition*" or TX "licens* in condition*" and TX "list* in condition*"
5. TX "only in research"
6. TX "only with research"
7. TX "in the context of research"
8. TX "condition* of coverage*"
9. TX "condition* of reimburse*"
10. TX "condition* fund*"
11. 1 or 2 or 3 or 3 or 5 or 6 or 7 or 8 or 9 or 10 (378)

Grey literature searches
New York Academy of Medicine Grey Literature Report

Searched via www.nyam.org/library/pages/grey_literature_report, 8 March 2010.

""ti,phr: access with evidence or ti,phr: coverage with evidence or ti,phr: approval with evidence or ti,phr: approved with evidence or ti,phr: reimbursement with evidence or ti,phr: reimbursed

with evidence or ti,phr: reimburse with evidence or ti,phr: funding with evidence or ti,phr: funded with evidence or ti,phr: licenced with evidence or ti,phr: licensed with evidence or ti,phr: licensing with evidence or ti,phr: licencing with evidence or ti,phr: licence with evidence or ti,phr: license with evidence or ti,phr: list with evidence or ti,phr: lists with evidence or ti,phr: listed with evidence or ti,phr: access in evidence or ti,phr: coverage in evidence or ti,phr: approval in evidence or ti,phr: reimbursement in evidence or ti,phr: reimbursed in evidence or ti,phr: reimburse in evidence or ti,phr: funding in evidence or ti,phr: funded in evidence or ti,phr: licenced in evidence or ti,phr: licensed in evidence or ti,phr: licensing in evidence or ti,phr: licencing in evidence or ti,phr: licence in evidence or ti,phr: license in evidence or ti,phr: list in evidence or ti,phr: lists in evidence or ti,phr: listed in evidence mc-ccode:GREYLIT" "
"ti,phr: access with conditions or ti,phr: coverage with conditions or ti,phr: approval with conditions or ti,phr: approved with conditions or ti,phr: reimbursement with conditions or ti,phr: reimbursed with conditions or ti,phr: reimburse with conditions or ti,phr: funding with conditions or ti,phr: funded with conditions or ti,phr: licenced with conditions or ti,phr: licensed with conditions or ti,phr: licensing with conditions or ti,phr: licencing with conditions or ti,phr: licence with conditions or ti,phr: license with conditions or ti,phr: list with conditions or ti,phr: lists with conditions or ti,phr: listed with conditions mc-ccode:GREYLIT"
""ti,phr: access in conditions or ti,phr: coverage in conditions or ti,phr: approval in conditions or ti,phr: approved in conditions or ti,phr: reimbursement in conditions or ti,phr: reimbursed in conditions or ti,phr: reimburse in conditions or ti,phr: funding in conditions or ti,phr: funded in conditions or ti,phr: licenced in conditions or ti,phr: licensed in conditions or ti,phr: licensing in conditions or ti,phr: licencing in conditions or ti,phr: licence in conditions or ti,phr: license in conditions or ti,phr: list in conditions or ti,phr: lists in conditions or ti,phr: listed in conditions mc-ccode:GREYLIT" "
"ti,phr: only in research or ti,phr: only with research or ti,phr: OIR mc-ccode:GREYLIT"
"ti,phr: in the context of research or ti,phr: condition of coverage or ti,phr: conditions of coverage or ti,phr: condition of reimbursement or ti,phr: conditions of reimbursement or ti,phr: conditional funding or ti,phr: conditionally funded or ti,phr: conditional funds mc-ccode:GREYLIT"

IDEAS – University of Connecticut Department of Economics

Searched via http://ideas.repec.org/search.html, 9 March 2010.
Phrase searching restricted to papers, chapters, books and software components, for following phrases:

access with evidence
coverage with evidence
approval with evidence
approved with evidence
reimbursement with evidence
reimbursed with evidence
reimburse with evidence
funding with evidence
funded with evidence
licenced with evidence
licensed with evidence
licensing with evidence
licencing with evidence
licence with evidence
license with evidence
list with evidence
lists with evidence

listed with evidence
access in evidence
coverage in evidence
approval in evidence
approved in evidence
reimbursement in evidence
reimbursed in evidence
reimburse in evidence
funding in evidence
funded in evidence
licenced in evidence
licensed in evidence
licensing in evidence
licencing in evidence
licence in evidence
license in evidence
list in evidence
lists in evidence
listed in evidence
access with conditions
coverage with conditions
approval with conditions
approved with conditions
reimbursement with conditions
reimbursed with conditions
reimburse with conditions
funding with conditions
funded with conditions
licenced with conditions
licensed with conditions
licensing with conditions
licencing with conditions
licence with conditions
license with conditions
list with conditions
lists with conditions
listed with conditions
access in conditions
coverage in conditions
approval in conditions
approved in conditions
reimbursement in conditions
reimbursed in conditions
reimburse in conditions
funding in conditions
funded in conditions
licenced in conditions
licensed in conditions
licensing in conditions
licencing in conditions
licence in conditions
license in conditions

list in conditions
lists in conditions
listed in conditions
in the context of research
condition of coverage
conditions of coverage
condition of reimbursement
conditions of reimbursement
conditional funding
conditionally funded
conditional funds

Note: the phrases 'only in research' and 'only with research' were not searchable as 'only', 'in' and 'with' are all stopwords and stopwords are applied even when phrase searching is selected. Search therefore searches for 'research' only and retrieves over 24,000 records.

In total, 64 results were saved to MRC OIR Project/Literature Search/IDEAS results.docx. These were screened and none was found to be relevant.

Results

Database	Results	Results after deduplication
MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations	258	253
EMBASE	182	50
CDSR	2	2
DARE	7	7
HTA	2	2
NHS EED	9	9
EconLit	37	35
CINAHL	378	317
New York Academy of Medicine	0	0
IDEAS	0	0
Total	875	675

Websites searched for policy documents

Organisation	Web link	Search terms
UK sources		
Department of Health	www.dh.gov.uk/en/Publicationsandstatistics/index.htm	Coverage with evidence Only in research Free scan of website
NICE	Via www.evidence.nhs.uk	Coverage with evidence Only in research Free scan of website
HM Treasury	http://archive.treasury.gov.uk/	Free scan of subject headings – linked to news releases Free scan of current news releases and reports
House of Commons Health Committee	www.parliament.uk/parliamentary_committees/health_committee/health_committee_reports_and_publications.cfm	Free scan of the reports and publication web pages

Organisation	Web link	Search terms
NICE Citizens Council	www.nice.org.uk/aboutnice/howwework/citizenscouncil/reports.jsp	All of the available reports (13) were scanned
Office of Fair Trading	www.oft.gov.uk/advice_and_resources/publications/	Free scan of reports section of publications
Office for Life Sciences	www.bis.gov.uk/publications	Only in research Coverage with evidence Free scan by subject
NHS Scotland and NHS Quality Improvement Scotland	www.show.scot.nhs.uk/publications/publication.asp www.nhshealthquality.org/nhsqis/CCC_FirstPage.jsp	Only in research Coverage with evidence Free scan of publications
Health Industry Forum	http://healthforum.brandeis.edu/publications/index.html	Free scan of publications list
Department for Business, Innovation & Skills	www.berr.gov.uk/publications/reports/index.html	Only in research Coverage with evidence Managed entry Also free scan by subject
Non-UK sources		
Centers for Medicare & Medicaid	www.cms.hhs.gov/home/rsds.asp	Free scan of the Research, Statistics, Data & Systems web pages
Center for Medical Technology Policy	www.cmtpnet.org/recent-articles	Free scan of publications list
Organisation for Economic Co-operation and Development (OECD)	www.oecd.org/publications/0,3353,en_2649_201185_1_1_1_1_1,00.html	Only in research Coverage with evidence Free scan by 'Health' and 'Economics' subjects
European network for Health Technology Assessment (EUnetHTA)	www.eunethta.eu	Free scan of website and publications section

Summary of search results

Table 27 summarises the total number of references identified, the number identified for full screening and the final number from each source included in the review. From the total number of references retrieved by the searches, potentially relevant references were initially identified by screening titles and abstracts. Full copies of all of these references were then obtained and subjected to full screening. References were included in the review if they explicitly discussed any of the following issues relating to OIR/AWR: terminology, taxonomy, general themes and/or principles including evidence collection, investment and reversal costs, changing prices and ethical considerations. In addition, case study applications were included if any of the above issues were discussed in detail as part of the application or interpretation of the results.

Of the 675 references identified using the traditional systematic search, full papers were obtained for only 56 despite a relatively inclusive approach being adopted (any mention of OIR/AWR or related terms in the titles or abstracts). Of these, 27 references were eventually included as part of the review (18 of which had been previously identified as part of the initial set of key papers used for the citation searches). Although this is fewer than the number of references identified in the traditional literature in the Stafinski *et al.* review[8] (*n* = 68), it simply reflects the fact that the literature discussing general principles and themes relevant to OIR/AWR is considerably smaller than the literature reporting specific applications and case studies.

TABLE 27 References identified by source

Source	Total no. of references retrieved	No. of relevant references (full screening)	No. of references included in the review
Original list of key references	31		
Traditional literature search	675	56	27 (18 of which already in original key paper list)
Grey literature search	64	0	0
Citation search	122	16	7 (2 of which already identified from original key paper list)
Policy websites	280	14	11 (9 of which already identified from original key paper list and systematic searches)
Advisory Group references	7	5	5 (4 of which are already in original key paper list)
Other sources	HTAi interest group = 5	HTAi interest group = 5	HTAi interest group = 3
	PATH interest group = 4	PATH interest group = 4	PATH interest group = 4
	Stafinski *et al.* review[8] = 68	Stafinski *et al.* review = 4	Stafinski *et al.* review[8] = 2

In total, 64 potentially relevant references were identified from the grey literature searches. However, after full screening, none of these references was included in the final review. The absence of grey literature discussing general principles and themes contrasts with the findings from the Stafinski review[8] in which the majority of case studies were located from grey literature sources. Although this difference could be due to the more limited range of grey literature databases searched here, it seems more likely that the reporting format of case study applications (typically in the form of presentation slides) means that these would be identified only through grey literature searches. In addition, it should be noted that several references from the grey literature were already identified from searches of other sources through the HTAi interest group, PATH website and additional Advisory Group suggestions.

The citation searches identified 122 potentially relevant references. After final screening, five additional references were deemed relevant for inclusion in the review. From the remainder of the searches undertaken, a further two references were identified from the policy websites[51,56] (*Table 28*), one from the Advisory Group suggestions, four from the PATH interest group and three from the HTAi interest group. In addition, by cross-referencing with the results from the Stafinski *et al.* paper,[8] four additional references were identified, two of which were journal articles not previously identified in the traditional systematic search. *Table 29* gives the full list of included references and areas covered.

TABLE 28 Results from policy website searches

Organisation	Potentially relevant references	Included references
UK based		
NICE	'Coverage with evidence' = 196 under NICE and economics subsections 'Only in research' = 45 under NICE subsection	1
NICE Citizens Council	1	1 (already included in list of key papers)
Department of Health	0	0
HM Treasury	0	0
House of Commons Health Committee	3	1 (already included in list of key papers)
Health Industry Forum	5	2 (1 of which already included in list of key papers)
Office of Fair Trading	0	0
Office for Life Sciences	1	1 (already included in list of key papers)
NHS Scotland and NHS Quality Improvement Scotland	0	0
Department for Business, Innovation & Skills	Search terms: 'only in research' and 'evidence' = 19 'Coverage evidence' = 0 Managed entry = 0 Free scan = 1	0
Non-UK based		
Centers for Medicare & Medicaid	1	1 (already included in list of key papers)
Center for Medical Technology Policy	7	7 (all of which already identified from key papers and traditional searches)
Organisation for Economic Co-operation and Development (OECD)	1	0
European network for Health Technology Assessment (EUnetHTA)	0	0

TABLE 29 Full list of included references and overview of issues covered in each reference

| Reference | Terminology and taxonomy | General themes | Specific issues | | | | Examples(s) of application |
			Evidence collection	Investment and reversal costs	Changing price of technology	Ethical issues/ social value judgements	
Journal articles							
1. Anon 2007[31]	✓	✓	✓	✓			
2. Bowen 2009[17]	✓	✓	✓				✓
3. Breckenridge 2008[36]		✓					✓
4. Brezis 2006[13]			✓			✓	
5. Brice 2008[35]		✓					✓
6. Briggs 2010[18]		✓	✓	✓	✓		✓
7. Carbonneil 2009[19]		✓					✓
8. Carino 2004[37]						✓	✓
9. Carino 2006[34]		✓	✓				✓
10. Carlson 2010[25]	✓	✓					✓
11. Chalkidou 2007[3]	✓	✓	✓	✓		✓	✓
12. Chalkidou 2008[29]		✓	✓	✓		✓	✓
13. Chalmers 2007[26]		✓				✓	✓
14. Dhalla 2009[38]		✓	✓	✓			✓
15. Gafni 2007[14]		✓		✓			✓
16. Garber 2008[39]		✓	✓				
17. Goeree 2006[20]		✓					
18. Goeree 2007[15]		✓					✓
19. Groeneveld 2006[16]						✓	
20. Hutton 2007[4]		✓	✓	✓	✓		✓
21. Kamerow 2007[32]						✓	✓
22. Krich 2007[40]		✓	✓			✓	
23. Lindsay 2007[41]		✓	✓			✓	✓
24. Macdonald 2008[42]		✓	✓	✓		✓	
25. McCabe 2010[21]	✓	✓	✓		✓		✓
26. Menon 2010[43]		✓	✓				
27. Miller 2006[22]		✓	✓			✓	✓
28. Mohr 2010[33]		✓	✓	✓		✓	✓
29. Neumann 2008[44]		✓	✓			✓	
30. Niezen 2007[45]	✓						
31. Pearson 2006[46]		✓	✓			✓	
32. Ramsey 2005[27]			✓				✓
33. Reed 2008[47]		✓		✓	✓		✓
34. Stafinski 2010[8]	✓	✓	✓	✓	✓		✓
35. Straube 2005[48]		✓	✓				✓
36. Thornton 2007[49]		✓	✓	✓	✓		
37. Towse 2010[23]	✓	✓	✓	✓	✓		✓
38. Trueman 2010[24]	✓	✓	✓	✓		✓	✓
39. Tunis 2006[2]	✓	✓	✓	✓		✓	✓
40. Tunis 2007[5]		✓	✓				✓
41. Tunis 2009[30]		✓	✓				
42. Wallner 2008[28]	✓	✓	✓			✓	✓
43. Whalley 2004[50]							✓

TABLE 29 Full list of included references and overview of issues covered in each reference (*continued*)

Reference	Terminology and taxonomy	General themes	Specific issues				
			Evidence collection	Investment and reversal costs	Changing price of technology	Ethical issues/ social value judgements	Examples(s) of application
Policy documents							
44. Centers for Medicare & Medicaid Services 2006[55]	✓	✓	✓				
45. Cooksey 2006[9]		✓	✓		✓		
46. Department of Health 1999[52]		✓	✓				
47. House of Commons Health Committee 2005[10]		✓		✓	✓		✓
48. Mechanic 2007[56]		✓	✓				✓
49. Miller 2006[22]	✓	✓					✓
50. NHS Quality Improvement Scotland 2008[53]		✓	✓	✓		✓	
51. NICE and AHRQ 2008[51]	✓	✓	✓	✓		✓	✓
52. NICE Citizens Council 2007[11]		✓	✓	✓		✓	
53. Office for Life Sciences 2009[54]		✓			✓		
54. Office of Fair Trading 2007[12]		✓		✓	✓		
Other sources							
55. Chalkidou 2006[59]			✓			✓	✓
56. CMTP 2009[133]		✓	✓				✓
57. Reeves 2008[60]		✓	✓				✓
58. Sansom 2007[58]		✓	✓	✓		✓	✓
59. Slutsky 2008[57]		✓	✓			✓	✓

Appendix 2

An algorithm for 'only in research' and 'approval with research' decisions

FIGURE 26 Part I: Technologies without significant irrecoverable costs.
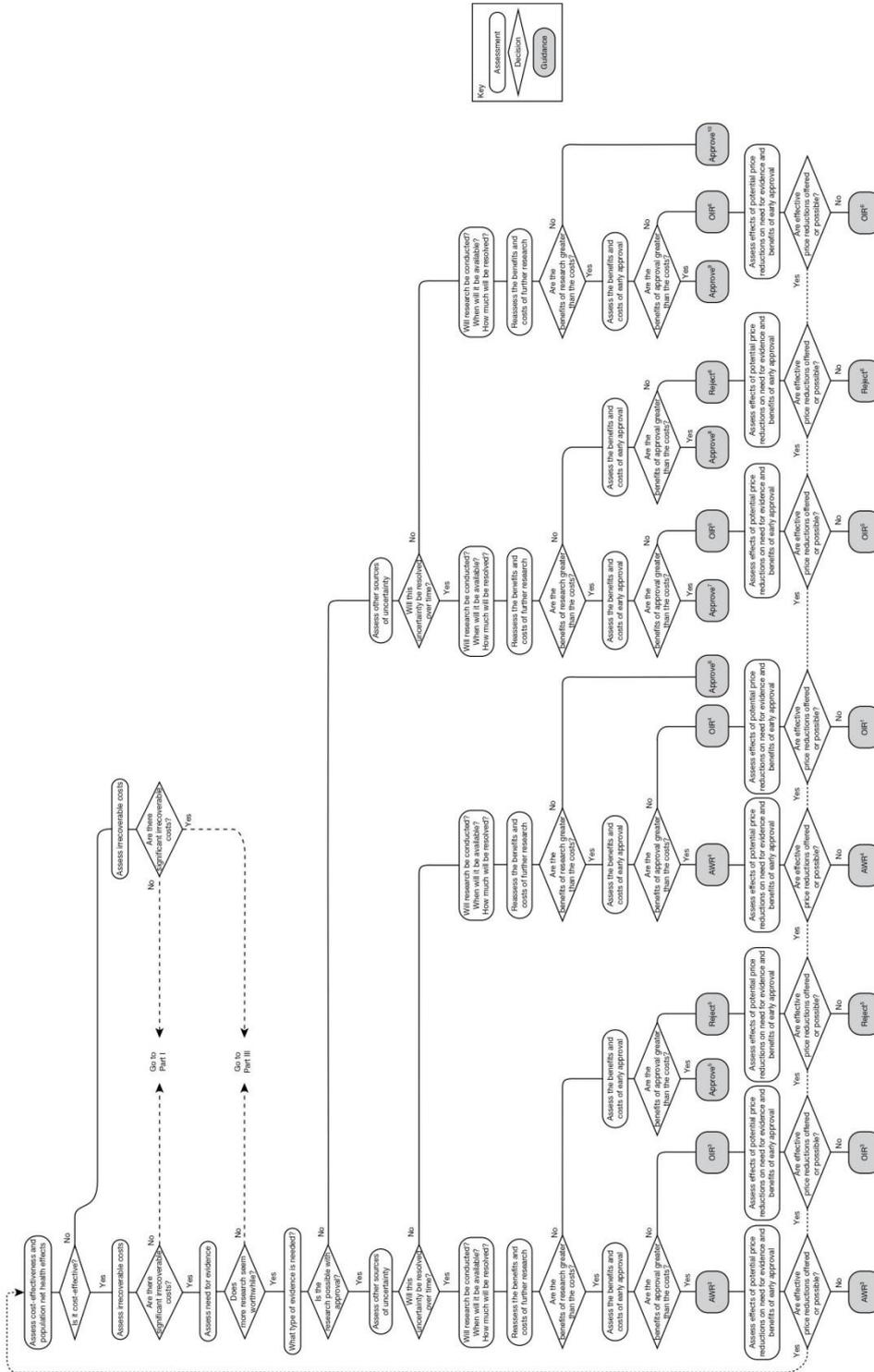
FIGURE 27 Part II: Technologies with significant irrecoverable costs, expected to be cost-effective and research needed.
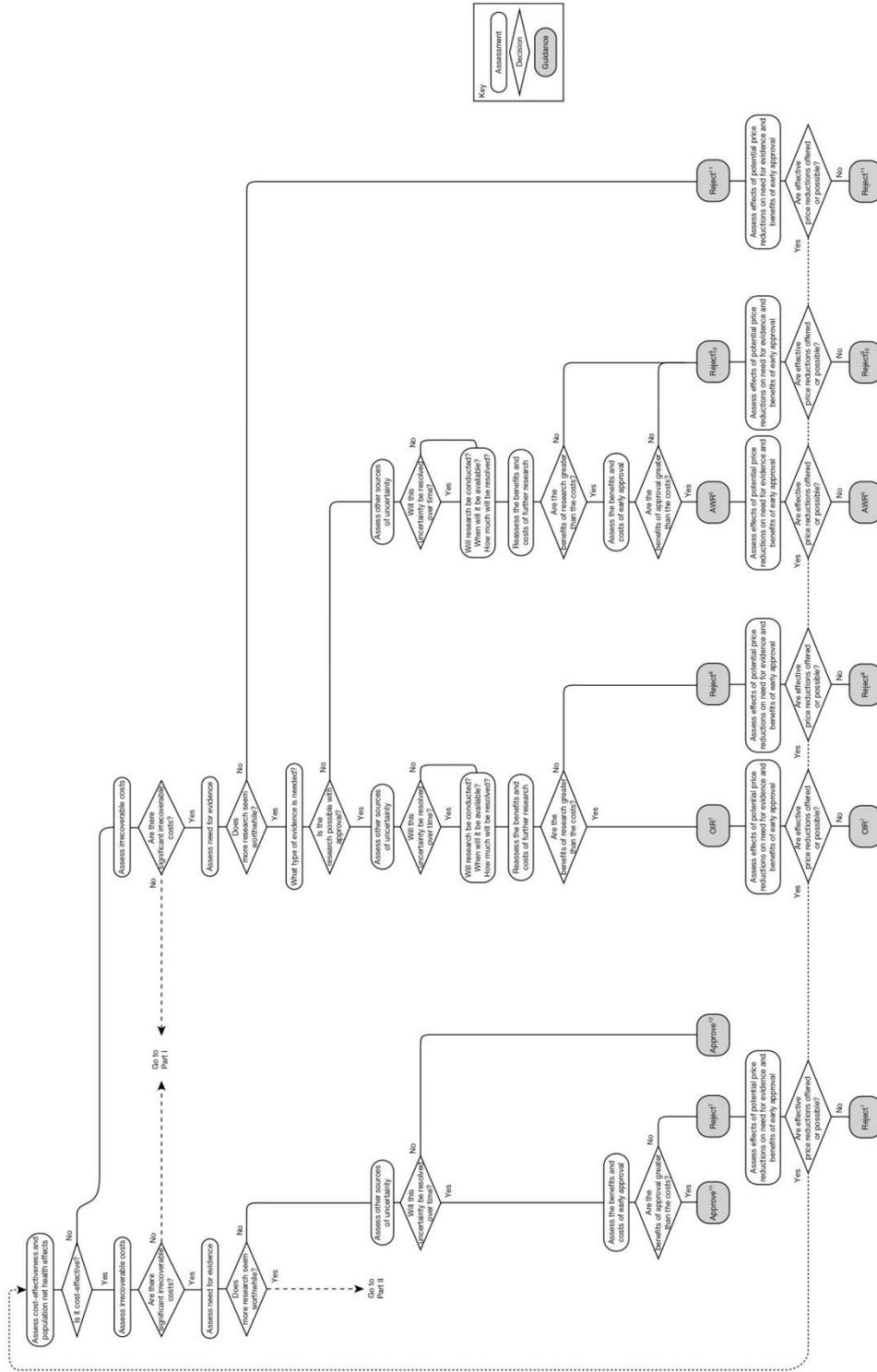
FIGURE 28 Part III: Technologies with significant irrecoverable costs, not expected to be cost-effective or research not needed.

Appendix 3

Data extraction

TABLE 30 Data extraction template

---

1. Appraisal information

(a) Appraisal number

(b) Appraisal title

(c) Publication date

(d) Is the appraisal a review of previous guidance?

(e) Has the guidance been subsequently reviewed?

(f) Appraisal process (STA or MTA)

(g) Disease area

(h) Monotherapy or combination therapy

2. Recommendations

(i) Type of recommendation (OIR or AWR)

(j) OIR recommendation

(k) Type of research recommended (experimental or observational)

(l) Applied to a subgroup?

(m) Patient access scheme offered?

(n) Patient access scheme accepted?

(o) Years until proposed review date

3. Background information on the condition

(p) Incidence and prevalence of the condition

(q) Total population affected

(r) Type of technology (drug, device, diagnostic, procedure, etc.)

(s) Alternative active treatment available?

4. Cost-effectiveness data

(t) Manufacturer/sponsor estimates of the:

　ICER for the whole population

　ICER for the OIR/AWR subgroup (where applicable)

　Probability cost-effective at £20,000 and £30,000

(u) AG/ERG estimates of the:

　ICER for the whole population

　ICER for the OIR/AWR subgroup (where applicable)

　Probability cost-effective at £20,000 and £30,000

(v) Committee's best estimate of the cost-effectiveness (mean ICER and/or range)

TABLE 30 Data extraction template

---

5. Committee rational for OIR/AWR recommendation

(w) Evidence base

    No evidence

    Poor or insufficient evidence

    Evidence suggests less effective than alternative

    No or limited data on target population

    No change in evidence base since previous appraisal

    No or limited evidence on safety or adverse effects

    No or limited evidence to support mechanism of action

(x) Cost-effectiveness

    Cost-effectiveness not assessed against appropriate comparator

    Considered unlikely to be cost-effective

    Uncertainty around cost-effectiveness too high

(y) Investment/irreversibility

    Overall costs too high

    Irreversible/sunk costs too high

    Concern about potential impact on ongoing or future research

---

TABLE 31 Technology appraisals including OIR or AWR recommendations in draft or final guidance

No.	Technology appraisal subject	Date	ACD/FAD
TA2[134]	Prostheses for primary total hip replacement	April 2000	FAD only[a]
TA5[88]	Cervical cancer – liquid based cytology	May 2000	FAD only[a]
TA6[105]	Taxanes for breast cancer	June 2000	FAD only[a]
TA8[90]	Hearing aid technology	July 2000	FAD only[a]
TA16[109]	Knee joints (defective) – autologous cartilage transplantation	December 2000	FAD only[a]
TA17[110]	Colorectal cancer – laparoscopic surgery	December 2000	FAD only[a]
TA23[135]	Recurrent malignant glioma (brain cancer) – temozolomide	April 2001	FAD only[a]
TA30[108]	Taxanes for breast cancer	September 2001	FAD only[a]
TA33[106]	Colorectal cancer (advanced) – irinotecan, oxaliplatin and raltitrexed (Tomudex®, Hospira)	March 2002	FAD only[a]
TA35[136]	Arthritis (juvenile idiopathic) – etanercept	March 2002	FAD only[a]
TA36[92]	Rheumatoid arthritis – etanercept and infliximab	March 2002	FAD only[a]
TA37[100]	Lymphoma (follicular non-Hodgkin's) – rituximab (MabThera®, Roche)	March 2002	FAD only[a]
TA44[98]	Hip disease – metal on metal hip resurfacing	June 2002	ACD and FAD
TA50[137]	Leukaemia (chronic myeloid) – imatinib (Gilvec®, Novartis)	May 2002	ACD only
TA51[96]	Depression and anxiety – computerised cognitive behaviour therapy	October 2002	ACD and FAD
TA60[93]	Diabetes – patient education models	November 2002	ACD only
TA65[97]	Aggressive non-Hodgkin's lymphoma – rituximab	September 2003	ACD and FAD
TA68[94]	Macular degeneration (age-related) – photodynamic therapy	September 2003	ACD and FAD
TA70[138]	Leukaemia (chronic myeloid) – imatinib	October 2003	FAD only
TA72[116]	Rheumatoid arthritis – anakinra (Kieret®, Swedish Orphan)	November 2003	ACD and FAD
TA75[99]	Hepatitis C – pegylated interferons, ribavirin and alfa interferon	January 2004	ACD and FAD
TA86[139]	Gastro-intestinal stromal tumours (GIST) – imatinib	May 2004	ACD only
TA89[102]	Cartilage injury – autologous chondrocyte implantation (ACI)	May 2005	ACD and FAD
TA92[140]	Tooth decay – HealOzone	July 2005	FAD only
TA93[113]	Colorectal cancer (advanced) – irinotecan, oxaliplatin and raltitrexed	August 2005	ACD and FAD
TA97[95]	Depression and anxiety – computerised cognitive behavioural therapy	February 2006	ACD and FAD
TA99[141]	Immunosuppressive therapy for renal transplantation in children	April 2006	FAD only
TA104[126]	Psoriatic arthritis – etanercept and infliximab	June 2005	ACD only
TA111[142]	Alzheimer's disease – donepezil (Aricept®, Eisai), galantamine (Reminyl® XL, Shire), rivastigmine (Exelon®, Novartis) and memantine (Ebixa®, Lundbeck)	November 2006 (update 2009)	ACD and FAD
TA113[91]	Diabetes (type 1 and 2) – inhaled insulin	December 2006	ACD and FAD
TA121[143]	Glioma – carmustine implants and temozolomide	December 2005	ACD only
TA129[103]	Multiple myeloma – bortezomib (Velcade®, Janssen)	July 2006	ACD only
TA130[114]	Rheumatoid arthritis – adalimumab, etanercept and infliximab	February 2006	ACD only
TA135[144]	Mesothelioma – pemetrexed disodium	March 2006	ACD only
TA142[145]	Anaemia – erythropoietin (alpha and beta) and darbepoetin	July 2005	ACD only
TA143[146]	Ankylosing spondylitis – adalimumab, etanercept and infliximab	July 2007	ACD only
TA159[147]	Pain (chronic neuropathic or ischaemic) – spinal cord stimulation	October 2008	ACD and FAD
TA163[89]	Ulcerative colitis (acute exacerbations) – infliximab	December 2008	ACD and FAD
TA166[104]	Hearing impairment – cochlear implants	December 2007	ACD only
TA167[101]	Abdominal aortic aneurysm – endovascular stent-grafts	February 2009	ACD and FAD

a ACDs were not publicly available for these appraisals.

Appendix 4

Types and categories of guidance for case studies

TABLE 32 Types and categories of guidance from using the checklist

| Assessment	1	2	3	4	5	6	7	Guidance	
1	Yes	No	Yes	Yes	Yes/no	Yes	–	AWR[1]	
2	Yes	No	Yes	Yes	Yes/no	No	–	Approve[1]	
3	Yes	No	Yes	No	Yes/no	Yes	Yes	Approve[2]	
4	Yes	No	Yes	No	Yes/no	Yes	No	OIR[1]	
5	Yes	No	Yes	No	Yes/no	No	–	Approve[3]	
6	Yes	No	No	–	–	–	–	Approve[4]	Part I of the algorithm
7	No	No	Yes	Yes	Yes/no	Yes	–	OIR[2]	
8	No	No	Yes	Yes	Yes/no	No	–	Reject[1]	
9	No	No	Yes	No	Yes/no	Yes	Yes	AWR[2]	
10	No	No	Yes	No	Yes/no	Yes	No	Reject[2]	
11	No	No	Yes	No	Yes/no	No	–	Reject[3]	
12	No	No	No	–	–	–	–	Reject[4]	
13	Yes	Yes	Yes	Yes	Yes	Yes	Yes	AWR[3]	
14	Yes	Yes	Yes	Yes	Yes	Yes	No	OIR[3]	
15	Yes	Yes	Yes	Yes	Yes	No	Yes	Approve[5]	
16	Yes	Yes	Yes	Yes	Yes	No	No	Reject[5]	
17	Yes	Yes	Yes	Yes	No	Yes	Yes	AWR[4]	
18	Yes	Yes	Yes	Yes	No	Yes	No	OIR[4]	
19	Yes	Yes	Yes	Yes	No	No	–	Approve[6]	
20	Yes	Yes	Yes	No	Yes	Yes	Yes	Approve[7]	Part II of the algorithm
21	Yes	Yes	Yes	No	Yes	Yes	No	OIR[5]	
22	Yes	Yes	Yes	No	Yes	No	Yes	Approve[8]	
23	Yes	Yes	Yes	No	Yes	No	No	Reject[6]	
24	Yes	Yes	Yes	No	No	Yes	Yes	Approve[9]	
25	Yes	Yes	Yes	No	No	Yes	No	OIR[6]	
26	Yes	Yes	Yes	No	No	No	–	Approve[10]	
27	Yes	Yes	No	–	Yes	–	Yes	Approve[11]	
28	Yes	Yes	No	–	Yes	–	No	Reject[7]	
29	Yes	Yes	No	–	No	–	–	Approve[12]	
30	No	Yes	Yes	Yes	Yes/no	Yes	–	OIR[7]	
31	No	Yes	Yes	Yes	Yes/no	No	–	Reject[8]	Part III of the algorithm
32	No	Yes	Yes	No	Yes/no	Yes	Yes	AWR[5]	
33	No	Yes	Yes	No	Yes/no	Yes	No	Reject[9]	
34	No	Yes	Yes	No	Yes/no	No	–	Reject[10]	
35	No	Yes	No	–	–	–	–	Reject[11]	
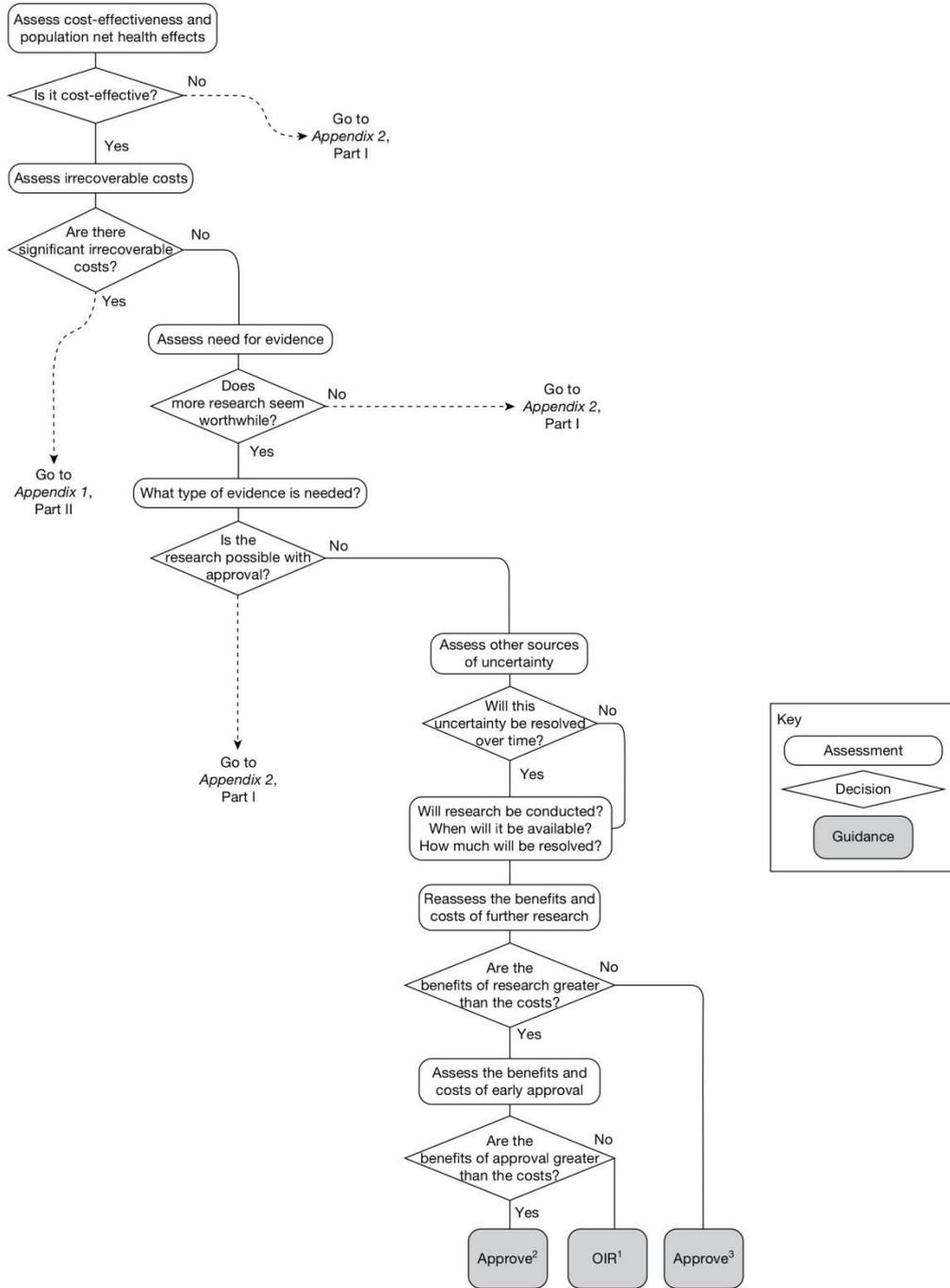
FIGURE 29 Possible pathways for CLOP.
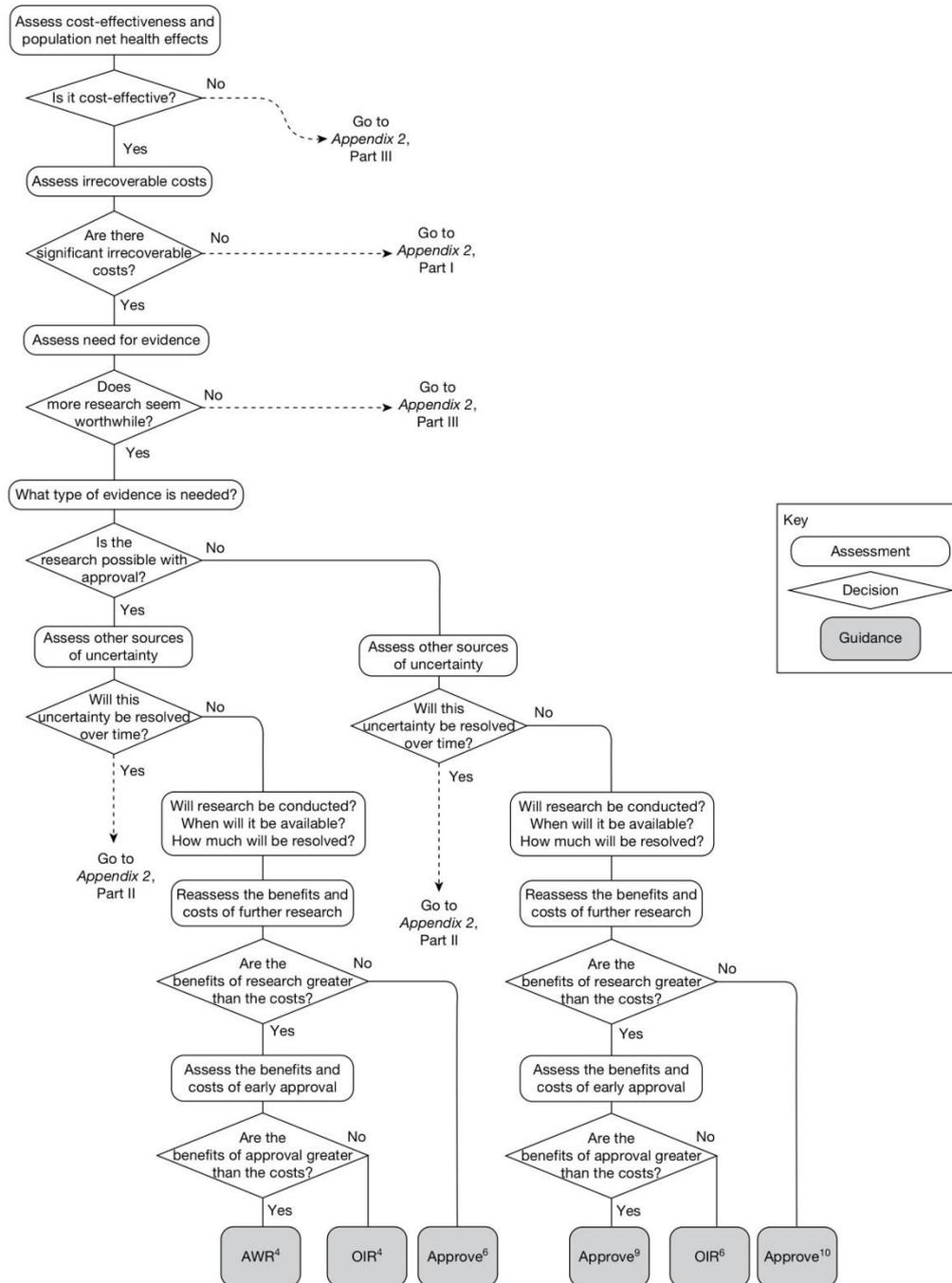
FIGURE 30 Possible pathways for EECP.

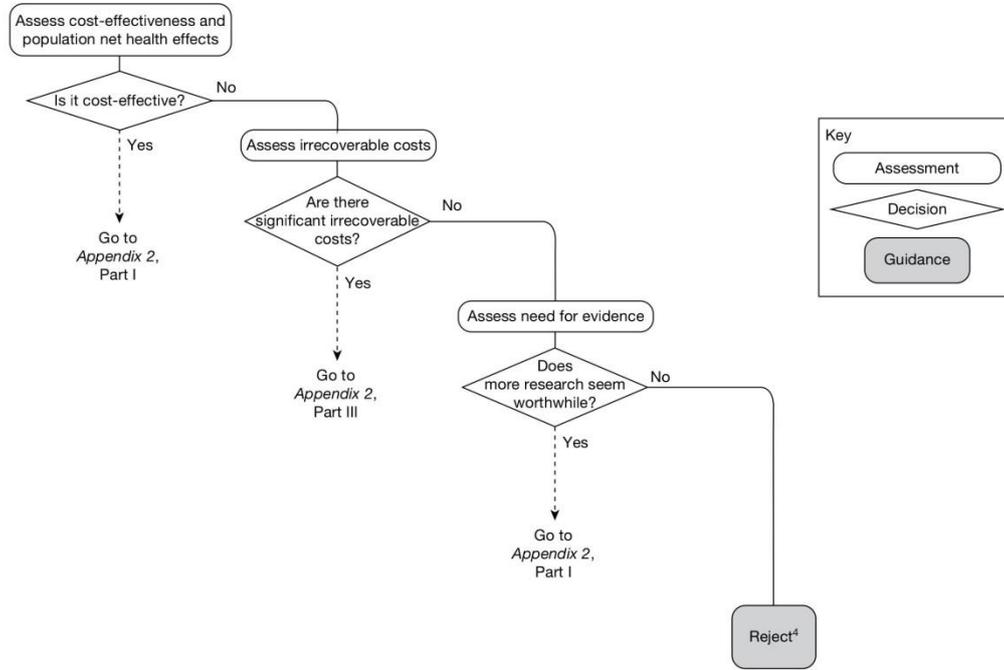**FIGURE 31** Possible pathways for OMAL.

FIGURE 32 Possible pathways for PsA.

Other sources of uncertainty: searches for sources of additional information

The searches for sources of information to complete the assessments in *Chapter 5* are detailed in the following sections. Specifically they describe sources for each type of information required, including access details.

Changes in the prices of the technology and its comparators

The anticipated time for drugs coming off patent is available from a number of web-based sources. These include:

1. MPA Business Services Limited (www.mpasearch.co.uk/mpadatasearch). This requires registration and a fee for generating information, which is approximately £25–500 for a single molecule.
2. Horizon scanning (www.ukmicentral.nhs.uk/pressupp/pe.htm) includes a patent database that details the dates of drug patent issue. This service is available only to NHS staff or through specific project permission.
3. The national patent database (www.ipo.gov.uk/patent/p-find/p-find-number.htm) provides the date of patent registration. The appropriate patent can be found using the INID code (on the patent page) or using http://gb.espacenet.com/search97cgi/s97_cgi. exe?Action=FormGen&Template=gb/en/quick.hts to search by drug name or summary of product characteristics. In the UK a patent typically lasts for 20 years from the application date shown on the front page of the patent; thus, 20 years from this date will give the full-term expiry date. One potential complication is that for pharmaceuticals this can be extended up to a further 5 years by gaining a Supplementary Protection Certificate (SPC).
4. In the USA Google Patent finder is available (www.google.com/googlepatents/about.html).

For non-drugs (medical interventions, etc.) there is less information. Many medical devices will have a trademark/CE mark (see www.ipo.gov.uk/tm/t-find/t-find-text/ to search by trademark number or free text). Trademarks can last indefinitely as long as they are renewed every 10 years (www.ipo.gov.uk/tm.htm). Any patent is likely to relate to some aspect of the device rather than the device itself. Although prices may change over time they can also be relatively stable but with incremental innovation of the original device. Again, this is likely to differ by health-care system, technology and indication.

Regarding the extent of change there is again very little published information on the likely impact of generics. The OFT estimated that, on average, generic prices tend to be 25% of the original brand price before patent expiry.[12] A recent HTA model[148] also used the assumption that the generic price is equal to 25% of the original product price. A US study[149] previously saw the cost of CLOP at 80% of the branded cost (this was deemed inappropriate for the UK). These general figures are, however, unlikely to be entirely accurate as reductions in prices are likely to be location specific and differ by type of technology. This cannot be quantified but needs to be recognised when interpreting the likely cost of generics.

Entry of new technologies

To identify any relevant new technologies that may be emerging there are a number of sources available:

1. The topic selection by NICE gives some indication as to what drugs will soon be given licences in the UK. The topic selection panel minutes are available at www.nice.org.uk/ getinvolved/topicselection/considerationpanels/minutes/minutes.jsp.

2. Also, proposed technology appraisals are given at www.nice.org.uk/ourguidance/ niceguidancebytype/technologyappraisals/proposedappraisals/nowave.jsp. Technology appraisal topics not being progressed by NICE are available at www.nice.org.uk/getinvolved/ suggestatopic/outcomes/Topics.jsp.

3. New drugs suggested by https://www.ukpharmascan.org.uk/login. These are available only to pharmaceutical companies and certain NHS staff.

4. Medical technologies in the Medical Technology Evaluation Programme (www.nice.org.uk/ aboutnice/whatwedo/aboutmedicaltechnologies/notifyaproducttoep.jsp). Contains a list of technologies reviewed by the Medical Technologies Advisory Committee (MTAC) up to a certain date.

5. Information about licence applications: new applications or change of use in progress are available through NHS horizon scanning (www.ukmi.nhs.uk/applications/NDO/dbSearch. asp). Requires registration to access 'new drugs online'. Can be searched by drug name or by BNF chapter. Search by developmental stage (licensed, Phase 1/2/3, licensed but not launched, application filled, etc.).

6. Non-UK regulators such as the FDA (www.fda.gov/Drugs/informationondrugs/ucm079436. htm) have databases containing upcoming competitors.

Many of the databases contain information on Phase II trials and so it may be necessary to try to predict the number of these that will then proceed to Phase III and approval. The OFT life cycle of drug development shows a 3-year period between initiation of Phase II and Medicines and Healthcare products Regulatory Agency (MHRA)/European Medicines Agency (EMEA) approval.[12] The same document also gives a period of 4.6 years from the initiation of Phase II to the end of phase III (taken from DiMasi *et al.*[150] but it is unclear which figures this is using). In the paper by DiMasi *et al.*[150] a period of 9.9 years from initiation of Phase II to the beginning of approval stage in the US market is quoted. The paper also presents the probability of approval by clinical phase (*Table 33*). Using these figures the OFT states that on average 21.5% of drugs entering clinical trials (Phases I, II and III) will go through to market approval.

Other research reporting the technology and its comparators

There are a number of sources available to search for clinical research ongoing at the time of appraisal. These include:

1. clinical trial registers (www.controlled-trials.com/, www.clinicaltrials.gov/)
2. the National Research Register (NRR) Archive (www.nihr.ac.uk/Pages/NRRArchive.aspx)
3. ClinicalTrials.gov (www.clinicaltrials.gov)
4. the World Health Organization (WHO) ICTRP (www.who.int/ictrp/en/).

These databases report on the sample size, end points, trial type, comparators, start date and anticipated reporting date.

TABLE 33 Probability of approval by clinical phase[150]

Clinical phase	Probability of entering phase
Phase I	100
Phase II	71
Phase III	31.4

In addition, there are a number of non-clinical registries and outcomes databases. These include:

1. The NRR Archive (www.nihr.ac.uk/Pages/NRRArchive.aspx) – this covers all UK NHS-funded research not just trials. The NRR stopped being updated in 2007.
2. The current equivalent is the UK Clinical Research Network (UKCRN) Study Portfolio (http://public.ukcrn.org.uk/search/). The type of research that is eligible to be included in the Portfolio database varies depending whether it is funded in England, Wales, Scotland or Northern Ireland, but details of eligibility criteria are available at www.crncc.nihr.ac.uk/about_us/processes/portfolio/p_eligibility.

Appendix 5

Funding brief

MRC-NIHR METHODOLOGY RESEARCH PROGRAMME: NEEDS-LED RESEARCH

FUNDING BRIEF

Informing a decision framework for when NICE should recommend the use of health technologies[a] only in the context of an appropriately designed programme of evidence development ('only in research')

1. BACKGROUND

Introduction

In its initial instructions to NICE, the Department of Health made provision for the Technology Appraisal Committee to 'recommend that further research is carried out to see whether the potential promise of the intervention can be realised, indicate in broad terms the questions this research should address and advise clinicians that, in the meantime, they should only use the new intervention as part of a well-designed programme of research intended to answer these questions'.[b]

There is controversy about the terminology used to refer to these decisions. To date the term 'only in research' (OIR) has been used. Also parallels have been drawn with the concept of 'Coverage with Evidence Development', which is used in the US (Hutton et al., [3]).

Of the 133 technology appraisal guidance documents issued by NICE between 1999 and the end of 2007, 21 (26%) contained one or more 'OIR' recommendations. In practice, such 'OIR' decisions have been made when there are important uncertainties about the effectiveness and cost-effectiveness of an intervention [1]. It is however acknowledged that 'OIR' decisions have potential costs, benefits and risks; both the individual decisions themselves and the principle of an 'OIR' decision option have been subject to international debate [2].

Importance of topic

Recently there have been a number of calls for NICE to use the 'OIR' decision option more often [2;4;5] [1]. There has also been criticism that 'OIR' is used as a money saving option or to avoid a clear 'no' decision [2].

While NICE's R&D Committee has advised on the issues that need to be addressed in making an OIR recommendation, there are at present no decision criteria as to when 'OIR' might be an appropriate option; each recommendation has been made on a case-by case basis. This research is very timely because of recent moves to assess new technologies closer to the point of licensing. At this point, the evidence base is likely to be immature and as a result greater uncertainty about the comparative effectiveness and overall value of the technology [1;3].

Research needs

The primary research need is to inform the key principles that could be taken into consideration by NICE Appraisal Committees to enable them to determine whether an 'OIR' decision is appropriate. The methodological and process implications of providing the necessary technology-specific data on each of these principles should also be considered.

The main research output that is required is a list of key principles and recommendations about how these principles could be operationalised within a structured decision framework that NICE could employ when considering an 'OIR' decision. The research should also explore the possibility of using these principles as the basis of a 'checklist' to provide transparency in 'OIR' decision-making. This could include one or more pilot studies based on appraisals in progress (or proposed) to test possible decision-making approaches for 'OIR'. It could also involve an evaluation of how Value of Information methods can be integrated into the decision-making framework.

A second research objective is to investigate how 'OIR' recommendations are currently implemented in practice and examine barriers and potential solutions including any changes to the research infrastructure. This will necessitate examination of previous 'OIR' recommendations.

An optional objective would be to explore the terminology used to describe 'OIR' recommendations. If opted for, then the acceptability of different options should be tested through consultation.

Other relevant initiatives and/or activities

Some preliminary work into the 'OIR' option has already been undertaken. NICE consulted its Citizen's Council; a group of 30 people who bring the views of the public to NICE decision-making. The council made a series of recommendations about the circumstances in which it felt it is justified for NICE to recommend that an intervention is used only in the context of research. NICE has also obtained the views of other stakeholders through a workshop, held in collaboration with the American Agency for Healthcare Research and Quality (AHRQ)[c] and a series of interviews with industry, academia, government, the NHS, independent research funding agencies, and consultants (publication in progress). Internationally the concepts have been discussed by participants at the Health Technology Assessment International (HTAI)_ Policy Forum in February 2007 [3] and the US Centres for Medicare and Medicaid Services [8]). A HTAI special interest policy group has been established.

There have been a number of academic publications exploring the 'OIR' option including that by Chalkidou *et al.* [1] and Hutton *et al.* [3]. There has also been research into potential methodologies including 'Value of Information' methods [10–13] and Options Theory [15]. Whilst not mandatory, the option to use VoI methods has been included in the NICE guide to methods of technology appraisal [14]. It is recognised that a quantitative value of information approaches cannot be used in isolation to determine whether an 'OIR' decision would be appropriate.

2. RESEARCH REQUIRED

Important issues to consider

Research is required to fulfil the needs identified in section 1.3 above.

Previous work has identified a number of potential themes that could be explored further. Note this is not an exhaustive list and further exploration of more recent research undertaken by NICE and others is required.

1. The potential research:
   - Will the research reduce uncertainties about the effects of a technology (including side effects) to a level at which a clear-cut decision could be made?
   - Will the research be timely, feasible and relevant (to the required patient groups)?
   - Would the research represent value for money?
   - What is the role of formal VoI methodology?
2. The consequences of making the wrong decision:
   - Are the sunk costs associated with a 'yes' decision particularly high such that it would be uneconomical to reverse the decision?
   - In terms of current research: would a 'yes' decision result in existing research being terminated early, possibly due to recruitment difficulties?
   - In terms of future research: what is the value of the research foregone as a result of a 'yes' or 'no' decision?
3. Equity and ethical concerns:
   - Would an 'OIR' decision effectively limit the technology to those who can access a research centre?
   - Should 'the benefit of the doubt' be given to technologies for life threatening conditions where no other treatment exists (the rule of rescue)?
   - Would a decision violate the social value judgements guidance applied by NICE [9]?

In addition, it will be necessary to consider the opportunity cost of delaying a coverage decision for further research in terms of the potential health gains foregone by delaying access to a cost-effective intervention.

Methodological approaches to be used

The MRP panel recognises the need to allow applicants to identify the most appropriate methodological approach for this study and the approach to be taken should be clearly justified. It is anticipated that a mixture of quantitative and qualitative primary and secondary research will be required. The research will need to address both the theoretical and practical considerations of 'OIR' decisions.

Applicants must ensure that the proposed research is congruent with the 2008 Methods for Technology Appraisal [14] and Social Value Judgements [9]. It is possible that more than one proposal will be funded.

Applicants may wish to consider undertaking a review of principles and methods currently used in, or advocated for, 'OIR' decisions, both in the UK and internationally.

Applicants should also include a retrospective 'case-study' analysis of 'OIR' decisions made by NICE to identify what principles have previously been taken into account when an 'OIR' decision has been made. It may also be necessary to explore instances when an 'OIR' decision was not determined to be appropriate.

Establishing and maintaining a dialogue with NICE and its Appraisal committees and other stakeholders, including patient groups and industry, will be essential. Where necessary, NICE will assist the research team to gain access to the necessary information and expertise.

This analysis may include consideration of how the 'OIR' decision-making process can help to define the nature of the research required (e.g. study design; key variables to be assessed). The pros and cons of possible approaches should be identified, in order to provide recommendations for the principles to be included in a structured decision-making process and how such a process can be operationalised by NICE Appraisal Committees. There is scope for this decision-making process to be supported by a 'checklist'.

3. MANAGEMENT OF STUDY

Applicants should specify and justify: (1) the resources requested for the research and (2) an appropriate timescale within which the research can be completed.

4. REFERENCES

1. Chalkidou K, Hoy A, Littlejohns P. Making a decision to wait for more evidence: When the National Institute for Health and Clinical Excellence recommends a technology only in the context of research. *J R Soc Med* 2007;**100**:453–60.

2. NICE Citizen's Council. *Only in research*. London: NICE; 2007.

3. Hutton J, Trueman P, Henshall C. Coverage with evidence development: An examination of conceptual and policy issues. *Int J Technol Assess Health Care* 2007;**23**:425–35.

4. Cooksey D. *A review of UK health research funding*. London: HM Treasury; 2006.

5. House of Commons Health Committee. National Institute for health and Clinical Excellence. London: The Stationery Office; 2008. Report No.: HC 27–1.

6. Tunis SR, Chalkidou K. Coverage with evidence development: A very good beginning, but much to be done. Commentary to Hutton *et al. Int J Technol Assess Health Care* 2007;**23**:432–5.

7. CMO. *On the state of the public health: Annual Report of the Chief Medical Officer 2005*. London: Department of Health; 2006.

8. Centers for Medicare and Medicaid Services. *National coverage determinations with data collection as a condition of coverage: Coverage with Evidence Development*. 2006 Dec 7.

9. NICE. Social Value Judgements Guideline. *Principles for the development of NICE guidance*. London: NICE; 2005.

10. Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. *Health Technol Assess* 2004;**8**(31).

11. Eckermann S, Willan AR. Expected value of information and decision making in HTA. *Health Econ* 2007;**16**:195–209.

12. Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis. *Health Technol Assess* 2004;**8**(27).

13. Girling A, Freeman G, Gordon JP, Poole-Wilson P, Scott DA, Lilford RJ. Modeling payback from research into the efficacy of left-ventricular assist devices as destination therapy. *Int J Technol Assess Health Care* 2007;**32**:269–77.

14. NICE. *Guide to the methods of technology appraisal*. London: NICE; 2004.

15. Palmer S, Smith PC. Incorporating option values into the economic evaluation of health care technologies. *J Health Econ* 2000;**19**:755–66.

Appendix 6

Technical appendix

Overview

The purpose of this appendix is to describe more formally the core principles of the framework developed to include OIR or AWR recommendations alongside approve and reject policies in deciding on the use of health technologies. This appendix will describe how to carry out the calculations that can inform each of the assessments described in *Chapter 5* in order to quantify the net health impact of alternative policy decisions. Established methods for cost-effectiveness analysis and value of information analysis are drawn together using a consistent set of notation throughout to describe a general algebraic framework. A simple numerical example is used to demonstrate the application of this general algebraic framework, illustrating how the pay-offs from the alternative policy options of approve, AWR, OIR and reject can be estimated. Special consideration is given to the impact of price at submission on the estimated pay-off and how this can alter the rank order of the alternative policy options.

Notation

This appendix makes use of formal notation throughout to illustrate how the pay-off from the alternative policy decisions can be calculated. *Table 34* shows the terms used, their definitions and the sections in this appendix where they are described in detail.

Is the intervention cost-effective and what are the risks?

Is the intervention cost-effective?

Within the health-care sector the outcome of interest for economic evaluation has generally been regarded as some measure of health, h. To facilitate comparisons between alternative health interventions, $j = 1,2,\ldots,J$, it is necessary to have a single index measure of health that captures both length of life and quality of life. The opportunity cost of investing resources in providing a particular health intervention can be characterised in terms of the health that could have been generated with the next best alternative use of those same resources. By utilising the additional cost that would displace one unit of health gain elsewhere in the health-care system, k, it is possible to describe costs that fall on the health-care sector budget, c, in terms of health benefits,

$$\frac{c}{k} = h,$$

forgone. The cost-effectiveness of a particular health-care intervention can be assessed by estimating its associated costs and health outcomes and determining whether or not, given the cost-effectiveness threshold k, reimbursing the intervention would lead to an expected increase in population health benefits. The assessment is the same regardless of whether the intervention is new or is already reimbursed as part of current activities, that is, the decision to discontinue funding an existing activity can be made on the same basis as the decision to reimburse a new intervention.

TABLE 34 Table of notation

Symbol	Definition	Section reference
$h(\ )$	Health benefits	*Is the intervention cost-effective?*
$j = 1,2,\ldots,J$	Each j represents an alternative out of a set of J mutually exclusive comparator interventions	*Is the intervention cost-effective?*
k	Cost-effectiveness threshold	*Is the intervention cost-effective?*
$c(\ )$	Costs of interventions that fall on the budget for health care	*Is the intervention cost-effective?*
θ	Set of uncertain parameters used to estimate costs and health effects and that are amenable to further research	*Is the intervention cost-effective?, Cost-effectiveness at the patient level*
$NHE(\ )$	Net health effects	*Is the intervention cost-effective?, Cost-effectiveness at the patient level*
Pop	The total effective number of patients who would receive an intervention	*Is the intervention cost-effective?, Cost-effectiveness at the population level*
P	The prevalent population eligible for treatment at the time of the reimbursement decision	*Is the intervention cost-effective?, Cost-effectiveness at the population level*
I_t	The incident population per unit time	*Is the intervention cost-effective?, Cost-effectiveness at the population level*
d	Discount rate	*Is the intervention cost-effective?, Cost-effectiveness at the population level*
T	Time horizon	*Is the intervention cost-effective?, Cost-effectiveness at the population level*
$C(\ )$	Investment cost	
$EVPI_\theta$	The maximum improvement in NHEs that could be achieved for an individual patient by further research on θ	*Is further research required?, Consequences of uncertainty at the patient level*
$\Pi(\ )$	Policy pay-off in terms of NHEs	*Is further research required?, Consequences of uncertainty at the population level*
s	Set of uncertain scenario assumptions used to estimate NHEs	*Is further research required?, Alternative scenarios*
R	The amount of health displaced as a result of the cost of further research	*Does more research seem worthwhile?*
τ	Time at which the results of research will be available	*How long until the research reports?*
α	Probability of further research	*Will the research be conducted?*
n	Number of patient recruited to a research study	*How much of the uncertainty will the research resolve?*
θ_i	Parameter i out of the set of uncertain parameters θ	*What type of evidence is needed?*
τ_i	Time at which the results of research on parameter i will be available	*What type of evidence is needed?*
R_i	The amount of health displaced as a result of the cost of further research to inform parameter i	*What type of research is possible if the technology is approved?*
γ	Uncertain parameter not able to be informed by further research but that will resolve over time	*Impact of other sources of uncertainty*

Figure 33 shows an intervention that generates two additional units of health compared with the next best alternative. For every extra £20,000 that is spent on this intervention there will be a loss of one unit of health from the displaced activities.

If the price of the intervention is £40,000, then approving it for use within the health-care system would generate two additional units of health but displace two units elsewhere and so the net effect would be no change in population health. If the price of the intervention is £60,000, then the decision to approve it for reimbursement would displace more health than would be gained, resulting in a net loss in population health. However, if the price is £20,000 then approving the intervention would displace less health than it generated, resulting in a net gain in health, and in these circumstances the intervention could be regarded as providing value for money.[a]
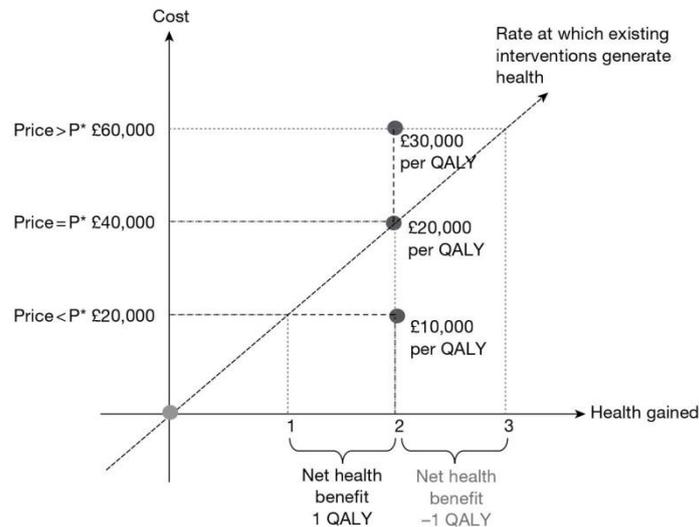
FIGURE 33 Value, cost-effectiveness and net benefit. P*, price that results in no change in population health.

Cost-effectiveness at the patient level

The estimation of expected costs and health outcomes will be informed by the available evidence, θ, and for each intervention j can be expressed in terms of NHEs:

$$NHE(j,\theta) = h(j,\theta) \hspace{4cm} \text{[Equation 1]}$$

Health and cost outcomes that occur in the future should be adjusted to net present values using an appropriate discount rate. The relevant time horizon over which to assess the health and cost impacts must be long enough to capture any differences between the alternative interventions.[150] The decision-maker in charge of the health-care budget can maximise population health by choosing the intervention that maximises expected NHEs:

$$\max_{j} E_{\theta} NHE(j,\theta) \hspace{4cm} \text{[Equation 2]}$$

Thus, an intervention, j, can be determined to be cost-effective if it would maximise expected NHEs[b] out of the range of mutually exclusive alternatives, J, relevant to the decision problem.[67] NHEs make cost-effectiveness comparisons and calculations numerically simple as they embody the objective to maximise health gains from available resources simply by maximising NHEs.[63,64] Of course, the caveat to this is that one must first be satisfied that costs, health outcomes and opportunity costs were calculated appropriately and that they take account of all factors relevant to the viewpoint of the decision-maker, such as equity considerations about the value of health gains to different beneficiaries.[c]

If we take as a starting point NHEs that are deemed to be accurate and reliable we can utilise a simple numerical example to demonstrate the processes and intuition behind conducting formal, evidence-based evaluations to support the range of assessments a decision-maker may wish to make to inform reimbursement decisions and recommendations for further research. The numerical example is introduced in *Table 35*, which presents the health outcomes and costs for two interventions ($j=1,2$). The health outcomes, h, and costs, c, are assessed according to their timing for the individual receiving each intervention by considering two consecutive time

TABLE 35 Net health effects per treated individual over time

Intervention 1							
	Period 1			Period 2			Total
θ	h(1,θ)	c(1,θ) (£)	NHE(1,θ)	h(1,θ)	c(1,θ)	NHE(1,θ)	NHE(1,θ)
1	0.35	2875	0.21	0.35	£2875	0.21	0.41
2	0.20	3250	0.04	0.20	£3250	0.04	0.08
3	0.45	3000	0.30	0.45	£3000	0.30	0.60
4	0.65	2750	0.51	0.65	£2750	0.51	1.03
E_θ	0.41	2969	0.26	0.41	£2969	0.26	0.53

Intervention 2							
	Period 1			Period 2			Total
θ	h(2,θ)	c(2,θ) (£)	NHE(2,θ)	h(2,θ)	c(2,θ)	NHE(2,θ)	NHE(2,θ)
1	0.50	10,250	0.01	0.50	£250	0.49	0.48
2	0.65	11,250	0.09	0.65	£1250	0.59	0.68
3	0.55	12,250	−0.06	0.55	£2250	0.44	0.38
4	0.70	11,500	0.13	0.70	£1500	0.63	0.75
E_θ	0.60	11,313	0.03	0.60	£1313	0.53	0.57

periods. NHEs are calculated given $k = £20,000$. The numbers in the table represent net present values. The evidence on which the estimates of cost and health outcomes are based is represented by an uncertain parameter θ, which can take one of four possible true underlying values, each of which would lead to a different estimation of the NHEs associated with each intervention. The final rows of each part of *Table 35* give the expected values given the uncertainty in how θ will resolve.

In this example there are differences between the interventions in both time periods and so the relevant time horizon for assessing the cost-effectiveness of either intervention comprises both periods one and two. *Figure 34* illustrates how the accrual of expected NHEs changes over time with each intervention from the perspective of an individual patient.

In *Table 35* it can be seen that with intervention 2 the costs are loaded upfront. A comparison of the expected NHEs in period 1 would suggest that intervention 1 is cost-effective. However, the NHEs in period 2 indicate that intervention 2 is cost-effective. The gains offered by intervention 2 in the second period are sufficient to ensure that, overall, intervention 2 is cost-effective. In *Figure 34* the area under the curve for intervention 2 is greater than the area under the curve for intervention 1. If treatment cannot or will not be withdrawn once initiated, then the timing of the costs is irrelevant to the reimbursement decision, and this can be based on the summation of the NHEs across the time periods, as shown in *Table 36*. In the following section (*Are there significant irrecoverable costs?*) we discuss the additional considerations that must be made when the NHEs that accrue vary over time and treatment may be withdrawn.

To apply *Equation 2* to determine which intervention is cost-effective we look at the final row of *Table 36*, which gives the expected NHEs of each intervention given the uncertainty in how θ will resolve. A comparison of these expected NHEs indicates that intervention 2 is cost-effective. Approving intervention 2 rather than intervention 1 would be expected to offer an additional 0.04 units of net health benefit per patient treated.
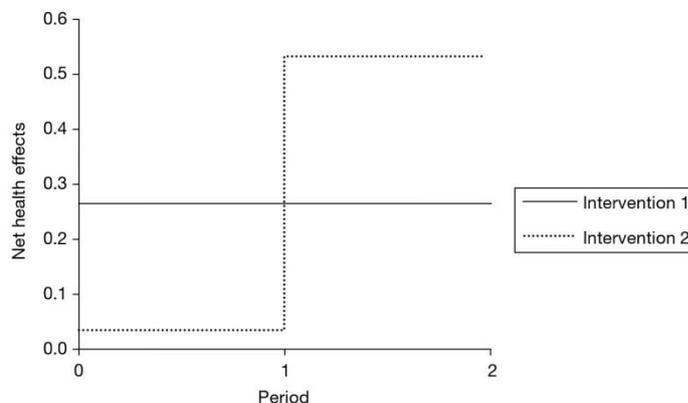
FIGURE 34 Accrual of expected NHEs over time.

TABLE 36 Net health effects per patient

θ	Intervention 1			Intervention 2		
	$h(1,\theta)$	$c(1,\theta)$ (£)	$NHE(1,\theta)$	$h(2,\theta)$	$c(2,\theta)$	$NHE(2,\theta)$
1	0.70	5750	0.41	1.00	£10,500	0.48
2	0.40	6500	0.08	1.30	£12,500	0.68
3	0.90	6000	0.60	1.10	£14,500	0.38
4	1.30	5500	1.03	1.40	£13,000	0.75
E_θ	0.83	5938	0.53	1.20	£12,625	0.57

Cost-effectiveness at the population level

To judge the scale of the improvement in health gains that might result from a decision to provide either intervention it is necessary to consider the number of individuals who would receive treatment. The total population health effects associated with each intervention are found by multiplying the per-patient NHE by the total number of patients who would receive each intervention during the time period for which they would be utilised. This population, *Pop*, is a function of the prevalence and incidence of the disease per unit of time, P and I_t, summed over the relevant time horizon, T_{Pop}, taking account of the decision-maker's rate of time preference for health gains, d:

$$\text{Pop} = P + \sum_{t=1}^{T_{Pop}} \frac{I_t}{(1+d)^t}$$ [Equation 3]

The time horizon, T_{Pop}, over which the interventions would be utilised may be uncertain, requiring that a range of possible values be considered.[53] The time horizon for determining the relevant population size and the scale of the NHEs is different from the time horizon for assessing the cost-effectiveness of an intervention for an individual as was described in the previous section.

The size of population does not alter the assessment of which treatment is cost-effective; however, it may be necessary to know the size of the population to estimate some elements of costs or health outcomes. The resources required to provide an intervention may include those purchased and utilised on a per-patient basis and those that are shared across individuals. The number of

patients who would utilise a joint resource determines how the costs can be apportioned at a per-patient level. For resource items that are shared between patients over successive time periods the total cost, C, can be converted to an equivalent cost per time period by dividing through by the annuity factor,

$$\frac{1-(1+d)^{-T_C}}{d},$$

which is a function of the lifetime of the resource item, T_C, and the decision-maker's rate of time preference, d. This cost per time period,

$$\frac{Cd}{1-(1+d)^{-T_C}},$$

can then be divided by the number of patients expected to utilise the resource in each time period to give the per-patient cost, c.[151] The per-patient costs of these joint resource items are thereby incorporated in the evaluation of cost-effectiveness at the individual level.

Table 37 uses the individual NHEs from *Table 36* and multiplies them by the population of patients who could receive treatment over three consecutive time periods. That is, it multiplies *Equation 2* by *Equation 3* evaluated over three consecutive time periods. Because the example deals with an acute disease the size of the population is determined by the incidence of the disease ($I_t = 100$). For simplicity we assume a discount rate of zero ($d = 0$). The time period over which the interventions could be utilised is 10 years ($T_{Pop} = 10$). This gives a total patient population, *Pop,* of 1000.

The time horizon over which the interventions could be utilised ($T_{Pop} = 10$) is split into three periods by considering two points in time at which reimbursement decisions can be revised ($t = 6,8$). For each population there remain the same two alternative interventions. In this example, applying *Equation 3*, population 1 consists of 600 patients and populations 2 and 3 each consist of 200 patients:

$$\mathrm{Pop}_1 = \sum_{t=1}^{6}\frac{I_t}{(1+d)^t} = 600\,;\ \mathrm{Pop}_2 = \sum_{t=7}^{8}\frac{I_t}{(1+d)^t} = 200\,;\ \mathrm{Pop}_3 = \sum_{t=9}^{10}\frac{I_t}{(1+d)^t} = 200$$

If nothing is expected to occur that would change the assessment of cost-effectiveness between each population the same decision will be made for each. The total overall population is $Pop = Pop_1 + Pop_2 + Pop_3$. Approving intervention 2 would be expected to offer $3413 + 1138 + 1138 = 5689$ units of net health benefit in the whole patient population, which represents an additional 408 units of population NHE compared with that offered by intervention 1 ($3169 + 1056 + 1056 = 5281$).

TABLE 37 Population NHEs

θ	Population 1		Population 2		Population 3	
	NHE(1,θ)	*NHE(2,θ)*	*NHE(1,θ)*	*NHE(2,θ)*	*NHE(1,θ)*	*NHE(2,θ)*
1	2475	2850	825	950	825	950
2	450	4050	150	1350	150	1350
3	3600	2250	1200	750	1200	750
4	6150	4500	2050	1500	2050	1500
$E_θ$	3169	3413	1056	1138	1056	1138

Analysis of subgroups

The relevant patient population is one that is homogeneous with respect to characteristics that affect the treatment decision. If the overall population is heterogeneous with respect to factors on which the treatment decision can be based and that affect the estimated NHEs of interventions it must first be split into homogeneous subgroups.[152] The cost-effectiveness of treatments must be evaluated separately in each distinct subgroup for which the treatment can be used, by applying the assessments described in *Chapter 5*. This allows different decisions to be made for each subgroup. When there are differences in the most cost-effective intervention between subgroups, basing decisions on the average across the whole population would lead to lower expected population net health benefits.

How does price affect cost-effectiveness?

Reducing price will make an intervention appear more cost-effective by increasing its expected NHEs. A straight discount will translate directly into increased NHEs. When more complex schemes are used to reduce the effective price of an intervention, the costs associated with setting up and utilising the scheme and the level of compliance must also be taken into account in the assessment of cost-effectiveness.

Are there significant irrecoverable costs?

When resources are committed in an earlier time period to that in which the ensuing benefits are realised these represent an investment. Similarly, investment may occur in health terms when patients incur initial health losses with the expectation of subsequent health gains. A typical example of this would be undergoing surgery to reduce the risk of death from coronary artery disease. Investment costs are irrecoverable if, once committed, they cannot be recovered should the decision to utilise an intervention be revised. This can be important if the decision is altered at a time earlier to that over which the expected future health gains were calculated in the assessment of cost-effectiveness described in *Chapter 5*. For example, if investment costs are irrecoverable and the decision is altered before the end of the anticipated life span of that investment the equivalent annual cost will have been calculated incorrectly and the number of patients who utilised the resource may be lower, meaning that the per-patient costs were underestimated when assessing cost-effectiveness. The irrecoverable investment costs will have been weighted against benefits that were not realised. However, in the case of surgery, once the patient has received the treatment a change of guidance does not alter his or her investment profile (the change of guidance does not modify the expected future change in health benefits attributed to having had surgery) and so does not have the same impact. The potential significance of irrecoverable costs can be judged:

1. according to whether or not estimates of cost-effectiveness would alter if the decision were to be revised sooner than anticipated
2. by assessing the probability that the decision might be altered earlier than expected
3. by comparing their size in health terms as a proportion of the total NHEs.

An assessment of irrecoverable costs requires some consideration of how health effects and costs accrue over time. This was shown in *Figure 34* at the patient level and an equivalent diagram can be constructed at the population level. It is then possible to determine a break-even point, T^*, the earliest time at which the treatment selected based on accumulated population NHEs up to T^* would match the treatment selected based on accumulated population NHEs up to T_{Pop}. Certain interventions may exhibit zero irrecoverable costs whereby health benefits are attained simultaneously with costs. If irrecoverable costs are a large proportion of overall NHEs this may imply a large initial negative NHE to be offset by future gains. If T^* is low then it will be less likely

that guidance will change before the cumulative incremental NHEs of adopting the intervention become positive. This would indicate that irrecoverable costs were not significant as they would have little influence on the alternative policy pay-offs. However, if T^* is large the impact of irrecoverable costs must be judged alongside an assessment of the likelihood that guidance will change, and this is addressed further later in this appendix (see *Impact of other sources of uncertainty* and *What is the impact of irrecoverable costs?*).

Suppose that the cost of intervention 2 includes an upfront irrecoverable investment of £20M. In health terms this upfront cost would be expected to displace health activities that would have generated 1000 net health benefits:

$$C(2) = \frac{£20,000,000}{k} = 1000.^\text{d}$$

In *Table 38* the NHE of this irrecoverable cost is separated out from the remaining costs and health effects.[e]

Irrecoverable costs as a proportion of total NHEs for intervention 2 are 1000/5689 = 18%. If Intervention 2 is selected and the decision is unchanged over the time horizon $T_{Pop} = 10 = T_C$, the impact of irrecoverable costs will have been adequately represented. The total NHEs from intervention 2 are 4013 + 1338 + 1338 – 1000 = 5689, which matches that calculated in *Cost-effectiveness at the population level* and is greater than that offered by intervention 1 (3169 + 1056 + 1056 = 5281). However, if the decision were changed earlier, at $t = 6$ (after population 1), the accumulated population NHEs for intervention 2 would be 4013 – 1000 = 3013, which is less than the NHEs that could have been realised with intervention 1 (3169).[f] If the guidance is altered at a time $t < T_C$, the investment costs will have been underestimated and the decision to approve intervention 2 may have been in error. If the decision is changed at $t = 8$ (after population 2), the accumulated population NHEs for intervention 2 would be equal to 4013 + 1338 – 1000 = 4351, which is greater than that offered by intervention 1 over the same time horizon (3169 + 1056 = 4225) and so intervention 2 has 'broken even'. In this example $T^* = 7.1$.

Is further research required?

The resource implications and health outcomes used to inform reimbursement decisions about the use of treatments within the health-care sector cannot be known with absolute certainty. Uncertainties stem from the reliance on sample information and the choice of methods to combine information from different sources and to translate what that body of information implies for the particular decision problem under consideration. Cost-effectiveness analysis can be used to estimate which of a set of mutually exclusive interventions might offer the greatest health gain and to describe the impact of uncertainty on those estimates.[153] *Table 35*

TABLE 38 Population NHEs and irrecoverable costs

θ	Irrecoverable costs		Population 1		Population 2		Population 3	
	NHE(1,θ)	NHE(2,θ)	NHE(1,θ)	NHE(2,θ)	NHE(1,θ)	NHE(2,θ)	NHE(1,θ)	NHE(2,θ)
1	0	−1000	2475	3450	825	1550	825	1550
2	0	−1000	450	4650	150	1950	150	1950
3	0	−1000	3600	2850	1200	1350	1200	1350
4	0	−1000	6150	5100	2050	2100	2050	2100
$E_θ$	0	−1000	3169	4013	1056	1338	1056	1338

shows a representation of the output from a cost-effectiveness analysis comparing two alternative interventions, $j = 1,2$. The range of plausible true values supported by the evidence used to calculate the costs, c, and health outcomes, h, for a typical patient who receives either intervention are described by θ, which can be labelled as a parameter or input that can take one of four possible values. As described in the previous section (see *Is the intervention cost-effective and what are the risks?*), cost-effectiveness can be determined by assessing expected NHEs. *Tables 35–37* also characterise the uncertainty associated with this decision by showing the range of values that θ could take and the NHEs that would result from those values.

A decision based on expected NHEs could be regarded as being in error if the true value of θ suggests that the alternative intervention to the one reimbursed would have offered greater health gains. The probability that a decision to recommend a particular intervention is in error is equal to the probability of observing a value of θ for which a different intervention would maximise NHEs. Knowing more about θ enables a decision-maker to reduce the cost of uncertainty by reducing the possibility of error. By avoiding reimbursing an intervention that turns out to be cost-ineffective, the decision-maker would expect better health gains overall.[154]

Consequences of uncertainty at the patient level

The best a decision-maker could do would be to select the intervention that maximised health gains for a particular realisation of θ, $\max_j NHE(j,\theta)$. Given the range of possible values for θ and the likelihood of observing those values it is possible to calculate the expected NHEs associated with this error-free choice:

$$E_\theta \max_j NHE(j,\theta) \qquad\qquad \text{[Equation 4]}$$

The difference between the NHEs that could have been achieved without error (see *Equation 4*) and the best choice based on current information (see *Equation 2*) describes the health consequences of this uncertainty at the individual patient level. This represents the maximum improvement in NHEs that could result from further research on θ and is often described as the EVPI:

$$EVPI_\theta = E_\theta \max_j NHE(j,\theta) - \max_j E_\theta NHE(j,\theta) \qquad\qquad \text{[Equation 5]}$$

Table 39 repeats the NHEs per patient treated from *Table 36* and in the final column presents the maximum NHE for each realization of θ. The final row shows the expected values assuming that each realisation of θ is equally likely.

As before, the expected value of intervention 2 is higher than that of intervention 1: $E_\theta NHE(1,\theta) < E_\theta NHE(2,\theta)$. This suggests that if θ is unknown it is better to use intervention 2. However, if θ were known to be 3 or 4 then it would be better to use intervention 1 because $NHE(2,3) < NHE(1,3)$, $NHE(2,4) < NHE(1,4)$, whereas if θ were known to be 1 or 2 then it would

TABLE 39 Expected and maximum NHEs per patient

θ	$NHE(1,\theta)$	$NHE(2,\theta)$	Max. $NHE(j,\theta)$
1	0.41	0.48	0.48
2	0.08	0.68	0.68
3	0.60	0.38	0.38
4	1.03	0.75	1.03
E_θ	0.53	0.57	0.69

Max., maximum.

be better to use intervention 2 because $NHE(1,1) < NHE(2,1)$, $NHE(1,2) < NHE(2,2)$. In this case the error probability is 0.5 (50%). The expected value of being able to make a decision knowing the exact value of θ is higher than the expected value of approving either intervention 1 or 2, suggesting that there is value in resolving θ:[g] $EVPI\theta = 0.69 - 0.57 = 0.12$.

Consequences of uncertainty at the population level

To judge the scale of the improvement in health benefits that might result from a decision to conduct further research it is necessary to consider the number of individuals who could benefit.[53] In other words, it is necessary to estimate the number of times a decision is made that utilises the information generated. This will be a function of the prevalence and incidence of the disease in question as shown in *Equation 3*, but it may also be a function of the extent of other diseases for which the information may be relevant. Any process for evaluating the value for money of interventions or research will involve assumptions and value judgements about the quality and relevancy of available evidence and which analytical methods are appropriate for determining what that evidence implies for the decision problem at hand. Also, there can be 'unknown unknowns', unexpected future events that we know can occur from past experience but that cannot be predicted or parameterised within the analysis. The relevant time horizon over which the information will be utilised, T_{EVPI}, can differ from the time horizon over which any particular intervention will be utilised, T_{Pop}.[53]

For simplicity, in the example we assume that the time horizon for information, T_{EVPI}, is coincident with the time horizon for the technology, T_{Pop}, and from this point denote both with T. We also assume that the relevant evidence is utilised only to make decisions between the two alternative interventions so far considered and therefore is utilised in the same number of decisions as the number of patients who receive treatment with either intervention. *Table 40* shows the individual-level NHEs from *Table 39* multiplied by the number of individuals who could benefit from either intervention or any additional information.

Using *Table 40* consider three simple policy options: to outright approve the new intervention (approve), to commission research (research) or to outright reject the new intervention (reject). The pay-offs from each policy in terms of the NHEs that accrue to each of the three patient populations $[\Pi(Pop_1), \Pi(Pop_2), \Pi(Pop_3)]$ and the total pay-off across all three populations (Π) are shown in *Table 41*.

The total NHE of approving and implementing $j = 2$ in all populations, in this example $3413 + 1138 + 1138 = 5688$, is more generally given by:

$$\Pi_{Approve} = \max_{j} E_{\theta} NHE(j,\theta).Pop \qquad \text{[Equation 6]}$$

TABLE 40 Expected and maximum population NHEs

θ	Population 1			Population 2			Population 3		
	$NHE(1,\theta)$	$NHE(2,\theta)$	Max. *NHE*	$NHE(1,\theta)$	$NHE(2,\theta)$	Max. *NHE*	$NHE(1,\theta)$	$NHE(2,\theta)$	Max. *NHE*
1	2475	2850	2850	825	950	950	825	950	950
2	450	4050	4050	150	1350	1350	150	1350	1350
3	3600	2250	3600	1200	750	1200	1200	750	1200
4	6150	4500	6150	2050	1500	2050	2050	1500	2050
E_{θ}	3169	3413	4163	1056	1138	1388	1056	1138	1388

Max., maximum.

TABLE 41 Policy pay-offs when research is immediately available

Policies	$\prod(Pop_1)$	$\prod(Pop_2)$	$\prod(Pop_3)$	$\prod$
1. Approve	3413	1138	1138	5688
2. Reject	3169	1056	1056	5281
3. Research	4163	1388	1388	6938

The total NHE of rejecting and continuing to implement $j = 1$ in all populations, in this example $3169 + 1056 + 1056 = 5281$, is more generally given by:

$$\prod\nolimits_{Reject} = E_\theta NHE(1,\theta).Pop \qquad \text{[Equation 7]}$$

As the NHE of $j = 2$ is higher than that of $j = 1$, the decision to approve is preferred to the decision to reject. The total NHE of being able to choose the appropriate treatment for each value of θ, in this example $4163 + 1388 + 1388 = 6938$, is given by:

$$\prod\nolimits_{Research} = E_\theta \max_j NHE(j,\theta).Pop \qquad \text{[Equation 8]}$$

The most beneficial policy option is therefore research, which allows the decision-maker to implement treatment based on knowing θ, resulting in total expected NHEs of 6938. The difference between the pay-off from research and the next best pay-off from approve describes the maximum value of further research and is equivalent to multiplying *Equation 5* by the size of the population to benefit (see *Equation 3*): $EVPI_\theta.Pop = \prod_{Research} - \prod_{Approve} = 6938 - 5688 = 1250$.

Subgroups

Just as cost-effectiveness can be assessed separately in different subgroups, so too can the value of further research. However, it may be necessary to consider whether information collected in one subgroup can be used to inform decisions in other subgroups. In these circumstances the number of patients who will benefit from the information can differ from the number of patients expected to benefit from each separate treatment decision. This should be reflected in calculating the value of further research.

Alternative scenarios

The example so far has considered uncertainty in the true value of θ, but uncertainty may also exist in how θ can be used to calculate cost and health outcomes. If there exists an alternative plausible characterisation of h and c as a function of θ this can produce modelling or structural uncertainty.[130,155,156] The alternative modelling or structural assumptions can be used to describe a series of scenarios within which expected NHEs will be estimated differently. From the decision-maker's perspective, the NHEs will now be a function of the choice of treatment, parameter uncertainty and the choice of scenario, $NHE(j,\theta,s)$. The assessment of cost-effectiveness in *Equation 2* is now made by taking the expectation of NHEs across the parameter uncertainty and the range and likelihood of the alternative scenarios, $\max_j E_\theta E_s NHE(j,\theta,s)$.

The best a decision-maker could do would be to select the intervention that maximised health gains for a particular realisation of θ and s. The expected NHEs associated with this error-free choice given in *Equation 4* become $E_\theta E_s \max_j NHE(j,\theta,s)$.

The added value of further research on parameter and scenario uncertainty can then be evaluated based on the results that integrate both parameter and scenario uncertainty.[h]

Table 42 presents the NHEs for each intervention based on competing plausible forms of the functions $h(j,\theta,2)$ and $c(j,\theta,2)$ to those that were presented in *Table 40* [which can be characterised as having presented $h(j,\theta,1)$ and $c(j,\theta,1)$].

In this alternative scenario the NHEs of intervention 2 are estimated to be smaller than those associated with intervention 1 and so it no longer appears cost-effective. Within this scenario, the value of resolving θ is $EVPI_{\theta|s=2}.Pop = 6406 - 5281 = 1125$, which differs to the value within scenario 1 (see *Table 40*) of $EVPI_{\theta|s=1}.Pop = 1250$.

If either scenario is equally plausible, the expected NHEs of each intervention overall can be found by averaging the NHEs from each scenario. For example, the expected NHEs of approving intervention 2, accounting for two equally likely scenarios, s, are calculated based on a 50% chance of achieving the NHEs from *Table 40* and a 50% chance of achieving the NHEs shown in *Table 42*: $\Pi_{Approve} = E_\theta E_s NHE(2,\theta,s) \times Pop = 0.5(3413 + 1138 + 1138) + 0.5(2513 + 838 + 838) = 4938$

As the expected population NHEs of intervention 1 are the same in both scenarios, $\Pi_{Reject} = E_\theta E_s NHE(1,\theta,s) \times Pop = 5281$; this would suggest that intervention 2 was not cost-effective ($\Pi_{Approve} < \Pi_{Reject}$).

Further research could be expected to resolve both parameter and structural uncertainty, meaning that the decision-maker could select the treatment that would provide the maximum NHEs given particular values of θ and s. When $\theta = 1$ the maximum NHE is associated with intervention 2 in scenario 1 (see *Table 40*) and intervention 1 in scenario 2 (see *Table 42*). When $\theta = 2$ the maximum NHE is associated with intervention 2 in both scenarios. When $\theta = 3$ the maximum NHE is associated with intervention 1 in both scenarios. Averaging across these maximum values gives $\Pi_{Research} = E_\theta E_s \max_j NHE(j,\theta,s) = 6672$. The value of research that would resolve both parameter and decision uncertainty is found by comparing $E_\theta E_s \max_j NHE(j,\theta,s)$ to the next best achievable pay-off of 5281. Thus, the value of research that would resolve both parameter and scenario uncertainty is $6672 - 5281 = 1391$.

For the remainder of this appendix we assume that there is no alternative plausible set of modelling assumptions.

Does more research seem worthwhile?

Although resolving decision uncertainty results in the most beneficial outcome, whether or not further research should be undertaken will depend on the costs of carrying out the research and who will be responsible for these costs (e.g. the manufacturer or public funding bodies). When the costs of research fall on the budget constraint for providing health care this will displace other activities that could have generated health. This research cost must be subtracted from the policy

TABLE 42 Population NHEs with alternative assumptions (scenario 2)

θ	Population 1			Population 2			Population 3		
	$NHE(1,\theta)$	$NHE(2,\theta)$	Max. *NHE*	$NHE(1,\theta)$	$NHE(2,\theta)$	Max. *NHE*	$NHE(1,\theta)$	$NHE(2,\theta)$	Max. *NHE*
1	2475	1950	2475	825	650	825	825	650	825
2	450	3150	3150	150	1050	1050	150	1050	1050
3	3600	1350	3600	1200	450	1200	1200	450	1200
4	6150	3600	6150	2050	1200	2050	2050	1200	2050
E_θ	3169	2513	3844	1056	838	1281	1056	838	1281

Max., maximum.

pay-off from conducting further research. If the maximum value of research does not exceed the expected costs this is a sufficient condition to decide against investing in further research.[67]

The value of research is established by comparing policy 3, research, with the next best alternative, which is policy 1, approve the new technology intervention 2. At the patient level this is equivalent to the $EVPI_\theta$ detailed in *Equation 5*. Scaling up by the size of the population that would benefit from any additional information generated gives a maximum value for research of 1250 ($EVPI_\theta \times Pop$). Using the same threshold, $k = £20,000$, the costs of research can be expressed in terms of health displaced, R. If the research costs are expected to exceed 1250 units of health benefit (or in monetary terms £25M), then policy 1 would be more beneficial than policy 3. Similarly, this would provide a sufficient condition for the decision to avoid the analysis time required for estimating more accurately the true value of research, which is discussed further in the following section (see *How long until the research reports?*). More generally, the pay-off from conducting additional research can be calculated as:

$$\prod_{Research} = E_\theta \max_j NHE(j,\theta).Pop\text{-}R \qquad \text{[Equation 9]}$$

How does price affect the need for evidence?

If the decision to immediately approve an intervention does not seem to be the most valuable policy option, the manufacturer can alter the balance between the alternative decisions by adjusting the price. Reducing price will make an intervention appear more cost-effective by increasing its expected NHEs. A price change will also affect the value of conducting further research.[70] The subtraction of a constant does not reduce the variance in NHEs for a given intervention, but it will shift the distribution of the NHEs expected from the new intervention relative to the next best alternative and hence alter the distribution of incremental NHEs. This can change the probability of error and/or the value of the health benefits forgone if the wrong intervention were reimbursed, and can make research appear either more or less valuable. When price change reduces (increases) the proportion of incremental NHEs that are less than zero, the decision uncertainty is reduced (increased) following a price reduction. For an intervention that is expected to be cost-effective a price reduction will generally reduce the value of research because the probability of error and the health forgone as a consequence of error will both fall. This is because the health forgone given that the decision to reimburse was wrong is calculated by comparing the NHEs of the intervention that was expected to be cost-effective with the NHEs of an alternative, and this difference will have been reduced by the price reduction. For an intervention that is not expected to be cost-effective a price reduction will initially increase the probability of error but may reduce the health forgone as a consequence of error, and the overall impact on the value of further research is more difficult to predict. This is because the alternative interventions compared to calculate the amount of health forgone may not always have included the intervention that was subject to a change in price. As price is bounded at zero there is a limit to how much price adjustments can be used to increase NHEs and alter estimates of the value of research.

For example, suppose a research design is proposed costing £20M. If this money were diverted from existing health-care activities it would be expected to displace 1000 units of health benefit ($R = £20M/k = 1000$). On the basis of *Table 41* research would seem to be the most valuable policy option, offering an expected pay-off in NHEs of $\prod_{Research} = 6938 - 1000 = 5938$. However, if the price of intervention 2 dropped from £1000 to £800 the pay-offs from the alternative policy options would alter. The impact of the price change is shown in *Table 43*.

The expected NHEs for an individual receiving intervention 2 have increased from 0.57 to 0.67. There is still value in knowing θ because if θ were known to be equal to 3 or 4 then intervention 1

offers greater NHEs. After multiplying by the size of the population to benefit, the pay-offs from the policy options incorporating the new expected NHEs are shown in *Table 44*.

After the price change the maximum value of research for the population has fallen to $EVPI_\theta \times Pop = 7438 - 6688 = 750$, and this is no longer sufficient to outweigh the health benefits displaced by imposing research costs on the budget constraint. In this particular example, a price reduction has created a sufficient condition for immediate approval without further research. The threshold price for intervention 2 at which the optimal policy decision changes from research to approve is £900.

Figure 35 displays the pay-offs in terms of NHEs from the policy options in *Table 41* as a function of the price of intervention 2, assuming that $R = 1000$.

The pay-off from reject is unaffected by the price whereas the pay-off from immediate approval is a linear function of the price of intervention 2. When the price of intervention 2 exceeds £1081 it is no longer cost-effective ($\Pi_{Reject} > \Pi_{Approve}$). The pay-off from research is a non-linear function of price in this example. The closer the price is to £1081, the more similar the NHEs from both treatments, hence the greater the decision uncertainty and the probability of error but the smaller the opportunity cost of selecting the wrong alternative. The policy with the highest pay-off switches as the price of intervention 2 increases. The figure shows how the optimal policy switches from approve to research at a price of £900. Once the price exceeds £1400 the optimal policy switches from research to reject.

Notice from *Tables 36* and *43* that intervention 2 is associated with better health gains than intervention 1 for every value of θ. In other words, intervention 2 is known to be more effective than intervention 1. The uncertainty around whether intervention 2 is more cost-effective stems from whether the magnitude of the health gains is sufficient to justify the additional expense of intervention 2. A price reduction directly addresses this uncertainty by reducing the price of intervention 2, but does not have any impact on the health outcomes expected for patients treated with each intervention. However, for some comparisons the uncertainty around whether or not the new intervention is cost-effective will stem from if it improves or worsens health in

TABLE 43 Net health effects per patient after price reduction for intervention 2

θ	Intervention 1			Intervention 2			Max. *NHE*
	$h(1,\theta)$	$c(1,\theta)$	$NHE(1,\theta)$	$h(2,\theta)$	$c(2,\theta)$	$NHE(2,\theta)$	
1	0.70	£5750	0.41	1.00	£8500	0.58	0.58
2	0.40	£6500	0.08	1.30	£10,500	0.78	0.78
3	0.90	£6000	0.60	1.10	£12,500	0.48	0.60
4	1.30	£5500	1.03	1.40	£11,000	0.85	1.03
E_θ	0.83	£5938	0.53	1.20	£10,625	0.67	0.85

Max., maximum.

TABLE 44 Policy pay-offs after price reduction for intervention 2

Policies	$\Pi(Pop_1)$	$\Pi(Pop_2)$	$\Pi(Pop_3)$	Π
1. Approve	4013	1338	1338	6688
2. Reject	3169	1056	1056	5281
3. Research	4463	1488	1488	7438
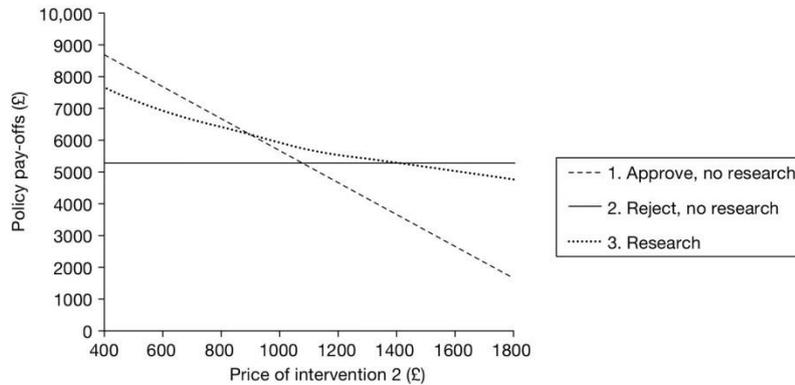
FIGURE 35 Impact of price on optimal policy decision (immediate research).

the patients that receive treatment. A price reduction will still have the impact of making such an intervention appear more cost-effective because it subtracts a constant from the cost of an intervention, and this constant could be equivalently expressed in health terms (by utilising the threshold, k).

How long until the research reports?

The results of further research will in general not be available immediately, meaning that there exists a population of patients for whom the treatment decision must be made based on current evidence. Thus, there is a need to make a reimbursement decision for this initial population (approve or reject) in combination with a decision about whether or not to collect further research. This means that research alone is not a policy option, and the relevant alternatives are approve (without further research), approve with further research (AWR), OIR (i.e. reject and conduct further research) and reject.

The following general notation is for calculating the policy pay-offs when research will report at time $t = \tau$. The pay-off from approve remains as described in *Equation 6*. If research was undertaken after approval (AWR) then the initial population receives the NHEs associated with intervention 2, but subsequent populations receive the maximum NHE because the uncertainty is resolved:

$$\Pi_{AWR} = \max_j E_\theta NHE(j,\theta).\sum_{t=1}^{\tau}\frac{I_t}{(1+d)^t} + E_\theta \max_j NHE(j,\theta).\sum_{t=\tau}^{T}\frac{I_t}{(1+d)^t} \qquad \text{[Equation 10]}$$

If research is undertaken after reject (OIR) then the initial population receives the NHE associated with intervention 1 whereas subsequent populations receive the maximum NHE:

$$\Pi_{OIR} = E_\theta NHE(1,\theta).\sum_{t=1}^{\tau}\frac{I_t}{(1+d)^t} + E_\theta \max_j NHE(j,\theta).\sum_{t=\tau}^{T}\frac{I_t}{(1+d)^t} \qquad \text{[Equation 11]}$$

If intervention 2 is rejected and no further research is undertaken then each population will receive the NHE associated with intervention 1 as described in *Equation 7*.

The numbers in *Table 45* are calculated from *Table 40* and show the policy pay-offs when research would report at $\tau = 6$ (results of research available for populations 2 and 3).[i] For example, if research was undertaken after approval then population 1 receives the NHEs associated with intervention 2, but populations 2 and 3 receive the maximum NHE because the uncertainty is resolved. If research is undertaken after rejecting intervention 2 then population 1 receives the

TABLE 45 Policy pay-offs when results of research are available for populations 2 and 3

Policies	$\prod(Pop_1)$	$\prod(Pop_2)$	$\prod(Pop_3)$	$\prod$
1. Approve (no research)	3413	1138	1138	5688
2. AWR (approve and research)	3413	1388	1388	6188
3. OIR (reject and research)	3169	1388	1388	5944
4. Reject (no research)	3169	1056	1056	5281

NHE associated with intervention 1 and populations 2 and 3 receive the maximum NHE. If intervention 2 is rejected and no further research is undertaken then each population will receive the NHE associated with intervention 1.

In this example research is better than not doing research as different values of θ imply different reimbursement decisions. As was noted earlier (see *Does more research seem worthwhile?*, *How does price affect the need for evidence?*) there is a 50% chance that once the results of research are known the intervention regarded as cost-effective will change. As $\tau < T$ this implies a 50% chance that guidance will change before the end of the lifetime of the technology following a policy decision of AWR or OIR. If we assume that research costs do not fall on the budget constraint for health care then AWR will always dominate approve and OIR will always dominate reject. The value in collecting further evidence is demonstrated by the fact that the total expected pay-off in terms of health effects from policy 2 exceeds that of policy 1 ($\Pi_{AWR} > \Pi_{Approve}$), and similarly the expected health benefits from policy 3 exceed those offered by policy 4 ($\Pi_{OIR} > \Pi_{Reject}$). Given the simple example presented so far, AWR will always dominate OIR when the new alternative is regarded as cost-effective (and vice versa when the new alternative is regarded as cost-ineffective).

However, given a cost of research of $R = 1000$ that does fall on the budget constraint for health care, the additional net health benefits from research are insufficient to outweigh the opportunity cost of gathering that information. When research costs are subtracted from the policies that incorporate research in *Table 45* the pay-off from policy 1 exceeds that from policy 2 $\Pi_{Approve} = 5688$ compared with $\Pi_{AWR} - R = 6188 - 1000 = 5188$) and the pay-off from policy 4 exceeds that from policy 3. It is still better to approve than reject because intervention 2 is cost-effective in this example.

How does price affect the choice between approve, 'approval with research', 'only in research' and reject?

Recall from *Does more research seem worthwhile?*, *How does price affect the need for evidence?* that for a cost-effective intervention a price reduction will generally reduce the value of research and hence improve the pay-off from approve relative to AWR. Also, a price reduction makes an intervention appear more cost-effective, improving the pay-off from approve relative to reject (see *Is the intervention cost-effective?*, *How does price affect cost-effectiveness?*). The NHEs that accrue to patients after research reports are assumed to be the same in AWR and OIR (although this will be relaxed later; see *Will the research be conducted?*), and so the threshold price at which the pay-offs from AWR and OIR are equivalent is the same as the threshold price for cost-effectiveness, that is, where the pay-offs from approve and reject are equivalent. However, depending on the value of further research and how price change affects this value, for a technology that is not expected to be cost-effective the initial decision may be reject or OIR and as the price falls the optimal decision could switch between reject and OIR before changing to AWR or straight to approve when the price is low enough for the technology to be regarded as cost-effective.

The pay-offs in *Table 45* are based on a price of £1000 for intervention 2. *Figure 36* shows how the optimal policy changes as the price of intervention 2 changes, assuming research costs of $R = 1000$.

The policies that entail research (AWR and OIR) do not offer the highest pay-off for any price of intervention 2 and hence the optimal policy switches from approve to reject at a price of £1081. Above this price approve (AWR) ceases to dominate reject (OIR) because intervention 2 is no longer cost-effective. When the difference between AWR and approve, or between OIR and reject, reaches its nadir it is equal to minus the cost of research (e.g. $\Pi_{AWR} - \Pi_{Approve} = -R$), which indicates that there is no cost of uncertainty. When the difference is not equal to minus the cost of research, this indicates that there were benefits of research but these were insufficient to outweigh those costs.

How does time to research affect the choice between approve, 'approval with research', 'only in research' and reject?

The numbers in *Table 46* are also calculated from *Table 40*. In this case further research takes longer to report ($\tau = 8$) and so only population 3 benefits from resolving θ.

Table 46 demonstrates that the benefits of research decrease if it takes longer to report. As the population that benefits from research decreases the decision-maker is more likely to select policy 1 over policy 2 or policy 4 over policy 3.

Recall that in the simple example presented so far, AWR will always dominate OIR when the new alternative is regarded as cost-effective and OIR will always dominate AWR when the new
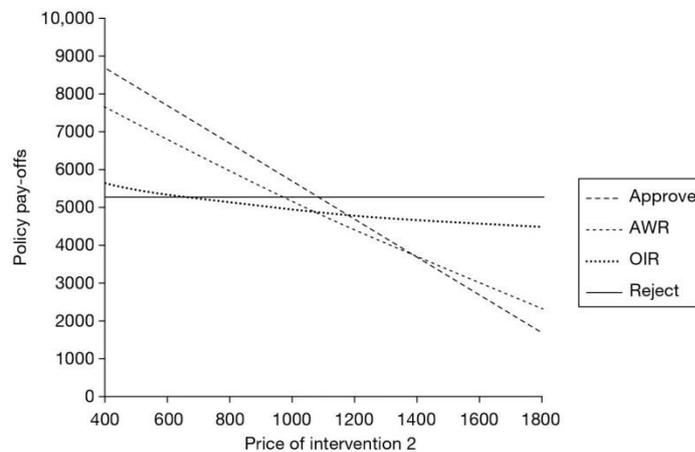
FIGURE 36 Impact of price on optimal policy decision (research available Population 2).

TABLE 46 Policy pay-offs when results of research are available for population 3

Policies	$\Pi(Pop_1)$	$\Pi(Pop_2)$	$\Pi(Pop_3)$	Π
1. Approve (no research)	3413	1138	1138	5688
2. AWR (approve and research)	3413	1138	1388	5938
3. OIR (reject and research)	3169	1056	1388	5613
4. Reject (no research)	3169	1056	1056	5281

alternative is regarded as cost-ineffective. This will not always be the case. In the next section we consider the interaction between the reimbursement decision and the ability to conduct the required type of research.

Is the research possible if the intervention is approved?

The prospects of acquiring information through research may be affected by the decision to approve for a number of reasons: (1) the adoption of a technology removes incentives on the manufacturer to conduct further research, (2) the early diffusion of a technology means that future clinical trials are less likely to be supported or regarded as ethical by the clinical community, even when public funds are made available for such research and (3) patients are unlikely to enrol in clinical trials once they have unrestricted access to the new technology. By rejecting a new technology in favour of current practice, the prospects for future research are unaffected: the manufacturer of the new technology retains an incentive to conduct research to obtain approval and no change in clinical practice occurs to damage existing trials or remove the ethical basis for future research.[70]

If the new intervention is expected to be cost-effective but research is not possible if the new intervention is approved then policy 2 in *Table 45* is no longer available. Even though the new intervention is cost-effective it may be preferable to reject and then carry out research (OIR) rather approving with no research (approve). This is the case in *Table 45* where the pay-off from policy 3 exceeds the pay-off from policy 1. In general, the balance between OIR and approve will depend on whether or not the increase in population health benefits that arises from resolving uncertainty exceeds the increase in population health benefits from approving the new technology in the population of patients who receive treatment before research reporting. The size of this trade-off is determined by the extent of cost-effectiveness, the amount of time it takes for research to report or the size of the populations affected and the value of resolving uncertainty. If research takes longer to report, as in *Table 46*, then policy 3 no longer dominates policy 1 and a sufficient condition for immediate approval is met. In the case in which the new technology is not cost-effective the optimal decision is unchanged by the fact that evidence cannot be gained after approval, as it depends only on whether there is an additional benefit of carrying out further research (see *Is further research required?*).

Some types of uncertainty, such as long-term safety, are easier to resolve if the new technology is approved. If it is the case that research cannot be undertaken without approval then policy 3 in *Table 45* is not an option. If the new intervention is cost-effective the decision does not change and the choice between policies 1 and 2 depends on the value of resolving the uncertainty. If, however, the new intervention is not cost-effective it might be the case that policy 2, AWR, is preferred to policy 4, reject with no research. This would occur if the value of resolving the uncertainty is greater than the health benefits lost from displaced activities as a result of approving an intervention that is not expected to be cost-effective. In the later section *What type of evidence is needed?* we consider the different types of uncertainty that may require different research designs and how these may be affected differently by the decision to approve a technology for widespread use.

Will the research be conducted?

As already discussed, the potential value of research depends on the likelihood that it will be conducted and the timing of when it will report. For reasons just described, in some cases further research is not possible if the new technology is approved. In other cases the research might be possible but may still not be conducted if the decision-maker must rely on other agents to commission or complete the required research. For example, the manufacturer may not have an incentive to conduct socially valuable research if it would not increase profits, and this is more likely for a technology that has already been approved. Even once research is under way studies may fail or remain unreported because of reasons outside the decision-maker's control.

To incorporate the uncertainty of whether research will be conducted and/or reported we introduce α, the probability that the results of further research will become available. In *Table 45* it is assumed that α = 1, but if it is < 1 then the reduced probability of research means that the expected NHEs from policies 2 and 4 will be lower. In reality policy 2 is weighted by the probability of no research being undertaken.

The expected NHEs for policies that entail further evidence collection (AWR and OIR) incorporate the expected benefits received by patients after research reports. The pay-offs from these policies were calculated in *How long until research reports?* in which the benefits of research expected to report at time τ were estimated within *Equations 10* and *11* by

$$E_\theta \max_j NHE(j,\theta).\sum_{t=\tau}^{T}\frac{I_t}{(1+d)^t}.$$

Incorporating the probability that research will be conducted, this expression now becomes, in the case of AWR:

$$\Pi_{AWR} = \max_j E_\theta NHE(j,\theta).\sum_{t=1}^{\tau}\frac{I_t}{(1+d)^t} +$$

$$(1-\alpha)\max_j E_\theta NHE(j,\theta).\sum_{t=\tau}^{T}\frac{I_t}{(1+d)^t} + \alpha.E_\theta \max_j NHE(j,\theta).\sum_{t=\tau}^{T}\frac{I_t}{(1+d)^t} \qquad \text{[Equation 12]}$$

and in the case of OIR:

$$\Pi_{OIR} = E_\theta NHE(1,\theta).\sum_{t=1}^{\tau}\frac{I_t}{(1+d)^t} +$$

$$(1-\alpha)E_\theta NHE(1,\theta).\sum_{t=\tau}^{T}\frac{I_t}{(1+d)^t} + \alpha.E_\theta \max_j NHE(j,\theta).\sum_{t=\tau}^{T}\frac{I_t}{(1+d)^t} \qquad \text{[Equation 13]}$$

Given that the number of potentially worthwhile research proposals exceeds available research funding, the probability of any particular proposal being funded will depend on its value relative to the relevant alternative research proposals. Because of the interaction between the adoption decision and the prospects for further research, as discussed in *Is the research possible if the intervention is approved?*, the probability of research following a decision to approve, $\alpha_{Approve}$, may differ from the probability of research following a decision to reject, α_{Reject}. If $\alpha_{Approve} > \alpha_{Reject}$ then AWR would be expected to yield higher NHEs for patients after research has reported than OIR. The section *Is the research possible if the intervention is approved?* considers the extreme situation in which $\alpha_{Approve} = 0$ and $\alpha_{Reject} = 1$ and AWR is not an option, but if $0 < \alpha_{Approve} < \alpha_{Reject}$ the pay-off from AWR could be either higher or lower than that of OIR depending on the trade-off between greater upfront benefits for patients receiving a more cost-effective intervention and lower expected benefits from further research for later populations.

When might it be commissioned and report?

As previously discussed (see *How long until research reports?*), the length of time until research reports, τ, affects the size of the population that benefits from immediate approval or rejection and the size of the population that benefits from further research. Previously we considered τ to be at *t* = 6 (research available for population 2) in *Table 45* or at *t* = 8 (research available for population 3) in *Table 46*. As discussed in the previous section, future research may be regarded as a chance event, with a probability of occurrence, α, that can be < 1. Based on an assessment of cost-effectiveness and an assessment of the value of further research it is possible to describe the range of possible values of τ and α for which the NHEs of immediate approval will exceed those of initial rejection of a given intervention.

Figure 37 shows an example in which the possibility of research if the new technology is approved is zero ($\alpha_{Approve} = 0$). When this is the case policy option 2 in *Tables 45* and *46* is no longer available. An initial decision to reject the new intervention (e.g. reject or OIR) is assumed to not affect the prospect for research, but research is not certain to proceed for the reasons discussed earlier. For combinations of τ and α_{Reject} located to the north-east of the boundary in *Figure 37*, the benefits of immediate approval exceed the health benefits forgone from the reduced possibility of research, and a necessary condition for immediate approval is met. This is because the pay-off from approve exceeds the policy pay-off from OIR based on the maximum value of further research that would resolve all of the uncertainty that was characterised in the evaluation of NHEs. So in *Table 46*, when research is available for population 3 and is certain to occur, the policy to approve exceeds the value of OIR. The point at which $\tau = 2$ and $\alpha_{Reject} = 1$ is to the north-east of the boundary and a sufficient condition for immediate approval is met. If research takes a shorter time to report, as in *Table 45*, then as more patients stand to benefit from the additional information OIR increases in value relative to approve. The point at which $\tau = 1$ and $\alpha_{Reject} = 1$ is located inside the boundary and a sufficient condition for immediate approval is not met.

The value of α can be inferred from system factors and previous experience. The value of α will be higher if the decision-making body determining approval also has the remit to commission further research and higher still if it has access to funding. In the scenario in which only the manufacturer has funding for research, incentives must be properly aligned to assume that α will be high. Even when research is under way, α may still not be equal to 1 if there is a chance that the research will not be reported. These factors will also influence the time until research reports. Research requiring large populations or undertaken in diseases with a low incidence and prevalence may be hindered by recruitment. Certain research questions will require longer follow-up and different types of research may be more complicated and time-consuming to execute (see *What type of evidence is needed?*).

Figure 37 allows decision-makers to incorporate their own values for τ and α. Alternatively, a quantitative analysis could describe the values of τ and α supported by available evidence by assigning to them probability distributions, which could be correlated if appropriate.[70] These could be utilised in the calculation of policy pay-offs described in *Will the research be conducted?* to estimate the pay-off from AWR or OIR conditional on τ and α. The decision-maker can then compare the expected policy pay-offs. However, this type of analysis provides a sufficient but not necessary condition for approval, that is, if approve offers the highest pay-off then it will be optimal to approve the intervention, but in some cases in which approve does not offer the highest pay-off it would still have been optimal to approve. Equivalently, it only provides a necessary condition for OIR because the research, if and when it does report, will not resolve all uncertainty. In other words, when OIR or AWR offers the highest pay-off it is not always the
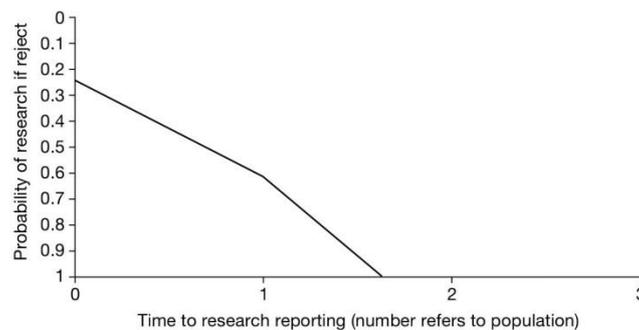

FIGURE 37 A sufficient boundary for immediate approval.

case that it is the optimal policy. Therefore, a sufficient condition for AWR or OIR would require consideration of how much uncertainty is likely to be resolved by the research.

How much of the uncertainty will the research resolve?

So far we have discussed the assessments that can be made about the value of further evidence. The assessments describe the value of further evidence that would eliminate all uncertainty. This is the maximum value that additional information could be expected to provide. A single study, no matter the size, is unlikely to take all of the measurements of interest to inform a cost-effectiveness analysis that draws together information on costs, quality of life, prognosis and efficacy, and this issue is addressed in the following section, which discusses the assessments required to describe the value of further evidence that would eliminate all uncertainty for only some of the elements of the decision problem. In practice, of course, new evidence would be expected to lessen rather than eliminate uncertainty.[157] When taking measurements from a sample of the relevant population, a larger sample would produce a more certain result, but unless we measure the entire population some measurement uncertainty will remain. However, it is important to note that this value uncertainty does not necessarily result in decision uncertainty.[153] Regarding *Table 36*, if further research reduced the uncertainty in θ so that it was known to take a value of either 1 or 2, then measurement error would remain but the decision uncertainty would be eliminated. Intervention 2 would have higher NHEs than intervention 1 no matter how θ resolved and the probability of error would be 0. Finally, there exist sources of uncertainty that cannot be resolved through further research. Although some of these may be resolved over time (see *Impact of other sources of uncertainty*), others will remain.

The process of assessing the value of sample information is simply that of asking what changes we could expect in our decisions if we obtained further information from additional research. The first step is to predict the possible outcome of this hypothetical new research. The second step is to incorporate the new evidence with the existing evidence to get an updated estimate of expected costs and health outcomes in order to record the difference in the NHEs we would expect from a revised decision compared with our original decision. Finally, as we could imagine many possible outcomes for the new study, we must repeat steps one and two until the full range of possible outcomes is adequately described. We can then compare the expected value of our revised decision with additional sample information with the value of the decision based on current evidence.[157] Although this process may be conceptually straightforward, in current practice with current computing capabilities it can entail very long analysis times, making it sometimes impractical for informing decision-making in a timely manner. This is especially true when considering the task of evaluating the value of the full range of possible study designs that could potentially inform the decision problem.

The expected gain in health from the additional information provided by future studies must be balanced against the cost of those studies. The notion of who would bear the costs of research was discussed in *Does more research seem worthwhile?*, but regardless of this the opportunity costs to patients included in research must be accounted for. Additional experimental research may assign recruited patients to receive interventions that they would otherwise not have had access to. The benefits that accrue to these research participants may differ from the benefits that accrue to those in the wider population, and so should also be assessed separately. If a RCT comparing two interventions recruits a total of n patients with a 1:1 allocation procedure then $0.5n$ patients will have different expected NHEs from the remaining $Pop - 0.5n$, and this should be reflected in any calculations. It may also be necessary to consider whether additional investment is required to ensure that the results of research will be translated into clinical practice.[73,158]

Recall from the section *Is further research required?* that it is possible to establish a necessary condition for immediate approval on the basis of establishing the value of research that would

eliminate all uncertainty. Thus, in *Figure 37*, if the decision-maker suspects that the likelihood and timing of further research are such that the pay-off from approve would be above the boundary, the intervention should be approved for immediate use. However, for points inside the boundary a sufficient condition for OIR is not met. Points inside a boundary that was drawn based on the expected value of sample information would provide a sufficient condition for OIR, but this boundary may not be estimable within reasonable analysis time for reasons already described. Nonetheless, an assessment of the additional value that research would generate could be used to draw a boundary that provides a necessary and sufficient condition for immediate approval, as shown by the dashed line in *Figure 38*. Such information could be a useful guide to decision-makers in possession of the expected value of a particular proposed study and would allow them to incorporate different views on the timing and likelihood of that research study.

Alternatively, given a pair of values for τ and α, it is possible to calculate what value of sample information would generate a boundary line that would pass through that point. This describes the minimum additional value that further research must generate to make OIR the optimal policy decision. In other words, when the sufficient condition for approval is not met we could ask how few additional health benefits would have to be generated from further sample information to mean that a given point would be to the north-east of a boundary based on sample information.

In *Figure 38*, if the decision-maker believes that research would be available for population 2 and thinks that there is an 100% chance of research if they reject the new technology, they would not be convinced that the benefits of immediate approval could exceed the value of evidence forgone should approval reduce the probability of research to zero (see *Table 45* and *Figure 37*; $\tau = 1$ and $\alpha_{Reject} = 1$ lies to the south-west of the boundary). However, this boundary is drawn considering the maximum benefit of further evidence, with which a policy of OIR generates $5944 - 5688 = 256$ additional units of NHE relative to approve. As can be seen from *Table 45*, the policy of OIR generates an opportunity cost of $3169 - 3413 = -244$ in the first period. However, after research has reported for populations 2 and 3 it generates additional expected NHEs of $(1388 + 1388) - (1138 + 1138) = 500$, outweighing the initial loss. New evidence must generate additional value of at least 244 units of NHE to outweigh the initial opportunity cost for OIR to dominate approve. The decision-maker must then assess whether they believe that new research would provide greater health benefits than this cut-off.
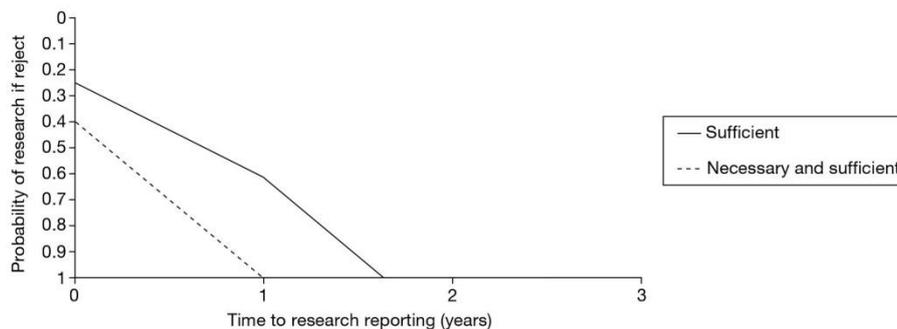
FIGURE 38 Necessary and sufficient boundaries for immediate approval.

How does price affect the need for evidence?

In *Does more research seem worthwhile?* it was shown that reducing the price of a new technology would increase the benefits of immediate approval and reduce the benefits of further research, shifting the optimal policy decision from OIR or AWR to approve. The impact of a price reduction is to shift the sufficient boundary for immediate approval to the south-west. This is shown in *Figure 39*.

The price change from £1000 to £800 means that the point at which $\tau = 1$ and $\alpha_{Reject} = 1$ goes from being south-west of the original boundary to being north-east of the new boundary, and a sufficient condition for immediate approval is met.

How does current uncertainty affect the need for evidence?

The manufacturer of an intervention can also move this boundary for immediate approval by providing additional evidence to support the use of the technology. Reducing current uncertainty causes the added value of gathering further evidence to diminish, shifting the boundary to the south-west. Suppose that the manufacturer of intervention 2 supplied the results of an additional study that reduced the uncertainty about the increase in NHEs relative to intervention 1. *Table 47* shows the NHEs associated with each intervention incorporating the additional information.

The expected health outcomes and costs of intervention 2 are similar to those in *Table 36* but the uncertainty is reduced. Additional information would now lead to a change in decision only if $\theta = 4$, and so the error probability has fallen from 0.5 to 0.25. The consequences of this uncertainty have fallen to less than 0.06 units of NHE per individual. *Figure 40* shows the impact of this on the boundary for immediate approval, after multiplying these individual NHEs by the relevant population to benefit.

In this example the additional evidence in *Table 47* shifts the boundary to the south-west, but not by enough to make immediate approval appear optimal when $\tau = 1$ and $\alpha_{Reject} = 1$. In other words, the policy of OIR would dominate the policy of approve. However, this boundary is drawn based on the maximum value of further research, and as discussed earlier it is shown that a boundary based on the actual value of further research would lie further to the south-west. Unless the decision-maker believed that further research would provide NHEs very close to this maximum value, the reduction in current uncertainty would result in a sufficient and necessary condition for immediate approval.
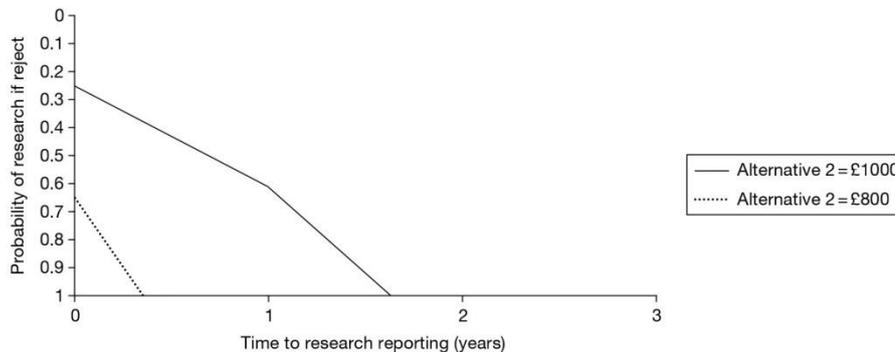
FIGURE 39 Impact of price change on sufficient boundary for immediate approval.

TABLE 47　Net health effects per patient with additional evidence for intervention 2

θ	Intervention 1			Intervention 2			Max. *NHE*
	h(1,θ)	*c*(1,θ)	*NHE*(1,θ)	*h*(2,θ)	*c*(2,θ) (£)	*NHE*(2,θ)	
1	0.70	£5750	0.41	1.0	£11,000	0.45	0.45
2	0.40	£6500	0.08	1.2	£15,000	0.45	0.45
3	0.90	£6000	0.60	1.3	£14,000	0.60	0.60
4	1.30	£5500	1.03	1.3	£10,500	0.78	1.03
E_θ	0.83	£5938	0.53	1.2	£12,625	0.57	0.63

Max., maximum.
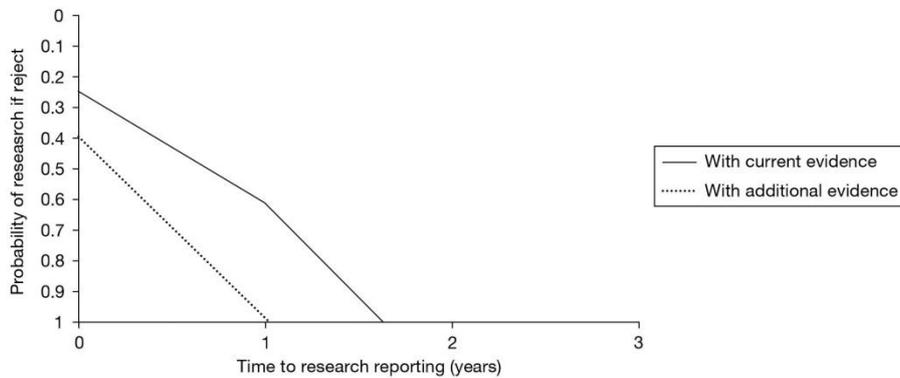
FIGURE 40　Impact of additional evidence on sufficient boundary for immediate approval.

What type of evidence is needed?

Earlier we discussed how to determine whether further research is worthwhile (see *Is further research required?*). It would also be useful to know what type of additional evidence would be most valuable. In this section we expand our established framework to show the value of different types of information and the importance of the sequence of data collection. This information can be used to prioritise data collection but also becomes important in determining a decision to approve if certain types of research are not possible with approval (see *Is the research possible if the intervention is approved?*).

We can expand *Table 39* to describe a situation in which the available evidence is made up of two uncertain parameters, $\theta = \theta_1, \theta_2$, resulting in *Table 48*. The first row of *Table 48*, when $\theta = 1$, is a result of $\theta_1 = 1$ and $\theta_2 = 1$. In this example, if $\theta_2 = 1$ then intervention 2 is preferred if $\theta_1 = 1$ but intervention 1 is preferred if $\theta_1 = 2$.

As was done previously in *Table 37*, in *Table 49* we multiply the individual NHEs by the size of the relevant patient population (given by *Equation 3*), split into three time periods.

The expected values assume that each value is equally likely for both θ_1 and θ_2. With more than one type of uncertainty it is possible that further research would gather information on a subset

TABLE 48 Net health effects per patient for two types of uncertainty

θ	θ₁	θ₂	NHE(1,θ)	NHE(2,θ)	Max. NHE
1	1	1	0.41	0.48	0.48
2		2	0.08	0.68	0.68
3	2	1	0.60	0.38	0.60
4		2	1.03	0.75	1.03
$E_θ$			0.53	0.57	0.69

Max., maximum.

TABLE 49 Population NHEs for two types of uncertainty

θ	θ₁	θ₂	Population 1			Population 2			Population 3		
			NHE(1,θ)	NHE(2,θ)	Max. NHE	NHE(1,θ)	NHE(2,θ)	Max. NHE	NHE(1,θ)	NHE(2,θ)	Max. NHE
1	1	1	2475	2850	2850	825	950	950	825	950	950
2		2	450	4050	4050	150	1350	1350	150	1350	1350
3	2	1	3600	2250	3600	1200	750	1200	1200	750	1200
4		2	6150	4500	6150	2050	1500	2050	2050	1500	2050
$E_θ$			3169	3413	4163	1056	1138	1388	1056	1138	1388

Max., maximum.

of the available evidence. To determine the benefit associated with resolving only θ₁ we find the expected NHEs of each intervention for each possible value of θ₁. In this case the benefit associated with θ₁ = 1 in population 1 for intervention 1,

$$E_{θ_2|θ_1=1}NHE(1,θ_1=1,θ_2).\sum_{t=1}^{6}\frac{I_t}{(1+d)^t},$$

is the average of 2475 and 450, which is equal to 1463 and the NHE associated with θ₁ = 1 in population 1 for intervention 2,

$$E_{θ_2|θ_1=1}NHE(2,θ_1=1,θ_2).\sum_{t=1}^{6}\frac{I_t}{(1+d)^t},$$

is the average of 2850 and 4050, which is equal to 3450. The full range of effects is calculated and shown in *Table 50* for the three patient populations. Knowing whether θ₁ would resolve at 1 or 2 would allow the decision-maker to obtain the maximum net benefits of

$$E_{θ_1}\max_j E_{θ_2|θ_1}NHE(j,θ_1,θ_2).Pop = 4163+1388+1388 = 6938.$$

The benefit of resolving uncertainty in θ₂ can be calculated in the same way and is shown in *Table 51*. Knowing whether or not θ₂ would resolve at 1 or 2 would allow the decision-maker to obtain the maximum NHEs of 3656 + 1219 + 1219 = 6094. In both tables the expected NHEs of approving the cost-effective intervention 2, based on current evidence, is unchanged at 3413 + 1138 + 1138 = 5688. However, the gain in expected NHEs from research on θ₁ is greater than the expected gain from further research on θ₂.

Previously, *Tables 44–46* considered the potential pay-offs from research that would resolve all uncertainty, θ, that is, both θ₁ and θ₂. Assuming that research designs to resolve θ₁ or θ₂

TABLE 50 Population NHEs conditional on θ_1

θ_1	Population 1			Population 2			Population 3		
	$NHE(1,\theta_1)$	$NHE(2,\theta_1)$	Max. NHE	$NHE(1,\theta_1)$	$NHE(2,\theta_1)$	Max. NHE	$NHE(1,\theta_1)$	$NHE(2,\theta_1)$	Max. NHE
1	1463	3450	3450	488	1150	1150	488	1150	1150
2	4875	3375	4875	1625	1125	1625	1625	1125	1625
E_θ	3169	3413	4163	1056	1138	1388	1056	1138	1388

Max., maximum.

TABLE 51 Population NHEs conditional on θ_2

θ_2	Population 1			Population 2			Population 3		
	$NHE(1,\theta_2)$	$NHE(2,\theta_2)$	Max. NHE	$NHE(1,\theta_2)$	$NHE(2,\theta_2)$	Max. NHE	$NHE(1,\theta_2)$	$NHE(2,\theta_2)$	Max. NHE
1	3038	2550	3038	1013	850	1013	1013	850	1013
2	3300	4275	4275	1100	1425	1425	1100	1425	1425
E_θ	3169	3413	3656	1056	1138	1219	1056	1138	1219

Max., maximum.

TABLE 52 Policy pay-offs when two types of uncertainty are considered

Policies	$\prod(Pop_1)$	$\prod(Pop_2)$	$\prod(Pop_3)$	$\prod$
1. Approve	3413	1138	1138	5688
2. AWR(θ_1)	3413	1388	1388	6188
3. AWR(θ_2)	3413	1219	1219	5850
4. AWR(θ)	3413	1388	1388	6188
5. OIR(θ)	3169	1388	1388	5944
6. OIR(θ_1)	3169	1388	1388	5944
7. OIR(θ_2)	3169	1219	1219	5606
8. Reject	3169	1056	1056	5281

individually would be available for population 2 ($\tau_1 = \tau_2 = 6$), and that research is certain to report ($\alpha = 1$), we consider an expanded range of potential adoption and research decisions in *Table 52*.

The policies and results for 1, 4, 5 and 8 are the same as those presented in *Table 45*. The NHEs for policies 2 and 6 are calculated by assuming that only θ_1 is resolved in populations 2 and 3 (see *Table 50*). The NHEs for policies 3 and 7 are calculated by assuming that only θ_2 is resolved in populations 2 and 3 (see *Table 51*). There are now a range of possible OIR and AWR decisions that differ in the type of additional evidence sought. For example, the policy to reject intervention 2 and conduct further research on θ_1 can be denoted OIR(θ_1). The policy to reject intervention 2 and conduct research on all (both) parameters is OIR(θ). Policies 2–4 represent AWR decisions whereas policies 5–7 represent OIR decisions.

The results reported in *Table 52* suggest that there is no additional benefit in resolving θ_2 over and above that realised if θ_1 is resolved. The most beneficial policy is to approve and then research θ_1, that is, an AWR(θ_1) decision.

The formal notation for the additional reimbursement and research decisions described in *Table 52* are:

1. Approve: see *Equation 6*
2. Approve and research θ_1:

$$\prod_{AWR(\theta_1)} = \max_j E_\theta NHE(j,\theta).\sum_{t=1}^{\tau_1} \frac{I_t}{(1+d)^t} + E_{\theta_1} \max_j E_{\theta_2} NHE(j,\theta).\sum_{t=\tau_1}^{T} \frac{I_t}{(1+d)^t} \qquad \text{[Equation 14]}$$

3. Approve and research θ_2:

$$\prod_{AWR(\theta_2)} = \max_j E_\theta NHE(j,\theta).\sum_{t=1}^{\tau_2} \frac{I_t}{(1+d)^t} + E_{\theta_2} \max_j E_{\theta_1} NHE(j,\theta).\sum_{t=\tau_2}^{T} \frac{I_t}{(1+d)^t} \qquad \text{[Equation 15]}$$

4. Approve and research θ (both θ_1 and θ_2): see *Equation 10*
5. Reject and research θ_1 and θ_2: see *Equation 11*
6. Reject and research θ_1:

$$\prod_{OIR(\theta_1)} = E_\theta NHE(1,\theta).\sum_{t=1}^{\tau_1} \frac{I_t}{(1+d)^t} + E_{\theta_1} \max_j E_{\theta_2} NHE(j,\theta).\sum_{t=\tau_1}^{T} \frac{I_t}{(1+d)^t} \qquad \text{[Equation 16]}$$

7. Reject and research θ_2:

$$\prod_{Approve} = \max_j E_\theta NHE(j,\theta).Pop \qquad \text{[Equation 17]}$$

8. Reject: see *Equation 7*.

What type of research is possible if the technology is approved?

The section *Is the research possible if the intervention is approved?* discussed the fact that some types of research may not be possible once a new intervention is approved for widespread use. In this example if research to inform θ_1 is not possible once the new intervention is approved, then policies 2 and 4 in *Table 52* are not options. For example, this type of trade-off may arise if a RCT is the appropriate research design for gathering further information on θ_1, whereas an observational study could inform θ_2. The decision-maker must choose between immediate approval, with further research possible only on θ_2, or rejection and the potential to conduct further research on either or both types of uncertainty. In these circumstances the expected pay-offs in health effects associated with policies 5 and 6 ($\prod_{OIR(\theta)}, \prod_{OIR(\theta_1)}$) are greater than those of policies 1 and 3 ($\prod_{Approve}, \prod_{AWR(\theta_2)}$), that is, it is more beneficial to utilise OIR and reject in order to allow research about θ_1 than to approve or utilise AWR restricted to research that would inform only θ_2. The expected health gains from conducting the more valuable research design exceed the opportunity cost of delaying access to the new intervention.

If the observational study design to inform θ_2 is viable only following approval, then policies 5 and 7 in *Table 52* are no longer options. The decision-maker may be faced with the choice between approving the new intervention in order to obtain information on θ_2 and rejecting in order to obtain information on θ_1. In this example, policy 6 ($\prod_{OIR(\theta_1)} = 5944$) is more valuable than policy 3 ($\prod_{AWR(\theta_2)} = 5850$). However, this balance could be altered if the research designs impose different costs on the budget for health care. For example, if research to inform θ_2 costs £200,000

$$\left(R_2 = \frac{£200,000}{k} = 10\right)$$

whereas research to inform θ_1 costs £4M ($R_1 = 200$) then the pay-off from AWR with further research on θ_2, $\Pi_{AWR(\theta_2)} - R_2 = 5850 - 10 = 5840$, would offer greater expected NHEs than policy 6, $\Pi_{OIR(\theta_1)} - R_1 = 5944 - 200 = 5744$.

How does price affect the type of research required?

Figure 41 shows the frontier of optimal policy choices as a function of the price of intervention 2 and with research costs of $R_1 = 200$, $R_2 = 10$ and $R_{1,2} = 210$. The frontier shows that the choice between approve, reject, AWR and OIR is sensitive to the price of intervention 2.

When the price of intervention 2 is less than £700, the benefits from additional research would not exceed the costs and approve is the optimal policy. As the price of intervention 2 rises to between £700 and £1081 the consequences of decision uncertainty increase to the point that the benefits of further research on θ_1 outweigh the costs and AWR(θ_1) is the optimal policy. For prices between £1081 and £1145 intervention 2 is no longer expected to be cost-effective, but further research to inform θ_1 still appears worthwhile so that OIR(θ_1) is the optimal policy. The type of research required remains the same for prices between £700 and £1145, but the initial decision to approve or reject prior to research reporting is altered. For prices between £1145 and £1780 intervention 2 is still not expected to be cost-effective, but the benefits of further research are altered such that the type of research required under OIR is now research into both θ_1 and θ_2. When the price of intervention 2 exceeds £1780 the consequences of uncertainty in the decision to reject intervention 2 are reduced to the point at which further research no longer appears worthwhile.

Does the sequence of research have implications for 'approval with research' and 'only in research'?

The policy set evaluated in *Table 52* describes the choice between no research or a single research study of various designs. However, the possible research decisions also include the different sequences in which different types of research could be conducted. The choice of sequence will depend on the benefits and costs of different types of research and the impact of resolving some uncertainties on the importance of others that remain unresolved. For example, quick and relatively cheap research about a key parameter (which influences the importance of others) might, in some circumstances, show that a very expensive trial which would take a considerable time to report might no longer be required. The benefits of a sequential research design are that the latter pieces of research will be conducted only if it still appears valuable conditional on the
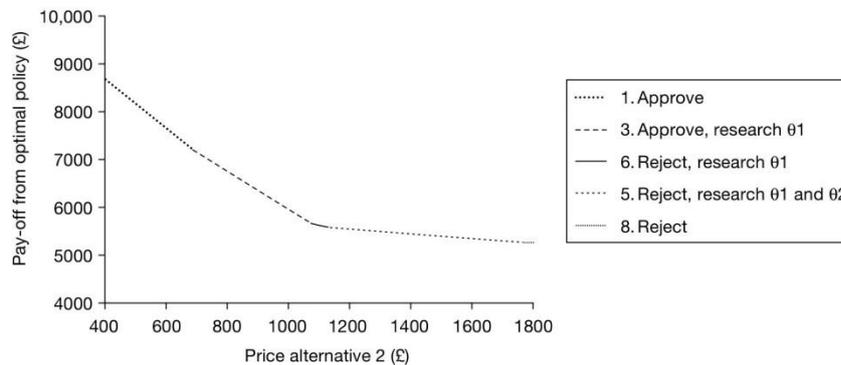

FIGURE 41 Optimal policy frontier.

results of the earlier pieces, thus allowing the decision-maker to avoid some research costs, R.[131] Note that the time until research reports and the likelihood of it reporting must be calculated for each potential research design. The choice of sequence will also be modified by whether particular types of research are possible with and without approval.

With two types of uncertainty there are two possible sequences for research that can be evaluated alongside one-off research decisions. Research initiated immediately would report at τ_1 for θ_1, or τ_2 if instead it was designed to inform θ_2. It is also possible to consider the probability of research separately for each possible research design (see *Is further research required?*). However, in this example we assume that research is guaranteed to report ($\alpha_1 = \alpha_2 = \alpha_{1,2} = 1$). If further research into θ_2 is initiated after research on θ_1 had reported at time τ_1, that research would report at time $\tau_1 + \tau_2$. So decisions earlier than τ_1 must be made before the results of any research are known. Decisions made between τ_1 and $\tau_1 + \tau_2$ can utilise the additional information generated for θ_1. Decisions made after $\tau_{1+}\tau_2$ can utilise the additional information generated for both θ_1 and θ_2.

If research on parameter i will cost R_i then after having conducted research into θ_1 the decision to proceed with research on θ_2 is informed by:

$$E_{\theta_1} \max \left\{ \max_j E_{\theta_2} NHE(j,\theta). \sum_{t=\tau_1+\tau_2}^{T} \frac{I_t}{(1+d)^t}, E_{\theta_2|\theta_1} \max_j NHE(j,\theta). \sum_{t=\tau_1+\tau_2}^{T} \frac{I_t}{(1+d)^t} - R_2 \right\}$$

[Equation 18]

That is, the benefits of research into θ_2 will be available at time $\tau_1 + \tau_2$, and these must be compared with the benefits of continuing with a decision based on current information available at time τ_1, that is, to maximise expected NHEs with perfect knowledge of θ_1.

After having conducted research into θ_2 the decision to proceed with research on θ_1 is informed by:

$$E_{\theta_2} \max \left\{ \max_j E_{\theta_1} NHE(j,\theta). \sum_{t=\tau_1+\tau_2}^{T} \frac{I_t}{(1+d)^t}, E_{\theta_1|\theta_2} \max_j NHE(j,\theta). \sum_{t=\tau_1+\tau_2}^{T} \frac{I_t}{(1+d)^t} - R_1 \right\}$$

[Equation 19]

Let us suppose that the costs of research are $R_1 = 200$ and $R_2 = 10$. *Table 52* indicates that there is no added value to further research on θ_2 following research into θ_1 (the pay-offs from policies 2 and 4 are the same, as are the pay-offs from policies 5 and 6). Therefore, the probability of proceeding to incur research costs for θ_2 in a sequential design to investigate first θ_1 then θ_2 is zero:

$$E_{\theta_1} \max_j E_{\theta_2} NHE(j,\theta). \sum_{t=\tau_1+\tau_2}^{T} \frac{I_t}{(1+d)^t} > \left[E_{\theta} \max_j NHE(j,\theta). \sum_{t=\tau_1+\tau_2}^{T} \frac{I_t}{(1+d)^t} - R_2 \right].$$

The pay-off from a sequential design to inform θ_1 then θ_2 is simply the same as the pay-off from a one-off research design for θ_1.

However, a comparison of the pay-offs from policies 3 and 4 or policies 5 and 7 suggests that there is additional value from resolving θ_1 after θ_2 is known. To calculate whether research

would proceed on θ_1 we need to return to *Table 49* to calculate the value of further research and compare this with the costs of research, $R_1 = 200$. We must establish whether further research to resolve θ_1 for population 3 would proceed once research had identified the true value of θ_2 for population 2. First, we consider the case in which research identifies that $\theta_2 = 1$. Notice from *Table 50* that if the value of θ_2 is known to be 1, the NHEs of intervention 1 in population 3 are the average of 825 and 1200,

$$E_{\theta_1} NHE(1, \theta_1, \theta_2 = 1) \sum_{t=8}^{T} \frac{I_t}{(1+d)^t} = 1013,$$

the NHEs of intervention 2 in population 3 are the average of 950 and 750,

$$E_{\theta_1} NHE(2, \theta_1, \theta_2 = 1) \sum_{t=8}^{T} \frac{I_t}{(1+d)^t} = 850,$$

and the maximum NHEs in population 3 are the average of 950 and 1200,

$$\max_j E_{\theta_1} NHE(j, \theta_1, \theta_2 = 1) \sum_{t=8}^{T} \frac{I_t}{(1+d)^t} = 1075.$$

The additional NHEs from research are not sufficient to outweigh the opportunity cost of that research,

$$\left[\max_j E_{\theta_1} NHE(j, \theta_1, \theta_2 = 1) - E_{\theta_1} NHE(1, \theta_1, \theta_2 = 1) \right] \sum_{t=8}^{T} \frac{I_t}{(1+d)^t} < R_1,$$

and the optimal policy is to reject. Now we make the same calculations for the case in which research identifies that $\theta_2 = 2$. The NHEs of intervention 1 in population 3 are the average of 150 and 2050,

$$E_{\theta_1} NHE(1, \theta_1, \theta_2 = 2) \sum_{t=8}^{T} \frac{I_t}{(1+d)^t} = 1100,$$

the NHEs of intervention 2 in population 3 are the average of 1350 and 1500,

$$E_{\theta_1} NHE(2, \theta_1, \theta_2 = 2) \sum_{t=8}^{T} \frac{I_t}{(1+d)^t} = 1425,$$

and the maximum NHEs in population 3 are the average of 1350 and 2050,

$$\max_j E_{\theta_1} NHE(j, \theta_1, \theta_2 = 2) \sum_{t=8}^{T} \frac{I_t}{(1+d)^t} = 1700.$$

The additional NHEs from research now outweigh the opportunity cost of that research,

$$\left[\max_j E_{\theta_1} NHE(j, \theta_1, \theta_2 = 1) - E_{\theta_1} NHE(1, \theta_1, \theta_2 = 1) \right] \sum_{t=8}^{T} \frac{I_t}{(1+d)^t} = 1700,$$

and the optimal policy is to continue to research θ_1. In summary, if $\theta_2 = 1$ research on θ_1 will not proceed, whereas if $\theta_2 = 2$ research on θ_1 will proceed. Each value of θ_2 is equally likely and so the probability of continuing to research on θ_1 is 0.5 (50%). This means that the expected research costs for the sequential research design ($R_2 + 0.5R_1$) are less than the expected research costs for a decision to conduct both research studies ($R_2 + R_1$). *Table 53* shows the full set of expected research costs alongside the policy pay-offs from the full range of policies that incorporate AWR and OIR with sequential research designs.

The general notation for calculating the pay-offs from the additional policies described in *Table 53* is:

TABLE 53 Policy pay-offs with different sequences of research and expected research costs

Policies	$\prod(Pop_1)$	$\prod(Pop_2)$	$\prod(Pop_3)$	R	$\prod - R$
1. Approve	3413	1138	1138	0	5688
2. AWR(θ_1)	3413	1388	1388	200	5988
3. AWR(θ_2)	3413	1219	1219	10	5840
4. AWR(θ_1 then θ_2)	3413	1388	1388	200	5988
5. AWR(θ_2 then θ_1)	3413	1219	1378	110	5899
6. AWR(θ)	3413	1388	1388	210	5978
7. OIR(θ)	3169	1388	1388	210	5734
8. OIR(θ_1 then θ_2)	3169	1388	1388	200	5744
9. OIR(θ_2 then θ_1)	3169	1219	1378	110	5655
10. OIR(θ_1)	3169	1388	1388	200	5744
11. OIR(θ_2)	3169	1219	1219	10	5596
12. Reject	3169	1056	1056	0	5281

1. Approve: see *Equation 6*
2. Approve and research θ_1: see *Equation 14*
3. Approve and research θ_2: see *Equation 15*
4. Approve and research θ_1 then θ_2: utilise *Equations 14* and *18*:

$$\prod_{AWR(\theta_1 \text{ then } \theta_2)} = \max_j E_\theta NHE(j,\theta).\sum_{t=1}^{\tau_1}\frac{I_t}{(1+d)^t} + E_{\theta_1}\max_j E_{\theta_2} NHE(j,\theta).\sum_{t=\tau_1}^{\tau_2}\frac{I_t}{(1+d)^t} - R_1 +$$
$$E_{\theta_1}\max\left\{\max_j E_{\theta_2} NHE(j,\theta).\sum_{t=\tau_2}^{T}\frac{I_t}{(1+d)^t}, E_{\theta_2|\theta_1}\max_j NHE(j,\theta).\sum_{t=\tau_2}^{T}\frac{I_t}{(1+d)^t} - R_2\right\}$$

[Equation 20]

5. Approve and research θ_2 then θ_1: utilise *Equations 15* and *19*:

$$\prod_{AWR(\theta_1 \text{ then } \theta_2)} = \max_j E_\theta NHE(j,\theta).\sum_{t=1}^{\tau_1}\frac{I_t}{(1+d)^t} + E_{\theta_2}\max_j E_{\theta_1} NHE(j,\theta).\sum_{t=\tau_1}^{\tau_2}\frac{I_t}{(1+d)^t} - R_2 +$$
$$E_{\theta_2}\max\left\{\max_j E_{\theta_1} NHE(j,\theta).\sum_{t=\tau_2}^{T}\frac{I_t}{(1+d)^t}, E_{\theta_1|\theta_2}\max_j NHE(j,\theta).\sum_{t=\tau_2}^{T}\frac{I_t}{(1+d)^t} - R_1\right\}$$

[Equation 21]

6. Approve and research θ (both θ_1 and θ_2): see *Equation 10*
7. Reject and research θ_1: see *Equation 16*
8. Reject and research θ_2: see *Equation 17*
9. Reject and research θ_1 then θ_2: utilise *Equations 16* and *18*:

$$\prod_{OIR(\theta_1 \text{ then } \theta_2)} = E_\theta NHE(1,\theta).\sum_{t=1}^{\tau_1}\frac{I_t}{(1+d)^t} + E_{\theta_1}\max_j E_{\theta_2} NHE(j,\theta).\sum_{t=\tau_1}^{\tau_2}\frac{I_t}{(1+d)^t} - R_1 +$$
$$E_{\theta_1}\max\left\{\max_j E_{\theta_2} NHE(j,\theta).\sum_{t=\tau_2}^{T}\frac{I_t}{(1+d)^t}, E_{\theta_2|\theta_1}\max_j NHE(j,\theta).\sum_{t=\tau_2}^{T}\frac{I_t}{(1+d)^t} - R_2\right\}$$ [Equation 22]

10. Reject and research θ_2 then θ_1: utilise *Equations 17* and *19*:

$$\Pi_{OIR(\theta_2 \text{ then } \theta_1)} = E_\theta NHE(1,\theta).\sum_{t=1}^{\tau_1}\frac{I_t}{(1+d)^t} + E_{\theta_2}\max_j E_{\theta_1} NHE(j,\theta).\sum_{t=\tau_1}^{\tau_2}\frac{I_t}{(1+d)^t} - R_2 +$$

$$E_{\theta_2}\max\left\{\max_j E_{\theta_1} NHE(j,\theta).\sum_{t=\tau_2}^{T}\frac{I_t}{(1+d)^t}, E_{\theta_1|\theta_2}\max_j NHE(j,\theta).\sum_{t=\tau_2}^{T}\frac{I_t}{(1+d)^t} - R_1\right\} \qquad \text{[Equation 23]}$$

11.   Reject and research θ (both θ_1 and θ_2): see *Equation 11*
12.   Reject: see *Equation 7*.

The consideration of research costs must be incorporated in order to determine the pay-offs from the sequential research designs. If manufacturers bear the research costs such that they no longer fall on the budget for health care, further knowledge on θ_1 would be regarded as valuable whether $\theta_2 = 1$ or 2. However, when research costs do fall on the budget for health care then if $\theta_2 = 1$ research on θ_1 would not proceed and the decision-maker would not be able to obtain the maximum possible NHEs for population 3, $E_\theta \max NHE(j,\theta)$, only the maximum conditional on θ_2, $E_{\theta_2}\max E_{\theta_1} NHE(j,\theta)$. In this example the smaller expected research costs from a sequential design to research first θ_2 then θ_1 (AWR policy 5 or OIR policy 10) are not sufficient to outweigh the benefit of earlier research into θ_1 (AWR policies 2, 4 and 6 or OIR policies 7, 8 and 10).

Recall from earlier that some types of research might not be possible once a technology is approved for widespread use (see *Is the research possible if the intervention is approved?*). If it is not possible to conduct research into θ_1 following approval then policies 2, 4, 5 and 6 in *Table 53* are no longer available. In this example, the sequential OIR designs that would provide additional information on θ_1 do not alter the conclusion that AWR(θ_2) offers the greatest expected pay-off.

Impact of other sources of uncertainty

The section *Is further research required?* explained why there may be value in resolving uncertainty through further research, and the section *What type of evidence is needed?* discussed the range of assessments that would have to be made about the appropriate research design. However, some uncertain elements that determine the expected costs and health outcomes of technologies will not be resolvable through further research. If it is possible that further information may be revealed over time on these other sources of uncertainty then they too can have an impact on the appropriate policy choice and so this possibility must be assessed. One example of a source of uncertainty that is not resolvable through research is the price of a health-care technology following patent expiry. The decision-maker may have good knowledge of the time at which the patent is due to expire but be uncertain of the impact that this and the potential for generic entry will have on the price. In this section we consider the impact of such an event on the appropriate policy choice.

For example, the patent for intervention 2 is known to expire at time $t = 6$ and at this point one of two possible outcomes will occur. Either there will be immediate generic entry to the market (a state of the world represented by $\gamma = 1$) or the manufacturer might successfully protect its exclusive market position (a state of the world represented by $\gamma = 2$). In this simple example both states of the world are deemed to be equally likely. As before, if the decision-maker must select an intervention before further information about generic entry is known, they should choose the technology with the greatest expected NHEs: $\max_j E_\gamma NHE(j,\gamma)$.

After the patent has expired and the decision-maker knows whether or not a lower-priced version of the drug will be available the intervention can be chosen conditional on the outcome

of the event max $NHE(j,\gamma)$, and the expected value of this decision is $E_\gamma \max_j NHE(j,\gamma)$. *Table 54* shows the expected NHEs for each intervention depending on the resolution of the uncertain parameter, γ. Note that when cost-effectiveness is assessed as described earlier (see *Is the intervention cost-effective and what are the risks?*) it is assessed on the basis of either the monopoly price or the generic price. It would be inappropriate to incorporate the expected price (i.e. the average of these two prices) in the assessment of cost-effectiveness, as basing decisions on this expected price would in general offer fewer NHEs than basing decisions on the known current price within each decision period. The example so far has assumed that the price of intervention 2 remains unchanged at the monopoly price and so the numbers equate to those for $\gamma = 2$.

Table 55 presents the pay-offs from possible policy options. As the uncertainty about whether or not generic entry will occur cannot be resolved through research, the decision to initiate research is no longer relevant. However, as the uncertainty is expected to resolve over time the reimbursement decision could be re-evaluated and reversed.

The general notation for the pay-offs included in *Table 55* are as follows:

1. Approve with no change of guidance:

$$\prod\nolimits_{Approve} = \max_j E_\gamma NHE(j,\gamma).Pop \qquad \text{[Equation 24]}$$

2. Reversal (approve then revise guidance after patent expiry):

$$\prod\nolimits_{Reversal} = E_\gamma NHE(2,\gamma).\sum_{t=1}^{\tau}\frac{I_t}{(1+d)^t} + E_\gamma \max_j NHE(j,\gamma).\sum_{t=\tau}^{T}\frac{I_t}{(1+d)^t} \qquad \text{[Equation 25]}$$

3. Delay (reject then revise guidance after patent expiry):

$$\prod\nolimits_{Delay} = E_\gamma NHE(1,\gamma).\sum_{t=1}^{\tau}\frac{I_t}{(1+d)^t} + E_\gamma \max_j NHE(j,\gamma).\sum_{t=\tau}^{T}\frac{I_t}{(1+d)^t} \qquad \text{[Equation 26]}$$

TABLE 54 Population NHEs with other sources of uncertainty

θ	Population 1			Population 2			Population 3		
	NHE(1,θ)	NHE(2,θ)	Max. NHE	NHE(1,θ)	NHE(2,θ)	Max. NHE	NHE(1,θ)	NHE(2,θ)	Max. NHE
1	3169	3413	3413	1056	1738	1738	1056	1738	1738
2	3169	3413	3413	1056	1138	1138	1056	1138	1138
E_γ	3169	3413	3413	1056	1438	1438	1056	1438	1438

Max., maximum.

TABLE 55 Policy pay-offs with other sources of uncertainty

Policies	$\prod(Pop_1)$	$\prod(Pop_2)$	$\prod(Pop_3)$	$\prod$
1. Approve	3413	1438	1438	6288
2. Reversal (approve then revise)	3413	1438	1438	6288
3. Delay (reject then revise)	3169	1438	1438	6044
4. Reject	3169	1056	1056	5281

4. Reject with no change of guidance:

$$\textstyle\prod_{Reject} = E_\gamma NHE(1,\gamma).Pop \qquad \text{[Equation 27]}$$

When a new intervention is initially regarded as cost-effective, the potential for a price reduction will not change that assessment: intervention 2 goes from being cost-effective to being even more cost-effective. The NHEs of approve in *Table 55* exceed the value in *Table 40* whereas the NHEs of reject are unaffected. It is for this reason that the option to initially approve the new intervention and then reverse the decision after the price change is not valuable. Likewise, the option to delay approval until after the price change is known generates fewer expected net health benefits than approve. However, the option to revise the decision may be relevant if the resolution of the other source of uncertainty would make intervention 2 appear less, rather than more, cost-effective.

How does price affect the impact of other sources of uncertainty?

If the monopoly price is higher (£1500) as shown in *Table 56* and intervention 2 is not cost-effective then a policy to reject the new intervention appears more valuable than a policy to approve. However, the potential change in price means that once the uncertainty is resolved approval may appear cost-effective, and so policy 3 to initially reject and then approve if the price falls after patent expiry offers the greatest pay-off: $\prod_{Delay} = 3169 + 1397 + 1397 = 5963$.

Some might regard this as a policy to delay the decision whereas it is in fact reject followed by a reassessment of the value of the new intervention should the price drop. This policy to delay approval (policy 3) offers $5963 - 5281 = 682$ additional expected NHEs compared with the next best policy of reject. In this example there is a 50% chance that guidance will change at $t = 6$, that is, before the end of the expected lifetime of the intervention, $T_{Pop} = 10$.

Figure 42 shows how the optimal policy changes with the monopoly price of intervention 2 (the plotted values correspond to those in *Table 55* with a 50% probability of a generic price of £400 after one period).

For prices below £1081 the optimal policy is to approve, and the opportunity to revise the decision (Reversal) does not offer additional value because intervention 2 is cost-effective regardless of whether or not there is generic entry. Under these circumstances the resolution of the other source of uncertainty would not be expected to lead to a change in guidance. Once the price exceeds £1081 intervention 2 is not cost-effective unless the price falls with generic entry, and so the policy of reject then revise (delay) offers the greatest pay-off. In this price range the resolution of the other source of uncertainty could potentially lead to a change of guidance. Thus, the impact of a price change can be to alter the probability that guidance will be changed in the future in response to the resolution of other sources of uncertainty over time.

TABLE 56 Policy pay-offs with other sources of uncertainty: monopoly price intervention 2 = £1500

Policies	$\prod(Pop_1)$	$\prod(Pop_2)$	$\prod(Pop_3)$	$\prod$
1. Approve	1913	1188	1188	4288
2. Reversal (approve then revise)	1913	1397	1397	4706
3. Delay (reject then revise)	3169	1397	1397	5963
4. Reject	3169	1056	1056	5281

How does the impact of other sources of uncertainty vary with their magnitude?

So far the probability of generic entry has been assumed to be 50%. If the probability of generic entry is lower, for example 10% as in *Table 57*, the additional expected health effects from delaying the decision will be reduced.

However, notice that the pay-offs from all policies in which the intervention might be approved have fallen because there is less chance of a large improvement in expected NHEs from intervention 2 when generic entry is less likely. The impact of altering the level of uncertainty about the possibility of generic entry is greatest for approve (no revision), for which the decision cannot be altered in the light of additional information.

How do other sources of uncertainty impact on the need for evidence?

When regarded in isolation, a source of uncertainty that will resolve over time illustrates that decision-makers may have an opportunity to maximise health benefits by revising their reimbursement decision in light of additional information. However, other sources of uncertainty can also impact on decisions about whether further research is valuable. Thus, when such a source of uncertainty is identified, it is necessary to reassess the value of further research in the presence of these uncertain events that may be resolved over time. That is, the assessments described in *Is further research required?* and *What type of evidence is needed?* must be combined with those described in *Impact of other sources of uncertainty*. To illustrate how this can be achieved, *Table 58* combines *Tables 49* and *54* to show all possible value combinations for θ conditional on different values for γ.
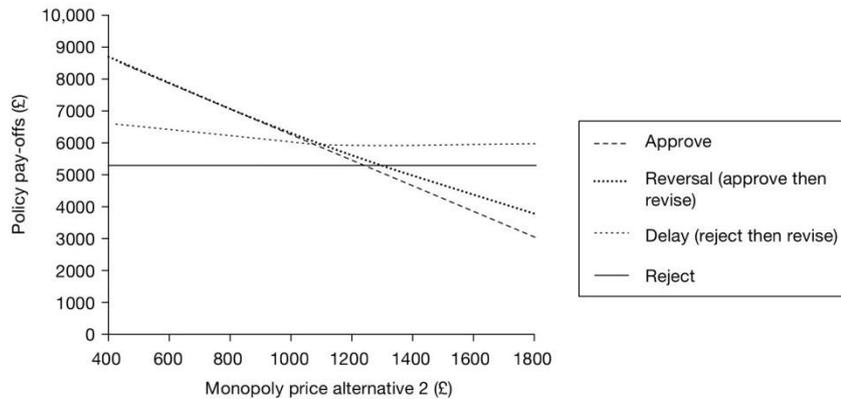
FIGURE 42 Impact of monopoly price intervention 2 on optimal policy decision.

TABLE 57 Policy pay-offs with other sources of uncertainty: monopoly price intervention 2 = £1500, probability generic entry = 0.1

Policies	$\prod(Pop_1)$	$\prod(Pop_2)$	$\prod(Pop_3)$	$\prod$
1. Approve	1913	748	748	3408
2. Reversal (approve then revise)	1913	1124	1124	4161
3. Delay (reject then revise)	3169	1124	1124	5418
4. Reject	3169	1056	1056	5281

TABLE 58 Population NHEs in the presence of uncertainty that can be resolved through research and uncertainty that is resolvable over time

θ	θ	θ_1	θ_2	Population 1 NHE(1,θ,θ)	NHE(2,θ,θ)	Max. NHE	Population 2 NHE(1,θ,θ)	NHE(2,θ,θ)	Max. NHE	Population 3 NHE(1,θ,θ)	NHE(2,θ,θ)	Max. NHE
1	1	1	1	2475	2850	2850	825	950	950	825	950	950
	2		2	450	4050	4050	150	1350	1350	150	1350	1350
	3	2	1	3600	2250	3600	1200	750	1200	1200	750	1200
	4		2	6150	4500	6150	2050	1500	2050	2050	1500	2050
2	1	1	1	2475	2850	2850	825	1550	1550	825	1550	1550
	2		2	450	4050	4050	150	1950	1950	150	1950	1950
	3	2	1	3600	2250	3600	1200	1350	1350	1200	1350	1350
	4		2	6150	4500	6150	2050	2100	2100	2050	2100	2100
$E_{\theta\gamma}$				3169	3413	4163	1056	1438	1563	1056	1438	1563

Max., maximum.

Table 59 repeats the policy options from *Table 53* in light of the other source of uncertainty regarding the potential price reduction. However, in this example we assume that the costs of research do not fall on the budget for health care and so $R_1 = R_2 = R_{1,2} = 0$.

When a new intervention is regarded as cost-effective, the potential for a price reduction can increase the benefits of adoption relative to the benefits of further research. Once other sources of uncertainty are resolved, the probability of a wrong decision (see *Is further research required?*) is altered and hence the value of research will change. For example, in *Table 49* additional information about θ_1, θ_2 would lead to a different reimbursement decision for two out of the four possible value combinations, giving an error probability of 50%. However, when also considering the possibility of a price reduction, there are eight possible value combinations shown in *Table 58*, and once that other source of uncertainty is resolved (for populations 2 and 3) additional information about θ_1, θ_2 would lead to a different reimbursement decision for only two of these, leading to an error probability of 25%. However, a reduction in the probability of a wrong decision does not necessarily imply that the value of further evidence will fall, as the resolution of other sources of uncertainty can also impact on the consequences of a wrong decision.

Whereas in *Table 53* policy 10 [OIR(θ_1)] was regarded as more valuable than approve without evidence collection (policy 1), the prospect of a price reduction means that the ranking of these policies is reversed in *Table 59*. The additional information about θ_1 generated by policy 10 would be available only after the price reduction was known. At this point additional information about θ_1 is less valuable relative to approval and the gain in expected NHEs is insufficient to outweigh the opportunity cost of denying earlier access to intervention 2 in order to conduct the research. Hence, the resolution of this other source of uncertainty makes an OIR decision to collect further information on θ_1 no longer valuable.

Other sources of uncertainty can also have the effect of increasing the value of certain types of research. For example, in *Table 59* there was no additional value to researching θ_2 once research on θ_1 is completed. However, given a reduction in price ($\gamma = 1$) further information about θ_2 would lead to different adoption decisions for population 3 in *Table 59*, and hence the value of policies 4 and 8 is increased such that $\prod_{AWR(\theta_1 \text{ then } \theta_2)} > \prod_{AWR(\theta_1)}$ and $\prod_{OIR(\theta_1 \text{ then } \theta_2)} > \prod_{OIR(\theta_1)}$.

TABLE 59 Policy pay-offs with uncertainty that can be resolved through research and uncertainty that is resolved over time

Policies	$\prod(Pop_1)$	$\prod(Pop_2)$	$\prod(Pop_3)$	$\prod$
1. Approve	3413	1438	1438	6288
2. AWR(θ_1)	3413	1538	1538	6488
3. AWR(θ_2)	3413	1438	1438	6288
4. AWR(θ_1 then θ_2)	3413	1538	1563	6513
5. AWR(θ_2 then θ_1)	3413	1438	1563	6413
6. AWR(θ)	3413	1563	1563	6538
7. OIR(θ)	3169	1563	1563	6294
8. OIR(θ_1 then θ_2)	3169	1538	1563	6269
9. OIR(θ_2 then θ_1)	3169	1438	1563	6169
10. OIR(θ_1)	3169	1538	1538	6244
11. OIR(θ_2)	3169	1438	1438	6044
12. Reject	3169	1056	1056	5281

So although the prospect of a future price reduction could make early adoption appear more valuable, it could also increase the value of further research. When further research is not possible after the intervention is approved (see *Is the research possible if the intervention is approved?* and *What type of research is possible if the intervention is approved?*) it is possible that on balance this other source of uncertainty could make OIR policies more attractive relative to approve even though the intervention is expected to be more cost-effective. So, for example, if research is not possible on θ_1 following approval, then policies 2, 4, 5 and 6 in *Table 59* are no longer available. Now it is preferable to use OIR(θ_1, θ_2), that is, to reject in order to conduct further research on both θ_1 and θ_2 rather than to utilise AWR to approve and conduct research only on θ_2.

What is the impact of irrecoverable costs?

Irrecoverable costs were introduced in the section *Are there significant irrecoverable costs?* in which it was noted that the expenditure of investment costs occurs in an earlier time period to that in which the ensuing benefits are realised. Investment costs may be incurred when the decision is made to commence treatment for each individual patient. Alternatively they may be at the level of the health-care system when the decision is made to approve an intervention for widespread use. Reversal costs are incurred when the decision is made to discontinue using a particular health-care technology. Similar to investment costs they can be at the individual patient level or at the level of the health-care system. For example, the decision to cease a pharmacological treatment for a chronic condition may require additional health-care contacts to monitor withdrawal at an individual level. At the health-care system level the decision to disinvest in a technology, that is, to no longer reimburse it, may need to be disseminated in order to alter clinical practice. Investment or reversal costs may be in terms of health effects rather than expenditure on capital equipment.[6,7]

When investment costs, *C*, such as expenditure on capital equipment, are incorporated by annuitisation (see *Are there significant irrecoverable costs?*), they are weighed against the health outcomes in the same way as any other costs, that is, the total investment cost is allocated over time and the patient population as if it was actually incurred (in small parts) per patient in each period. The total expected costs of approving the intervention will be correctly estimated as long

as this decision is not revised earlier than expected because of other events that may change the expected costs and health outcomes of this or comparator technologies. Therefore, it should be apparent that investment and reversal costs will have a particular impact (other than on expected costs) only if research is likely to report or other sources of uncertainty resolve in the near or distant future.

Investment costs are incurred at the point of approving a technology for use within the health service, and their introduction reduces the benefits of approve relative to reject. If investment costs are high enough, this may lead us to reject an intervention that was regarded as cost-effective before considering those costs. If there is an additional irrecoverable cost of £20M associated with intervention 2, this would be expected to displace activities that could have produced 1000 health benefits

$$(C(2) = \frac{£20,000,000}{k} = 1000).$$

Thus, the NHEs from approving the new intervention in *Table 41* need to be reduced by this amount, from 5688 to 4688 (assuming that they have not been incorporated by annuitisation in the estimation of individual NHEs). This would mean that the benefits of approve no longer exceed the benefits of reject (5281) and consideration of investment costs means that the new alternative no longer appears cost-effective. Of course, additional costs that were previously not included in an analysis will increase expected costs. However, the question is, in what circumstances will the irrecoverable nature of these costs impact on the approval and research decisions?

The impact of irrecoverable costs with other sources of uncertainty

In the section *Impact of other sources of uncertainty* we discussed the impact of resolving uncertain events over time on the expected costs and health effects of alternative interventions, and the possibility that this may lead to changes in the reimbursement decision over time. This means that the health outcomes from approving a technology may accrue over a shorter time period and thus will be reduced in magnitude relative to the irrecoverable costs.

Reversal costs are incurred at the point of removing an existing intervention from use within the health service (e.g. discontinuing the intervention that is current practice). These are often omitted from cost-effectiveness analysis, which simply compares the marginal benefit of one intervention over another. However, once we consider that the reimbursement decision may change over time, these should be incorporated in the same way as investment costs before the point at which the decision to approve the intervention is made. The difference between investment and reversal costs relates to the time at which they are incurred, but also to the likelihood of their occurring. Although the decision to approve a technology will necessarily lead to the imposition of investment costs (probability = 1), it is only if that decision is later changed that reversal costs are incurred (probability ≤ 1).

Table 55 showed the benefit of alternative policy decisions in the presence of uncertainty surrounding the future price of intervention 2. With this price change, the most beneficial policy was approve with NHEs of 6288. *Table 60* shows the value of the same policy options from *Table 55* but with an irrecoverable cost of $C(2) = 1000$ attached to the decision to approve intervention 2.

It can be seen that overall the health effects across all populations from approving the new intervention outweigh the investment costs (approve is more valuable than reject). However, the policy to reject the intervention until the outcome of the price change is known (delay) is now

TABLE 60 Policy pay-offs with investment costs and other sources of uncertainty

Policies	$\prod(Pop_1)$	$\prod(Pop_2)$	$\prod(Pop_3)$	$E_{C(2)}$	$\prod$
1. Approve	3413	1438	1438	1000	5288
2. Reversal (approve then revise)	3413	1438	1438	1000	5288
3. Delay (reject then revise)	3169	1397	1397	500	5463
4. Reject	3169	1056	1056	0	5281

increased in value relative to approve. This is because if we reject initially, the health effects from approving the new intervention in populations 2 and 3 in the absence of a price change ($\gamma = 2$) are not enough to exceed the irrecoverable costs,

$$NHE(2, \gamma = 2) \sum_{t=6}^{T} \frac{I_t}{(1+d)^t} - C(2) < NHE(1, \gamma = 2) \sum_{t=6}^{T} \frac{I_t}{(1+d)^t}$$

($1138 + 1138 - 1000 < 1056 + 1056$), and so if this state of the world arises we still reject. However, if generic entry does occur ($\gamma = 1$), the increased health gains from approving the new technology are enough to exceed the irrecoverable costs,

$$NHE(2, \gamma = 1) \sum_{t=6}^{T} \frac{I_t}{(1+d)^t} - C(2) < NHE(1, \gamma = 1) \sum_{t=6}^{T} \frac{I_t}{(1+d)^t}$$

($1738 + 1738 - 1000 > 1056 + 1056$), and so we would choose to approve at this point. The decision to approve immediately is optimal if the price drops, and delaying approval allows the decision-maker to incur the irrecoverable costs only if this is actually what happens ($p = 0.5$), rather than incurring the irrecoverable costs regardless of how γ is resolved ($p = 1$). Therefore, the expected irrecoverable costs associated with the policy to reject then revise (delay), $E_{C(2)} = 500$, are lower than the expected irrecoverable costs associated with approve, where $E_{C(2)} = 1000$.

The pay-off from the optimal policy in the presence of irrecoverable costs can be found using:

$$\max_j \left[E_\gamma NHE(j, \gamma) . \sum_{t=1}^{\tau} \frac{I_t}{(1+d)^t} - C(j) \right] + E_\gamma \max_j \left[NHE(j, \gamma) . \sum_{t=\tau}^{T} \frac{I_t}{(1+d)^t} - C(j, j_{t<\tau}) \right]$$

[Equation 28]

where $C(j)$ represents investment costs; these fall to zero in the second time period ($t \geq \tau$) if intervention j was selected in the first time period ($t \leq \tau$). For $C(j)$ that are reversal costs these become positive in the second time period ($t \geq \tau$) only if intervention j is selected in the first time period ($t \leq \tau$) then rejected in the second.

The impact of irrecoverable costs when further research is conducted

In the section *Is further research required?* we have discussed how to assess if further evidence is valuable for reducing uncertainty and shown that further research is valuable when the results of that research could indicate a change in the reimbursement decision. This implies again (see *Impact of other sources of uncertainty*) that the health effects from approving an intervention may be accrued over a shorter time period, and will be reduced in value relative to the irrecoverable costs. The impact of irrecoverable costs must be incorporated at the point at which an intervention is approved, which may be after research reports. When the irrecoverable costs are incurred after research reports, the remaining intervention lifetime is shorter and so health effects from approval will accrue to a smaller population compared with immediate approval. *Table 61* shows the policy options previously shown in *Table 53* but now with irrecoverable costs included and when the research costs do not fall on the budget for health care.

TABLE 61 Policy pay-offs with irrecoverable costs and further research

Policies	$\prod(Pop_1)$	$\prod(Pop_2)$	$\prod(Pop_3)$	$E_{C(2)}$	$\prod$
1. Approve	3413	1138	1138	1000	4688
2. AWR(θ_1)	3413	1388	1388	1000	5188
3. AWR(θ_2)	3413	1219	1219	1000	4850
4. AWR(θ_1 then θ_2)	3413	1388	1388	1000	5188
5. AWR(θ_2 then θ_1)	3413	1219	1388	1000	5019
6. AWR(θ)	3413	1388	1388	1000	5188
7. OIR(θ)	3169	1356	1356	250	5631
8. OIR(θ_1 then θ_2)	3169	1388	1388	500	5444
9. OIR(θ_2 then θ_1)	3169	1056	1356	250	5331
10. OIR(θ_1)	3169	1388	1388	500	5444
11. OIR(θ_2)	3169	1056	1056	0	5281
12. Reject	3169	1056	1056	0	5281

The expected irrecoverable cost, $E_{C(2)}$, is determined by the probability that intervention 2 will be approved. In policies 1–6 intervention 2 is approved immediately and so irrecoverable costs are incurred with a probability equal to one. In policy 12 intervention 2 is never approved and so irrecoverable costs are never incurred.

In policy 7, OIR(θ), intervention 1 is provided to population 1 and the results of research on both parameters are available in population 2. Investing in intervention 2 at this point would allow the health effects to accrue to patients in populations 2 and 3. Comparing these benefits previously reported in *Table 49* with the irrecoverable costs indicates that for three of the four possible combinations of θ_1 and θ_2 the decision-maker would continue to reject intervention 2. It is only if $\theta_1 = 1$ and $\theta_2 = 2$ that the benefits of intervention 2 in populations 2 and 3 outweigh the irrecoverable costs and the decision-maker would approve. This outcome has a probability of 0.25 and so the expected irrecoverable costs of policy 7 are 250. However, note that the pay-offs from policy 7 differ from those from the same policy in *Table 52* because the probability of approval given perfect knowledge of θ is reduced from 0.5 to 0.25 once additional irrecoverable costs are introduced.

In policy 10, OIR(θ_1), intervention 1 is provided to population 1 and the results of research to inform θ_1 are available in population 2. From *Table 50* it would appear optimal at this point to change the decision to approve if $\theta_1 = 1$ but not if $\theta_1 = 2$; hence, the expected irrecoverable cost for OIR(θ_1) is $E_{C(2)} = 500$.

In policy 11, OIR(θ_2), intervention 1 is provided to population 1 and the results of research to inform θ_2 are available in population 2. The values reported in *Table 51* illustrate that, regardless of how θ_2 resolves, the decision-maker would not choose to approve intervention 2. This is because the additional NHEs from approving intervention 2 if $\theta_2 = 2$ are reduced by the irrecoverable costs $(1425 + 1425 - 1000 = 1850)$ and are less than the NHEs from continuing to provide intervention 1 $(1100 + 1100 = 2200)$. The expected irrecoverable costs of policy 11 are therefore equal to zero, the additional value generated by the research on θ_2 is zero and the pay-off is equivalent to that of policy 12 $(\prod_{Reject} = E_\gamma NHE(1,\gamma).Pop)$. The impact of the irrecoverable costs is to reduce the value of research that previously would have appeared worthwhile.

In policy 8, OIR(θ_1 then θ_2), the results of research to inform θ_1 are available in population 2. At this point it would appear optimal to change the decision to approve intervention 2 if $\theta_1 = 1$ but

not if $\theta_1 = 2$ (see *Table 50*). The results of research to inform θ_2 are available in population 3. If θ_1 is already known to be 1, the irrecoverable costs will have been sunk and the decision-maker must simply choose whether to continue to reimburse intervention 2 or opt for intervention 1. In this example, the decision-maker would choose to continue to reimburse intervention 2 if $\theta_1 = 1$ no matter how θ_2 resolves (see *Table 49*). However, if θ_1 is already known to be 2, the investment will not have yet been made. In this example, if $\theta_1 = 2$ the decision-maker would continue to reject intervention 2 no matter how θ_2 resolves. The probability of incurring the irrecoverable costs is therefore given by the probability that $\theta_1 = 1$, that is, 0.5.

Summary

Each section of this technical appendix describes how the pay-off from alternative policy decisions can be quantified in the presence of uncertainty that can be resolved through research, uncertainty that will resolve over time and irrecoverable costs. The impact of various combinations of these factors has been demonstrated in a simple framework. The general notation should allow readers to combine relevant sections of the report in order to make the assessments required to inform a range of reimbursement and research decisions.

Appendix 7

Enhanced external counterpulsation for chronic stable angina

Introduction

The NIHR HTA programme identified EECP as an important topic and commissioned a short report to examine the clinical effectiveness and cost-effectiveness of EECP as an adjunct to standard therapy in patients with chronic stable angina. Although the topic was not ultimately considered by NICE it was commissioned in the same way and with the same resources as other assessment reports that inform NICE guidance. The assessment followed the NICE reference case[1] and is consistent with the type of analysis that would have been required in a MTA appraisal. Like other MTA appraisal reports it was published in full as a HTA monograph.[121]

The case study of EECP is used to demonstrate how the key principles and assessments in *Chapter 5* could inform the development of guidance for the use of EECP. The following sections include a background to EECP and a detailed examination of the sequence of assessments and decisions that lead to a particular category and type of guidance for EECP.

Background to the case study

Angina is a condition most commonly caused by coronary artery atherosclerosis with flow-limiting plaques that impede blood flow to the myocardium. Its prevalence in the UK is estimated to be approximately 700,000 men aged between 55 and 75 years and 400,000 women, with about 52,000 new cases per year in men and 43,000 in women.[159] The use of EECP as a treatment for angina is increasing steadily worldwide following reports of sustained benefit.[160–163] EECP has been utilised mainly in patients not suitable for coronary revascularisation or in those who have chosen not to undergo revascularisation.[164,165] It is estimated that about 10% of angina patients could potentially benefit from EECP (Michael Chester, Liverpool Hope University, UK, 2008, personal communication), which implies a prevalence of about 110,000 and an annual incidence of 9500.

Enhanced external counterpulsation results in upfront costs of treatment but the potential quality-of-life benefits through improved symptoms of angina and long-term relief from symptoms may outweigh the costs compared with not giving the therapy. Although EECP was not ultimately considered by NICE, the NIHR HTA programme identified EECP as an important topic for research and commissioned it in the same way as other assessment reports.

Intervention and population

Enhanced external counterpulsation is a non-invasive procedure used to provide symptomatic relief from stable angina.[160,162,163] Long inflatable pressure cuffs are wrapped around a patient's calves, lower thighs and upper thighs. The cuffs are inflated and deflated to increase blood flow to the coronary arteries and decrease peripheral vascular resistance and cardiac workload.[165] It typically involves 35 1-hour treatment sessions over a period of 4–7 weeks.[166] EECP as an adjunct to standard therapy is compared with standard therapy alone.

The population of interest is patients with stable angina. Because stable angina is a chronic condition, the prevalent population while any research is being conducted can benefit from the information and switch treatment if they survive and remain eligible when the results of the research reports.

Evidence on clinical effectiveness

A review of clinical effectiveness identified only one RCT [the Multicenter Study of Enhanced External Counterpulsation (MUST-EECP) trial[160,161]] comparing active with inactive EECP. This trial showed evidence of improved HRQoL from EECP (see *Table 62*); however, follow-up was limited to 12 months and so the degree to which improvement in HRQoL from EECP is sustained beyond 12 months is uncertain. To characterise the uncertainty associated with possible durations of treatment effect, formal elicitation of expert clinical judgement was undertaken.[167] Five experts with experience and knowledge of EECP in the UK independently completed an Microsoft Excel-based exercise that elicited their judgements about the likelihood of sustaining HRQoL benefits from EECP in subsequent years.[a] Because the uncertainty associated with any judgement is critical, a frequency chart format to represent a distribution was adopted.[168] The results from each expert were linearly pooled, with equal weight, providing the probability of continuing to respond to treatment in subsequent years.[169] The uncertainty associated with the pooled estimates was characterised by fitting beta distributions to pooled responses. See *Table 62* for a summary of the mean values and distributions for the probability of sustaining HRQoL benefits in each year.

Decision model

A probabilistic decision-analytic model was developed to estimate the expected cost-effectiveness of EECP and associated uncertainty. The decision model was structured to capture the HRQoL benefits and costs associated with EECP up to a period of 12 months and to project these benefits to a lifetime time horizon relevant to this patient population. Three health states were defined: (1) a 'responders' state represented patients who continued to sustain HRQoL benefits from EECP in each yearly cycle, (2) a 'non-responders' state represented patients who lost their initial HRQoL benefits from treatment and reverted back to baseline HRQoL and (3) 'dead' represented deaths from disease-specific cardiovascular causes[170] and other causes informed by UK life tables.[171] In the first year after treatment it was assumed that patients achieved the HRQoL benefits reported in the 12-month follow-up of MUST-EECP.[161] After 12 months the proportion of patients responding to treatment was based on the results of the elicitation exercise. After year 4, experts did not expect the probability of sustained response to be different in subsequent years. Therefore, a beta distribution, equivalent to that fitted to pooled year 4 responses, was assigned to year 5 onwards.

Some patients require additional 'top-up' sessions, which are generally given to help sustain the long-term benefits of treatment. Based on the study by Michaels *et al.*,[172] 18% of patients required repeat EECP within 2 years of the initial course of treatment. This was converted to an annual probability and decreased exponentially over time to reflect the diminishing attempts to continue to retreat if the patient is not responding. Resource use and costs associated with EECP were based on published unit costs. The cost for repeat procedures was based on an average of 10 additional sessions. *Table 62* summarises the key parameters for the decision model, the distributions assigned and the sources of evidence used. The analysis is conducted from the perspective of the NHS and Personal Social Services (PSS) and costs and outcomes are discounted at a rate of 3.5% per annum.[1]

Key features and possible pathways

Enhanced external counterpulsation results in large upfront costs of treatment but the potential quality-of-life benefits through improved symptoms and long-term relief from symptoms may

outweigh the costs when compared with not giving the treatment. The technology has a range of interesting characteristics that can be used to explore the implications of the principles and assessments outlined in *Chapter 5*:

- EECP is a medical device rather than a pharmaceutical.
- There is an element of investment cost associated with the purchase of equipment.
- EECP has large initial upfront costs of treatment (£4347 per patient), which are irrecoverable once treated.
- The impact of investment and irrecoverable costs on guidance can be fully explored.
- Formal elicitation methods are combined with Bayesian decision theory to determine a necessary and sufficient condition for conducting further research. This provides an opportunity to examine the potential role of elicitation rather than extreme scenarios to characterise uncertainty. It also provides an opportunity to examine the impact of the research design, in terms of length of follow-up, on guidance.
- Angina is a chronic condition so while research is conducted the prevalent population can still benefit from the information and switch treatment based on the results of the research.

The possible pathways through the algorithm are reported in *Figure 30* in *Appendix 4*. EECP is expected to be cost-effective but with potentially significant irrecoverable costs. These irrecoverable costs include both (1) long-lived costs associated with the purchase of equipment and (2) large initial per-patient treatment costs, combined with a chronic condition in which a decision not to treat a particular patient with EECP can be changed at a later date (decisions are not irreversible) when research reports or other events occur. Consequently, these irrecoverable costs might influence the category of guidance, for example OIR rather than approve.

The following sections examine each of the seven points on the checklist relating to the possible sequence of assessments and decisions that lead to a particular category and type of guidance for EECP.

Is it cost-effective and what are the risks?

Point 1: Is it expected to be cost-effective?

The sequence of assessments starts with cost-effectiveness and the expected impact on population NHEs, that is, at the following point in the algorithm:
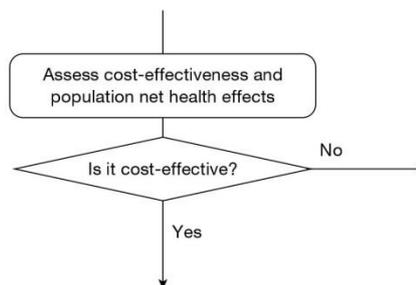
Cost-effectiveness at the patient level

Table 63 summarises the expected cost-effectiveness of EECP per patient treated. EECP as an adjunct to standard therapy is compared with standard therapy alone. Only the additional costs and effects of EECP over and above those of standard therapy are considered in the analysis. The expected incremental difference in lifetime costs and QALYs of EECP relative to standard

TABLE 62 Parameters of the decision model for EECP

Parameter	Mean value	Prior distribution	Source
Baseline patient characteristics			
Age	64 years	Fixed	MUST-EECP trial[161]
HRQoL improvement from baseline to 1 year after treatment			
EECP relative to standard therapy	0.0717	Beta($\alpha = 3.641$, $\beta = 47.139$)[a]	MUST-EECP trial[161]
Probability of sustaining HRQoL benefits in subsequent years[b]			
Year 2	0.757	Beta($\alpha = 8.028$, $\beta = 2.577$)	Expert elicitation[121]
Year 3	0.742	Beta($\alpha = 5.589$, $\beta = 1.943$)	Expert elicitation[121]
Year 4	0.719	Beta($\alpha = 4.413$, $\beta = 1.726$)	Expert elicitation[121]
Year 5 onwards	0.719	Beta($\alpha = 4.413$, $\beta = 1.726$)	Expert elicitation[121]
Probability of repeat EECP sessions			
Per year (decreased exponentially)	0.099	Beta($\alpha = 194$, $\beta = 884$)	Michaels *et al.*[172]
Resource use and unit costs			
Capital cost of EECP machine	£90,000	Fixed	K Miles, 2008, personal communication
Capital cost per annum (annuitised over 10 years)	£10,822	Fixed	
Equipment replacement costs			
One set of cuffs per year	£139	Fixed	Vasogenics[173]
One set of hoses per year	£76	Fixed	Vasogenics[173]
Pleth per every 2 years	£53	Fixed	Vasogenics[173]
Consumables per patient			
Ultrasound scan	£75	Fixed	Vasogenics[173]
Trousers	£16	Fixed	Vasogenics[173]
Gel	£8	Fixed	Vasogenics[173]
ECG electrodes	£110	Fixed	Vasogenics[173]
Staffing and overhead costs per patient (based on 12 patients per year)	£3214	Fixed	W Sheedy, Castle Hill Hospital, East Yorkshire, UK, 2008, personal communication
Total cost of EECP per patient	£4347	Fixed	

a   It is assumed that EECP always improves HRQoL in the first year after treatment compared with standard therapy based on the results of MUST-EECP, but this improvement may not be sustained in subsequent years.
b   Refers to probability of sustaining benefits during year *t* assuming that benefits have been sustained to year *t*−1.

TABLE 63 Expected cost-effectiveness of EECP per patient treated

Treatment	Cost (£)	QALYs	ICER (£/ QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£)	Incremental NHE, QALYs (£)	NHE, QALYs (£)	Incremental NHE, QALY (£)
EECP	4744	7.6045	19,391	7.3673 (147,346)	0.0074 (149)	7.4464 (223,391)	0.0865 (2595)
Standard care	–	7.3598	–	7.3598 (147,197)	–	7.3598 (220,795)	–

Results are presented as the mean of the probabilistic sensitivity analysis. See *Appendix 11, Model linearity and correlation between parameters* for assessments of linearity and correlation.

therapy is £4744 and 0.2447 QALYs, respectively, giving an ICER of £19,391. Therefore, EECP is just expected to be cost-effective at a threshold of £20,000 per QALY.[b] The NHE of EECP is greater than that of standard care but the difference per patient treated (the incremental NHE) is small. This difference of 0.0074 and 0.0865 QALYs at thresholds of £20,000 and £30,000 per QALY, respectively, represents the expected benefit per patient of immediate approval of EECP based on current evidence or, alternatively, the opportunity cost per patient of withholding approval of EECP if, for example, additional research is required that could not be conducted if it is approved.

As discussed in *Chapter 5, Is it cost-effective and what are the risks?*, it is also important to consider how NHEs accumulate over time or the investment profile per patient treated with EECP. *Figure 43* illustrates the cumulative incremental NHE over the patient time horizon of 50 years. The initial per-patient costs of EECP are high and are far in excess of the immediate health benefits in the initial period of treatment. These negative NHEs are gradually offset by positive NHEs in later periods. In this case, it is only after 14 years that the initial losses are compensated by later gains, that is, EECP doesn't break-even until 14 years from initial treatment. It is only beyond 30 years that the modest incremental NHEs reported in *Table 63* are eventually achieved.

Cost-effectiveness at the population level

Per patient NHEs can also be expressed for the population of current and future patients. This requires information about prevalence and future incidence of the target population. It also requires a judgement about the time horizon over which the technology will be used. An estimate of the scale of the total population NHEs and how they cumulate over time is important for subsequent assessments, including: (i) where the NHE for current patient populations must be compared with the benefits to future patients; and (ii) where the treatment decision can be changed such that the irrecoverable opportunity costs of initially negative NHE become significant, i.e., might influence the category of guidance.

The British Heart Foundation estimates that the prevalence of angina in the UK is just under 1.1 million and the annual incidence is around 95,000.[159] Assuming that 10% of angina patients can potentially benefit from EECP (Michael Chester, personal communication, 2008) this implies a prevalent population of 109,800 and an annual incidence of 9500. The total population
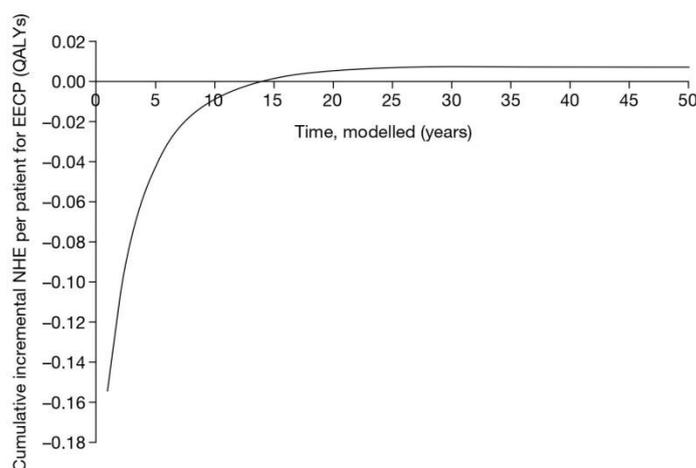
FIGURE 43 Cumulative incremental NHEs of EECP over the patient time horizon.
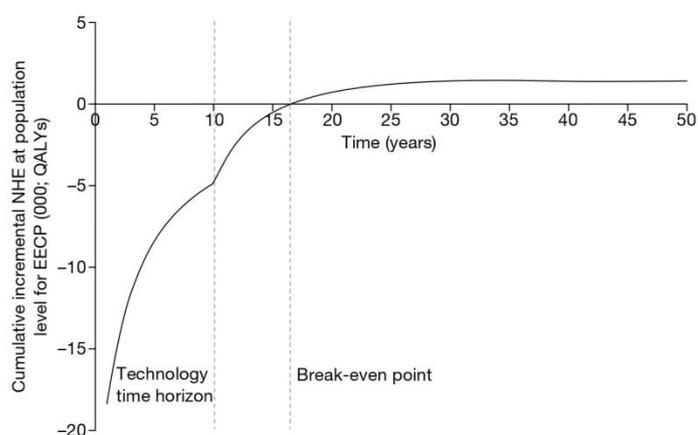
NHEs, assuming that the technology will be used to treat prevalent and incident patients over 10 years, are reported in *Table 64*. Expected cost-effectiveness is unchanged (ICER is the same as in *Table 63*) but the incremental NHEs, although small per patient, are more significant at a population level at 1405 and 16,334 QALYs at thresholds of £20,000 and £30,000 per QALY, respectively.

The investment profile for EECP when used to treat patients over 10 years is illustrated in *Figure 44*. At a population level it is not until 17 years (rather than 14 years at a patient level) that initial losses are compensated for by later gains and EECP breaks even. In other words, EECP appears a more risky investment when evaluated at a population rather than at an individual level. This is because, although each patient treated with EECP is expected to offer the same profile of NHEs shown in *Figure 43*, the negative NHEs associated with patients incident and treated in year 10 will not be offset by later gains until year 24. The population-level investment profile would exhibit greater risk (break-even later) if the prevalent population was smaller relative to the incident population and/or the technology time horizon was longer.

The time horizon of 10 years was chosen as a proxy for a complex and uncertain process of future changes in new technologies, prices and evidence.[53] The impact of different technology time horizons on NHE is illustrated in *Table 65*. The break-even point extends to 23 years when the technology time horizon is increased to 20 years. Therefore, the investment profile for EECP appears more risky the longer the time horizon of the technology.

TABLE 64 Expected cost-effectiveness of EECP for the population

| Treatment	Costs (£M)	QALYs	ICER (£/ QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£M)	Incremental NHE, QALYs (£M)	NHE, QALYs (£M)	Incremental NHE, QALYs (£M)
EECP	896	1,435,787	19,391	1,391,001 (27,820)	1405 (28)	1,405,930 (42,177)	16,334 (490)
Standard care	–	1,389,596	–	1,389,596 (27,792)		1,389,596 (41,688)	

FIGURE 44 Cumulative incremental NHEs of EECP for the population.

TABLE 65 Incremental NHEs of EECP over a range of technology time horizons

Technology time horizon (years)	Incremental NHE, QALYs (£M)	Break-even point (years)
5	1137 (22.7)	15
10	1405 (28.1)	17
15	1632 (32.6)	20
20	1822 (36.4)	23

Point 2: Are there significant irrecoverable costs?

The second point on the checklist requires (1) an assessment of whether or not there are irrecoverable costs and (2) a judgement of their potential significance, that is, at the following point in the algorithm:
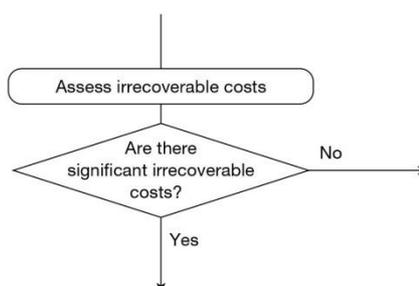
Identifying the irrecoverable costs

Irrecoverable costs are those that, once committed, cannot be recovered if guidance is changed at a later date. Irrecoverable costs are most commonly thought of as 'upfront' or capital costs of new facilities or equipment with long life expectancy. In the case of EECP, treatment requires capital investment in the EECP machines. The expected per-patient and population costs reported in *Tables 63* and *64* allocated this capital cost by annuitising the cost over the 10-year lifetime of the equipment at a rate of 3.5% per annum and allocating it to the number of patients treated each year (see *Table 61*). These irrecoverable costs are reported separately in *Table 66* and represent 19% of the total; however, this will have no influence on expected cost-effectiveness as long as guidance does not change during the lifetime of the equipment. *Figure 45* shows the effect on population NHEs when these capital costs are incurred in full in the first year rather than allocated per patient over the lifetime of the equipment. The cumulative incremental NHE for EECP is more negative in the first 10 years but the investment profile for EECP is no more risky when the capital cost is incurred up front. The investment profile would be expected to exhibit greater risk if the lifetime of the equipment was extended beyond 10 years.

Even in the absence of capital costs of EECP, NHEs accumulate over time. Therefore, approval or AWR commits opportunity costs of negative NHEs that are irrecoverable.

Are they likely to be significant?

The significance of the irrecoverable capital costs can be judged based on the proportion of the total population cost that they represent as well as their scale relative to the additional population NHEs offered. The irrecoverable capital cost of EECP represents 19% of the total cost (see *Table 66*) and could be considered significant. Furthermore, even in the absence of the capital costs EECP exhibits a profile of NHEs in which approval or AWR commits opportunity costs of negative NHEs that are irrecoverable. Irrecoverable costs include situations in which initial losses (negative NHEs) of a technology are offset by later gains (positive NHEs). If early approval is

TABLE 66 Capital costs associated with EECP

Treatment	Capital cost (£)	Non-capital cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY	
					NHE, QALYs (£M)	Incremental NHE, QALYs (£M)
EECP	170,304,591	725,408,798	1,435,787	19,391	1,391,001 (27,820)	1405 (28)
Standard care	–	–	1,389,596	–	1,389,596 (27,792)	–

FIGURE 45 Cumulative incremental NHEs with capital costs incurred in year 1.

revised, for example because research reveals that the technology is not as effective as expected, then initial losses will have been incurred but they will not be compensated for by later gains. For example, if research reports before the break-even point of 17 years (i.e. before losses are recouped) there is a chance that the results will indicate that EECP is not in fact cost-effective, approval will be withdrawn and patients will be switched from EECP to standard care; initial losses will have been incurred but they will not be compensated for by the later gains that were originally expected. The impact of this investment profile of cumulative incremental NHEs will be greater when a decision to approve is more likely to change and in the more immediate future, that is, when it is more uncertain and when research will be conducted and report in the near future.

Enhanced external counterpulsation is for a chronic condition in which the decision to treat a particular patient can be changed at some later date (decisions are not irreversible). Therefore, the type of investment profile of NHEs at a patient and population level is significant because, instead of committing irrecoverable costs by deciding to use the technology expected to be cost-effective now, the decision and commitment of costs can be made later, after research reports, other events occur and/or guidance changes.

Figure 45 illustrates the impact of accounting for the actual timing of expenditure on EECP machines rather than treating it as if it was paid when each patient was treated, that is, expenditure is treated like a consumable cost by spreading the capital cost over 10 years.[c] If approval of EECP might be withdrawn before 10 years, the potential losses in NHEs will be greater than initially indicated in *Figure 45* because the equipment costs allocated to treating future patients cannot be recovered. The earlier such a change might occur the greater the additional loss. The impact of these possibilities should be considered at point 7 of the checklist before guidance to approve or AWR commits both types of irrecoverable costs.

Types and categories of guidance resulting from points 1 and 2

Points 1 and 2 of the checklist do not lead directly to a category or type of guidance. The sequence of assessments and decisions that ultimately lead to guidance starts with cost-effectiveness, expected impact on population NHEs and significance of irrecoverable costs. In the case of EECP, the technology is expected to be cost-effective and has significant irrecoverable costs. *Table 67* summarises the categories and types of guidance that could ultimately result from the sequence of assessments up to point 2 (see *Table 32* in *Appendix 4*).

Is further research required?

Point 3: Does more research seem worthwhile?

The third point on the checklist requires an assessment of the potential benefits of conducting further research, that is, at the following point in the algorithm:
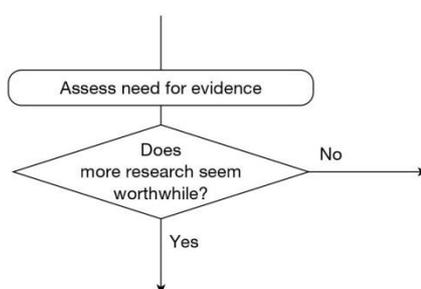
This requires judgements about (1) how uncertain a decision to approve or reject EECP might be based on the estimates of expected cost-effectiveness and (2) whether or not the scale of the likely consequences of this uncertainty might justify further research. Some assessment of the potential consequences of uncertainty is important because it indicates the population NHEs that could be gained if the uncertainty surrounding this decision could be resolved immediately, that is, it represents an expected upper bound on the benefits of more research.[d]

TABLE 67 Types and categories of guidance that could ultimately result from assessments 1 and 2

Assessment	1	2	3	4	5	6	7	Guidance
13	Yes	Yes	Yes	Yes	Yes	Yes	Yes	AWR[3]
14	Yes	Yes	Yes	Yes	Yes	Yes	No	OIR[3]
15	Yes	Yes	Yes	Yes	Yes	No	Yes	Approve[5]
16	Yes	Yes	Yes	Yes	Yes	No	No	Reject[5]
17	Yes	Yes	Yes	Yes	No	Yes	Yes	AWR[4]
18	Yes	Yes	Yes	Yes	No	Yes	No	OIR[4]
19	Yes	Yes	Yes	Yes	No	No	–	Approve[6]
20	Yes	Yes	Yes	No	Yes	Yes	Yes	Approve[7]
21	Yes	Yes	Yes	No	Yes	Yes	No	OIR[5]
22	Yes	Yes	Yes	No	Yes	No	Yes	Approve[8]
23	Yes	Yes	Yes	No	Yes	No	No	Reject[6]
24	Yes	Yes	Yes	No	No	Yes	Yes	Approve[9]
25	Yes	Yes	Yes	No	No	Yes	No	OIR[6]
26	Yes	Yes	Yes	No	No	No	–	Approve[10]

Assessing the consequences of uncertainty

Enhanced external counterpulsation is expected to be cost-effective compared with standard care (see *Tables 63* and *64*) but the estimates of cost and QALYs are uncertain so there is a chance that a decision to approve EECP based on existing evidence will be incorrect, that is, standard care might offer greater NHEs. Some assessment of the likely consequences of approving EECP when standard care might be better could be based on the difference in expected NHEs, that is, the expected incremental population NHEs reported in *Table 63*. The simplest approach would be to weight the average NHEs for EECP and standard care (reported in *Table 63*) by a judgement of the probability of an incorrect decision. For example, if the decision was judged to be 100% certain then there are no consequences and so there would be nothing to be gained by more research; however, as the probability that the decision is correct becomes less certain, the expected consequences (and hence potential value of more research) increase. *Table 68* shows the expected consequences of uncertainty based on a weighting of average NHEs.

A judgement of how uncertain a decision might be can be informed by the PSA already used to estimate costs and QALYs. The probability that EECP is cost-effective is 0.428 at £20,000 per QALY and 0.700 at £30,000 per QALY (*Table 69*).[e] This translates into approximately 800 QALYs (*Figure 46*) based on the expected or average difference in NHEs at a threshold of £20,000 per QALY.

However, the difference in NHEs when EECP is not the correct decision is not necessarily the average (see *Appendix 11, Why the consequences of uncertainty differ from mean incremental effects*). In fact, it is very unlikely to be the average and such estimates may substantially under- or overestimate the expected consequences of uncertainty. At £20,000 per QALY, the estimate of 800 QALYs based on mean incremental population NHEs is a substantial underestimate of the expected consequences of 9287 QALYs (see *Table 69* and *Figure 46*) based on the distribution of uncertainty from the PSA. Similarly, the estimate of 4900 QALYs at a threshold of £30,000 per QALY (*Figure 47*) is an overestimate of the expected consequences of 2774 QALYs (see *Table 69*) based on the distribution of uncertainty from the PSA. Therefore, it is only appropriate to conduct an analysis of the expected consequences of uncertainty based on the PSA rather than use a simple estimate of average NHEs, which can be misleading.

The same PSA can be used to record the frequency of errors in the decision to approve EECP. *Figure 48* shows the distribution of consequences of uncertainty for EECP. Most commonly there are no consequences because EECP is the correct decision 42.8% of the time. When EECP offers

TABLE 68 Expected consequences of uncertainty based on a weighting of average NHEs

Probability decision is correct	Cost-effectiveness threshold £20,000 per QALY	Cost-effectiveness threshold £30,000 per QALY
	Expected consequences, QALYs (£M)	Expected consequences, QALYs (£M)
1.00	0 (0)	0 (0)
0.99	14 (0.3)	163 (4.9)
0.95	70 (1.4)	817 (25)
0.90	141 (2.8)	1633 (49)
0.75	351 (7.0)	4083 (123)
0.50	703 (14)	8167 (245)
0.25	1054 (21)	12,250 (368)
0.10	1265 (25)	14,701 (441)
0.05	1335 (27)	15,517 (466)
0.01	1391 (28)	16,171 (485)
0.00	1405 (28)	16,334 (490)

TABLE 69 Expected consequences of uncertainty for EECP

Treatment	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY			Cost-effectiveness threshold £30,000 per QALY		
		Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)	Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
EECP	19,391	1405 (28.1)	0.428	9287 (185.7)	1,405,930 (490)	0.700	2774 (83.2)
Standard care		–	0.572		–	0.300	

FIGURE 46 Probability that EECP is cost-effective and the consequences of uncertainty at £20,000.
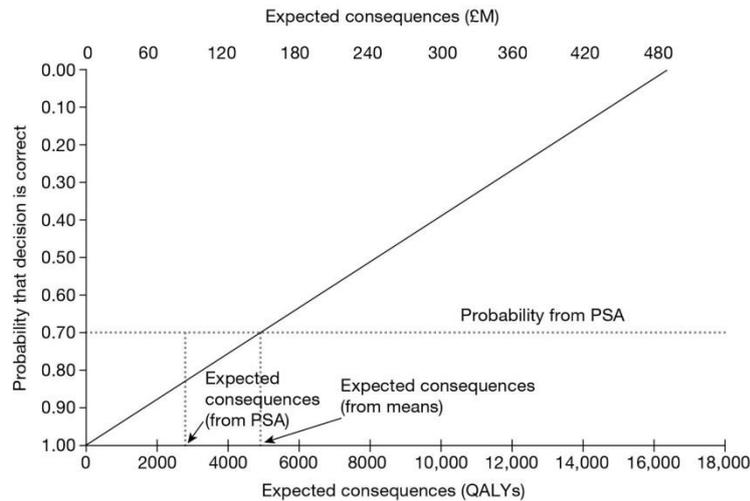
FIGURE 47 Probability that EECP is cost-effective and the consequences of uncertainty at £30,000.

FIGURE 48 Distribution of the consequences of uncertainty for EECP.

lower NHEs than standard care the consequences of error may be relatively small, for example 9% are < 5000 QALYs. However, they may be very large, although this is less likely, for example there is a small chance (5.7%) that they are > 30,000 QALYs. The average over this distribution provides the expected consequences of uncertainty (9287 QALYs).

These expected consequences can be interpreted as an estimate of the population NHEs over the technology time horizon that could be gained if the uncertainty surrounding this deci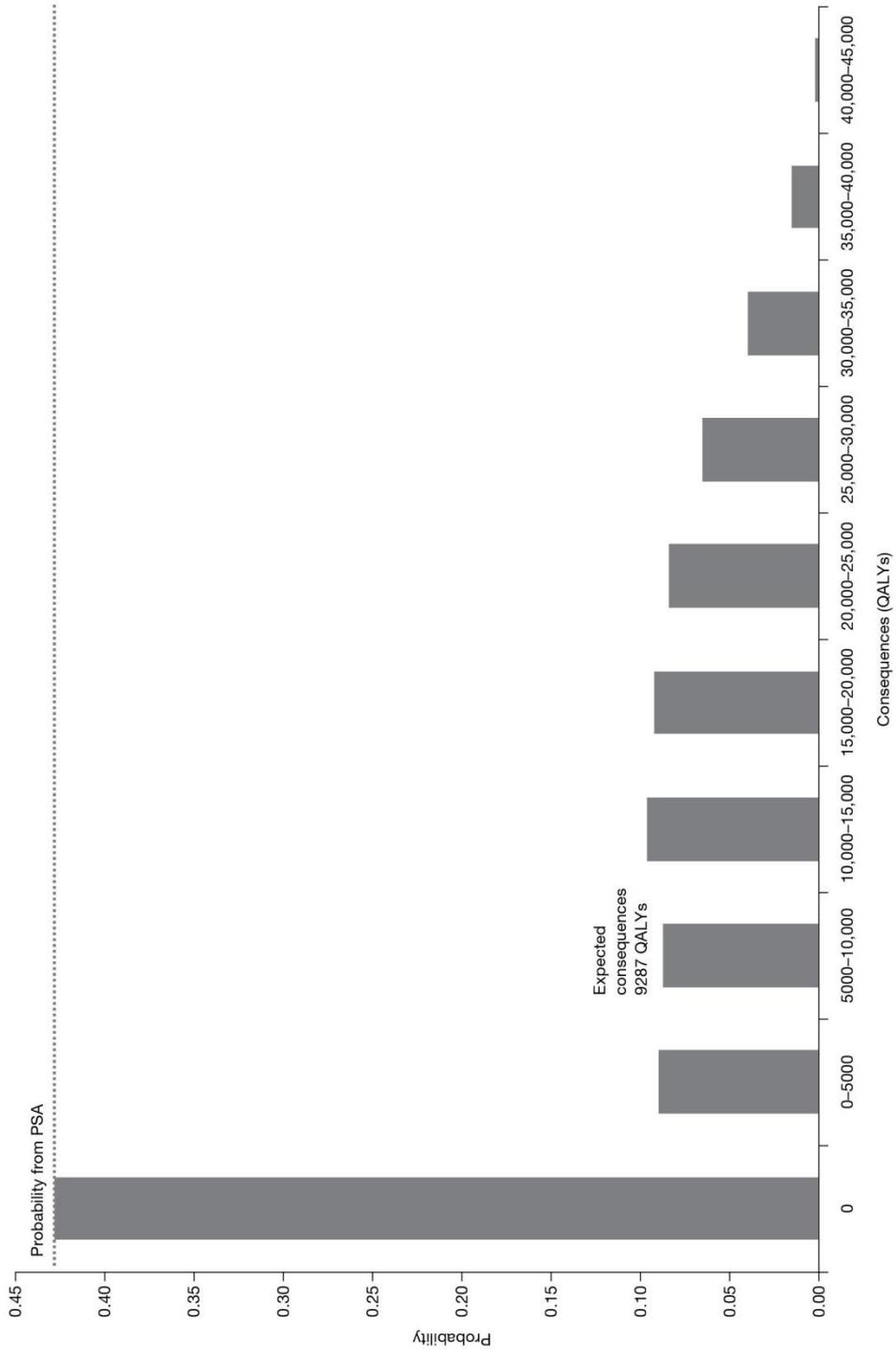sion could be resolved immediately, that is, it indicates an expected upper bound on the benefits of more research. The consequences can also be expressed as the equivalent NHS resources required to generate the same population NHEs (£185.7M in *Table 69*). In the case of EECP the consequences fall with the cost-effectiveness threshold because a decision to approve EECP will be less uncertain (see *Table 69*). A judgement at this point that more research might be worthwhile seems reasonable as the upper bound on its potential benefits exceeds the likely costs.

The time horizon over which evidence generated by research about a technology might be valuable may be longer (or shorter) than the period over which the technology is used. In this case, the technology time horizon is assumed to be equal to the time horizon for the benefits of research, that is, 10 years. *Table 70* shows the expected consequences of uncertainty for different technology time horizons. The consequences increase with the technology time horizon and will also increase with the size of the patient population.

Analysis of subgroups
There are no relevant subgroups in the EECP case study.

Alternative scenarios
The uncertainty described above reflects uncertainty within the set of assumptions used to estimate expected costs and QALYs. However, there are often alternative views about the assumptions, which are usually presented as separate scenarios. When more than one scenario might be credible and carry some 'weight', there will be uncertainty *between* as well as *within* scenarios. The single RCT of EECP showed evidence of improvements in quality of life at 12 months; however, the degree to which these are sustained in the long run is uncertain. Rather than make alternative assumptions and present extreme scenarios, formal elicitation of the judgements of clinical experts about the likelihood of QALY gains in subsequent years was undertaken. The uncertainty in these elicited values is included in the estimates of the expected consequences of uncertainty reported above. However, the analysis could be presented as alternative scenarios:

- scenario A – HRQoL benefits are sustained for only 1 year (first 12 months)
- scenario B – HRQoL benefits are sustained for a patient's lifetime
- scenario C – HRQoL benefits are sustained for 4 years.

TABLE 70 Expected consequences of uncertainty for different technology time horizons

Technology time horizon, years	Expected consequences, QALYs (£M)
5	7511 (150)
10	9287 (186)
15	10,783 (216)
20	12,042 (241)

Scenarios A and B are extreme scenarios whereas scenario C lies somewhere between the two extremes. *Table 71* presents the cost-effectiveness of EECP for the alternative scenarios. EECP is not expected to be cost-effective under scenario A whereas it is highly cost-effective under scenario B. For scenario C, EECP would not be regarded as cost-effective at a threshold of £20,000 per QALY but could be approved for use, based on current evidence, at a threshold of £30,000 per QALY. However, as above, some assessment of the likely consequences of approving EECP or standard care when the alternative treatment might be better should be informed by the PSA. *Table 72* shows the expected consequences if the decision to approve the technology expected to be cost-effective turns out to be wrong. Under scenarios A and B the expected consequences are small (1.4 and 394 QALYs at a threshold of £20,000 per QALY, respectively) because there is less uncertainty about the decision (99.95% and 96.5% chance of no error, respectively); however, scenario C is more uncertain with expected consequences of 3228 QALYs.

The expected consequences of uncertainty within each scenario are sufficient to indicate the potential benefits of research when only one scenario is regarded as credible. However, when more than one scenario might be credible and carry some 'weight', there will be uncertainty between as well as within scenarios. The 'weighting' of scenarios can be made explicit by assigning probabilities to represent how credible each is believed to be. The weighted average of costs and QALYs across scenarios can easily be calculated. It is tempting to take a simple weighted average of the expected consequences of uncertainty across the scenarios. However, a simple weighted average may under- or overestimate the combined consequences of uncertainty within and between scenarios. The correct estimate requires the probabilities (weights) to be applied directly to the simulated output from PSA rather than to the mean values. *Table 73* shows that the simple weighted average of expected consequences (linear combination of mean estimates) substantially underestimates the correct estimate based on weighting the output of PSA.

TABLE 71 Expected cost-effectiveness of EECP with alternative scenarios

Treatment	Cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£M)	Incremental NHE, QALYs (£M)	NHE, QALYs (£M)	Incremental NHE, QALYs (£M)
Scenario A							
EECP	861,758,393	1,401,330	65,292	1,358,242 (27,165)	−29,889 (−598)	1,372,605 (41,178)	−15,527 (−466)
Standard care	–	1,388,132	–	1,388,132 (27,763)	–	1,388,132 (41,644)	–
Scenario B							
EECP	966,199,093	1,555,685	5767	1,507,375 (30,147)	119,243 (2385)	1,523,478 (45,704)	135,347 (4060)
Standard care	–	1,388,132	–	1,388,132 (27,763)	–	1,388,132 (41,644)	–
Scenario C							
EECP	886,474,497	1,423,008	26,531	1,378,685 (27,574)	−10,911 (−218)	1,393,459 (41,804)	3864 (116)
Standard care	–	1,389,596	–	1,389,596 (27,792)	–	1,389,596 (41,688)	–

The results of the elicitation can be used to provide an estimate of the 'weighting' of scenarios. The elicitation implied probabilities of 0.243, 0.353 and 0.404 associated with scenarios A, B and C, respectively. A simple weighted average of the expected consequences within each scenario using these probabilities (1442 QALYs) significantly underestimates both the estimate of expected consequences based on all of the information from elicitation (9287 QALYs) and the estimate based on weighting the simulated output rather than the mean estimates (13,081 QALYs). This illustrates (1) that a simple weighted average of expected consequences may be misleading and (2) that elicitation may provide a richer characterisation of uncertainty as well as the probabilities associated with alternative assumptions (see *Appendix 11*, *Why averaging scenarios may be misleading*).

TABLE 72 Expected consequences of uncertainty with alternative scenarios

Treatment	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY			Cost-effectiveness threshold £30,000 per QALY		
		Incremental NHE, QALYs (£)	Probability cost-effective	Expected consequences, QALYs (£M)	Incremental NHE, QALYs (£)	Probability cost-effective	Expected consequences, QALYs (£M)
Scenario A							
EECP	65,292	−29,889 (−598)	0.0005	1.4 (0.028)	−15,527 (−466)	0.0278	117 (3.5)
Standard care	–	–	0.9995		–	0.9722	
Scenario B							
EECP	5767	119,243 (2385)	0.9649	394 (7.8)	135,347 (4060)	0.9905	67 (2.0)
Standard care	–	–	0.0351		–	0.0095	
Scenario C							
EECP	26,531	−10,911 (−218)	0.2256	3228 (64)	3864 (116)	0.5261	4715 (141)
Standard care	–	–	0.7744		–	0.4739	

TABLE 73 Expected consequences of uncertainty with alternative scenarios[a]

Probability of scenario C	Probability of scenario A	Probability of scenario B	Expected consequences, QALYs (£M)	
			Weighted	PSA
1.00	0.00	0.00	3228 (64)	3228 (64)
0.90	0.05	0.05	2925 (59)	8966 (179)
0.80	0.10	0.10	2622 (52)	14,369 (287)
0.70	0.15	0.15	2319 (46)	14,426 (289)
0.60	0.20	0.20	2016 (40)	14,480 (290)
0.50	0.25	0.25	1713 (34)	14,580 (292)
0.40	0.30	0.30	1410 (28)	14,700 (294)
0.30	0.35	0.35	1107 (22)	14,849 (297)
0.20	0.40	0.40	804 (16)	14,972 (299)
0.10	0.45	0.45	501 (10)	15,039 (301)
0.00	0.50	0.50	198 (4)	15,106 (302)

a Probability of scenario C with error probability equally distributed between scenarios A and B.

Point 4: Is research possible with approval?

The fourth point on the checklist requires an assessment of what type of evidence is needed and a judgement of whether or not the research required to generate it can be conducted while the technology is approved, i.e., at the following point in the algorithm:
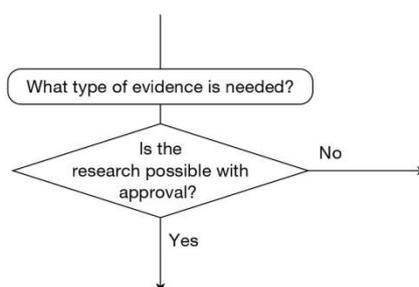
This requires judgements about (1) how important particular types of parameters are to estimates of cost and QALYs, (2) what values these parameters would have to take to change a decision based on expected cost-effectiveness, (3) how likely it is that parameters might take such values and (4) what would be the consequences if they did, that is, what might be gained in terms of population NHEs if the uncertainty in the values of these parameters could be immediately resolved.

Assessing the importance of parameters

The expected cost-effectiveness of EECP is based on the relationship between the input parameters (see *Table 62*) and outputs of cost and QALYs. A simple summary of the direction and strength of these relationships can be provided by the elasticity, that is, the proportionate change in NHEs due to a 1% change in the value of the parameter (*Table 74*). Parameters with high elasticities (especially with respect to differences in NHE) might be regarded as more 'important'.

Although these measures of importance are more instructive than a series of arbitrary one-way sensitivity analyses, they do not directly help the assessment of what values parameters must take to change decisions and how likely such values might be. A simple summary of the values that particular parameters must take to make each of the alternatives cost-effective can also be provided (*Table 75*); however, although instructive, such 'threshold values' do not indicate how likely it is that the threshold will be crossed.

TABLE 74 Elasticity for small changes (1%) in mean parameter values

Parameter	Elasticity for NHEs, QALYs		Elasticity for incremental NHEs, QALYs
	EECP	Standard care	
HRQoL increment in first year from EECP	−0.6419	−0.8899	0.2480
Probability of sustaining HRQoL benefits in year 2 (elicited)	0.1682	0.0000	0.1682
Probability of sustaining HRQoL benefits in year 3 (elicited)	0.1218	0.0000	0.1218
Probability of sustaining HRQoL benefits in year 4 (elicited)	0.0881	0.0000	0.0881
Probability of sustaining HRQoL benefits in subsequent years (group of elicited parameters)	0.3726	0.0000	0.3726
2-year probability of requiring repeat EECP sessions	−0.0165	0.0000	−0.0165
Cost of EECP	−0.2372	0.0000	−0.2372

Assessment of uncertainty

The judgement about how likely it is that parameters might take values that will change the technology expected to be cost-effective can be informed by the results of the PSA. The distributions assigned to the parameters in PSA describe how uncertain the parameter estimates are, such that they ought to reflect the amount and quality of existing evidence. The probability that each parameter might take values that would lead to each of the alternatives being cost-effective is reported in *Table 76* for the uncertain parameters.

What type of evidence is needed?

An assessment of the likely consequences of the uncertainty described above is required. This assessment can directly inform the judgement of what evidence is needed and whether or not the type of research required to generate it will be possible with approval. The expected consequences of uncertainty associated with each parameter are reported in *Table 77*. This decomposes the overall expected consequences into the contribution that each parameter (or group of parameters) makes. Note that the overall expected consequences of uncertainty will not, in general, equal the sum of the expected consequences for each of the parameters (or groups of parameters) separately. This is because the overall consequences take account of the joint effect of uncertainty in all parameters simultaneously. Even if parameters are independent they will be related to differences in NHEs in different ways (see *Table 74*); sometimes the effect of uncertainty in one parameter may, to some extent, substitute for or complement the effect of uncertainty in the others.

The most significant consequences of uncertainty associated with parameters relate to the relative treatment effect of EECP in the first year in terms of its improvement of HRQoL. Interestingly, although the probability of sustaining the QALY benefits of EECP in the long run is very uncertain, the greater part of potential value is in more precise estimates of QALY gains in the first 12 months (2709 QALYs or £54M and 8511 QALYs or £170M, respectively).

Because more precise estimates of the relative treatment effect are required research needed might not be possible for the reasons discussed in *Chapter 5*. If it is possible, more precise

TABLE 75 Range of parameter values to make each alternative cost-effective

Parameter	Mean value	EECP	Standard care
HRQoL increment in first year from EECP	0.0717	0.0686 to max.	0–0.0685
Probability of sustaining HRQoL benefits in year 2 (elicited)	0.7570	0.7123–1	0–0.7122
Probability of sustaining HRQoL benefits in year 3 (elicited)	0.7420	0.6782–1	0–0.6781
Probability of sustaining HRQoL benefits in year 4 (elicited)	0.7188	0.6419–1	0–0.6418
2-year probability of requiring repeat EECP sessions	0.1780	0–0.2680	0.2681–1
Cost of EECP	£4347	£0–4486	£4487 to max.

Max., maximum.

TABLE 76 Probabilities associated with parameter values

Parameters	EECP	Standard care
HRQoL increment in first year from EECP	0.454	0.546
Probability of sustaining HRQoL benefits in year 2 (elicited)	0.627	0.373
Probability of sustaining HRQoL benefits in year 3 (elicited)	0.653	0.347
Probability of sustaining HRQoL benefits in year 4 (elicited)	0.661	0.339
2-year probability of requiring repeat EECP sessions	1.000	0.000

estimates of treatment effect either in the short or in the longer term will require an experimental design if selection bias is to be avoided. The results in *Table 77* suggest that a further RCT may well be worthwhile; however, whether such a trial should have a follow-up of 12 months or a longer follow-up duration depends on if the additional benefits of a longer follow-up to more precisely estimate duration of effect exceed the additional opportunity costs of a more costly and lengthy trial (see later sections).

Implications of between-scenario uncertainty

Earlier the contribution that alternative scenarios might make to the overall expected consequences of uncertainty and therefore the potential gains from further evidence was considered. In situations in which more than one scenario might be regarded as credible there will be uncertainty *between* as well as *within* each of the scenarios. It was demonstrated in *Alternative scenarios* that an assessment of the combined consequences of both sources of uncertainty requires 'weights' (probabilities) to be assigned to represent their credibility, which can then be applied directly to the simulated output from PSA (see *Appendix 11, Why averaging scenarios may be misleading*). The same analysis can also be used to identify the expected consequences of uncertainty associated with the alternative scenarios themselves, that is, what might be gained if evidence could immediately distinguish which scenario was 'true'. This can help to inform the assessment of what type of evidence might be needed and whether or not the research required to generate it is likely to be possible once a technology is approved for widespread NHS use.

In *Alternative scenarios* three scenarios were considered: A – HRQoL benefits are sustained for only 12 months; B – HRQoL benefits are sustained for a patient's lifetime; and C – HRQoL benefits are sustained for 4 years. *Table 78* shows the expected consequences of uncertainty within scenarios, between scenarios and combining consequences within and between scenarios. The expected consequences of the uncertainty between scenarios are much greater than the expected consequences within scenarios, suggesting that further evidence about EECP that could help to distinguish between the scenarios is more valuable than the parameters associated with each.

Formal elicitation of the judgement of clinical experts about whether observed QALY gains at 12 months are likely to be sustained in subsequent years was undertaken. Because the uncertainty in these elicited values was incorporated into the analysis in the same way as other parameters, the use of alternative scenarios was not necessary. The results of elicitation implied probabilities of 0.243, 0.353 and 0.404 associated with scenarios A, B and C, respectively. Based on these 'weights' for each scenario the overall expected consequences of uncertainty (combining the consequences within and between scenarios) would be 14,146 QALYs. The expected consequences of uncertainty between the scenarios (13,202 QALYs) are much greater than what

TABLE 77 Consequences of uncertainty associated with parameter values

Parameters	Expected consequences, QALYs (£M)
HRQoL increment in first year from EECP	8511 (170)
Probability of sustaining HRQoL benefits in year 2 (elicited)	1644 (33)
Probability of sustaining HRQoL benefits in year 3 (elicited)	1398 (28)
Probability of sustaining HRQoL benefits in year 4 (elicited)	1058 (21)
Probability of sustaining HRQoL benefits in subsequent years (group of elicited parameters)	2709 (54)
2-year probability of requiring repeat EECP sessions	0 (0)
Overall expected consequences of uncertainty	9440 (189)

might be potentially gained from resolving the uncertainty within each scenario (1765 QALYs). Therefore, most of what might be gained from further evidence about EECP (in the absence of formal elicitation) would be evidence that could help distinguish between the scenarios rather than the parameters associated with each.

Types and categories of guidance resulting from points 3 and 4

Points 3 and 4 of the checklist are critical because if research is not judged to be worthwhile no further assessments are required. In the case of EECP, more research appears to be worthwhile; the expected consequences of uncertainty are high. Whether or not the research required to generate the evidence needed can be conducted while the technology is approved for widespread use will determine whether AWR or OIR is a possibility. *Table 79* summarises the categories and types of guidance that could ultimately result from the sequence of assessments up to point 4.

TABLE 78 Expected consequences of uncertainty within and between scenarios[a]

Probability of scenario C	Probability of scenario A	Probability of scenario B	Expected consequences, QALYs (£M)		Within and between scenarios
			Within scenarios	Between scenarios	
1.00	0.00	0.00	3228 (65)	0 (0)	3228 (65)
0.90	0.05	0.05	4281 (86)	5962 (119)	8887 (178)
0.80	0.10	0.10	6380 (128)	11,718 (234)	14,340 (287)
0.70	0.15	0.15	4406 (88)	12,121 (242)	14,440 (289)
0.60	0.20	0.20	3273 (65)	12,524 (250)	14,540 (291)
0.50	0.25	0.25	2637 (53)	12,928 (259)	14,641 (293)
0.40	0.30	0.30	2279 (46)	13,331 (267)	14,741 (295)
0.30	0.35	0.35	2078 (42)	13,735 (275)	14,841 (297)
0.20	0.40	0.40	1972 (39)	14,138 (283)	14,942 (299)
0.10	0.45	0.45	1940 (39)	14,541 (291)	15,042 (301)
0.00	0.50	0.50	1947 (39)	14,945 (299)	15,142 (303)

a Probability of scenario C with error probability equally distributed between scenarios A and B.

TABLE 79 Types and categories of guidance that could ultimately result from assessments 3 and 4

Assessment	1	2	3	4	5	6	7	Guidance
13	Yes	Yes	Yes	Yes	Yes	Yes	Yes	AWR[3]
14	Yes	Yes	Yes	Yes	Yes	Yes	No	OIR[3]
15	Yes	Yes	Yes	Yes	Yes	No	Yes	Approve[5]
16	Yes	Yes	Yes	Yes	Yes	No	No	Reject[5]
17	Yes	Yes	Yes	Yes	No	Yes	Yes	AWR[4]
18	Yes	Yes	Yes	Yes	No	Yes	No	OIR[4]
19	Yes	Yes	Yes	Yes	No	No	–	Approve[6]
20	Yes	Yes	Yes	No	Yes	Yes	Yes	Approve[7]
21	Yes	Yes	Yes	No	Yes	Yes	No	OIR[5]
22	Yes	Yes	Yes	No	Yes	No	Yes	Approve[8]
23	Yes	Yes	Yes	No	Yes	No	No	Reject[6]
24	Yes	Yes	Yes	No	No	Yes	Yes	Approve[9]
25	Yes	Yes	Yes	No	No	Yes	No	OIR[6]
26	Yes	Yes	Yes	No	No	No	–	Approve[10]

Do the benefits of research exceed the costs?

Point 5: Will other sources of uncertainty resolve over time?

The fifth point on the checklist requires an assessment of whether or not changes are likely to occur in the future that will influence the cost-effectiveness of the alternative technologies and the potential benefits of research, that is, at the following point in the algorithm:

This assessment requires information about (1) changes in the prices of the technology and its comparators, (2) the emergence of new technologies that might make existing ones obsolete or change their cost-effectiveness and (3) other relevant research reporting. A number of potential sources of information and evidence were examined to inform this assessment (see *Appendix 4* for full details of the sources and searches conducted).

Changes in the prices of the technology and its comparators

Changes in prices influence not only expected cost-effectiveness but also uncertainty and the potential benefits of research to future patients, for example if the price of a technology expected to be cost-effective is likely to fall significantly just before research reports the potential benefits will not be realised because approval of the technology will be less uncertain and there may be much less or little to gain from the results of the research. This assessment requires information about when major changes in prices are likely and some evidence about the likely extent of any change. It is very difficult to locate information on changes in price relevant to devices such as EECP as a device may have only a CE mark, which, unlike a patent, does not offer protection and can be renewed every 10 years. Any patent is likely to relate to some aspect of the device rather than the device itself. EECP was filed in 1995, renewed in 2005 and is up for renewal in 2015. The manufacturer (Vasogenics, Sapphire House, Albion Road, Bradford, West Yorkshire, UK) is still heavily marketing EECP so it is fair to assume that it will renew in 2015. There are no current comparators to EECP (apart from standard care). For these reasons a change in price for EECP is not anticipated and is not explored further.

Entry of new technologies

The entry of a new technology may make the existing technology that is expected to be cost-effective obsolete (no longer the most cost-effective alternative). Even when it does not, it will tend to change the relative cost-effectiveness of the alternatives, influencing how uncertain a decision to approve the original technology will be for future patients and the potential gains from research. A number of potential sources of information were examined to identify new technologies relevant to stable angina that were likely to become available. In particular, products considered by the MTAC as part of the Evaluation Pathway Programme for Medical Technologies at NICE were searched using the term 'angina'.[174] No relevant records were found. Although there is no evidence of new technologies emerging in EECP, different scenarios are explored in *Point 6:*
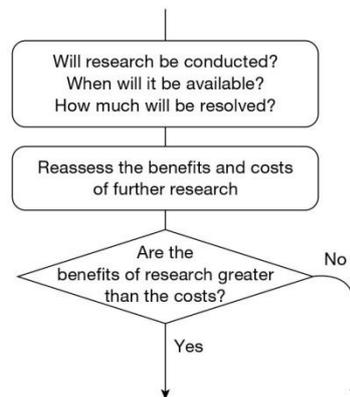
Are the benefits of research greater than the costs? as the development and launch of new devices are more difficult to identify in advance.

Other research reporting of the technology and its comparators

Research that is already under way, commissioned or likely to be undertaken, whether in the UK or elsewhere, is relevant. A number of potential sources of information were examined to identify clinical research under way at the time of the assessment for EECP including national and international trial registries (ClinicalTrials.gov, WHO ICTRP and Current Controlled Trials) as well as other databases that report NHS-funded research and not just clinical trials (e.g. NRR and UKCRN). Despite an assiduous search no records relevant to EECP were identified.

Point 6: Are the benefits of research greater than the costs?

The sixth point on the checklist requires a judgement of whether the potential benefits of conducting further research (initially considered at point 3) are likely to exceed the costs, that is, at the following point in the algorithm:

This requires an assessment of (1) whether or not the type of research that is required is likely to be conducted, (2) if conducted, when the results are likely to be available, (3) how much uncertainty is likely to be resolved and (4) the likely impact of any other sources of uncertainty on the longer-term benefits of research.

Will the research be conducted?

Even if research is recommended in OIR or AWR, it might not be undertaken by manufacturers or commissioned by research funders and there is no guarantee that research will be able to recruit or complete. The expected consequences of uncertainty for EECP reported in *Point 3: Does more research seem worthwhile?* are illustrated in *Figure 49* for a range of probabilities that research will be successfully undertaken. The potential gains from research depend on a judgement of whether the research recommended as part of OIR or AWR will be successfully completed. The cost of research (in this case considered to be either £1.5M or £10M) can be compared directly with the potential benefits by either expressing the potential gains in population NHEs as the equivalent NHS resources (i.e. the resources that would be required to generate the same NHEs) or expressing the cost of research in terms of the QALYs that could be gained elsewhere in the NHS by using the same resources to provide access to health care.

When will it be available?

Research, even if commissioned and successfully completed, will take time to complete and report. Therefore, any assessment of the potential benefits should account for the fact that patient
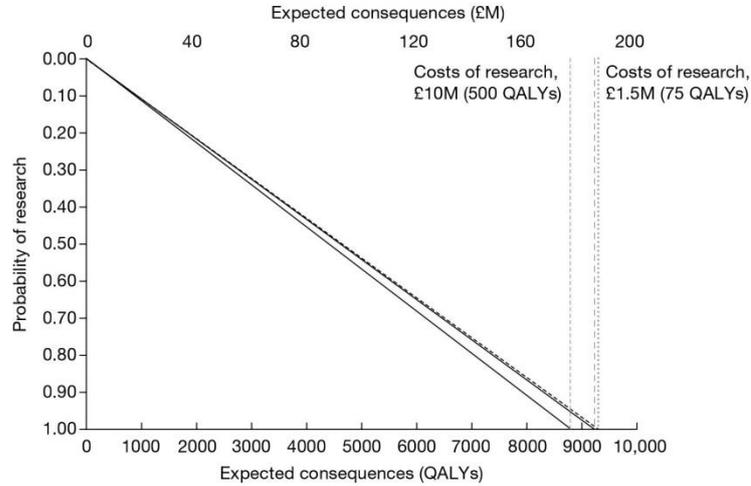
FIGURE 49 Expected potential benefits of research.

populations will not benefit from the results of research until they are available. If treatment decisions are irreversible (e.g. an acute indication) then it is only those patients who are incident after the research reports who will realise any of the potential benefits. However, for treatment decisions that are not irreversible, which is the case for EECP as it is a chronic condition, prevalent and incident patients can benefit from the results of the research. Patients prevalent while the research is undertaken will not benefit immediately but those who survive can benefit from the results once the research is completed. *Figure 50* shows the expected benefits of research for EECP for a range of probabilities that research will be successfully undertaken and for different times for research to report. Again, the cost of research (considered to be £1.5M) can be compared directly with the potential benefits by the time that it takes for research to report.

How long research might take to report will depend in part on the design (follow-up, sample size and end points), recruitment rates and size of the eligible patient population, as well as on how efficient the organisation and data collection might be. The potential value of research for EECP over a range of possible time horizons is reported in *Figure 51* for different research designs (length of follow-up). The potential value of further research declines with the time to research reporting. The value of research is always higher for designs with longer follow-up as the research is able to resolve more of the uncertainty surrounding the approval of EECP; however, the differences are small in this case, but the longer the follow-up design the longer the time it takes for the research to report. This relationship gives some indication of the value of improving the timeliness of research through, for example, investment in research infrastructure or adopting a research design.

How much will be resolved?

How much of the uncertainty will be resolved depends on the type of research likely to be undertaken. In *Point 4: Is research possible with approval?* the potential benefits of different types of evidence were assessed. In the case of EECP, most benefit can be gained by resolving the uncertainty in the improvement in quality of life at 12 months. This would require a 1-year follow-up design, with research reporting after 1 year. Resolving more of the uncertainty would require a longer follow-up and hence longer follow-up designs. For example, a trial with a 2-year follow-up will provide information about improvements in quality of life at 12 months and about
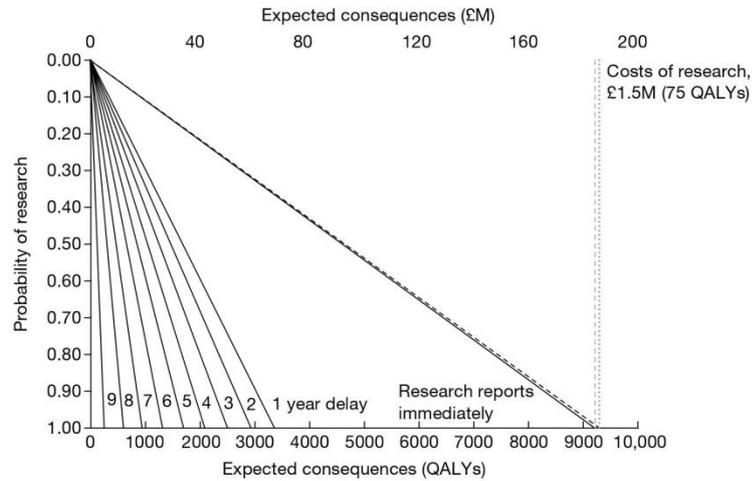
FIGURE 50 Expected potential benefits of research by time for research to report.
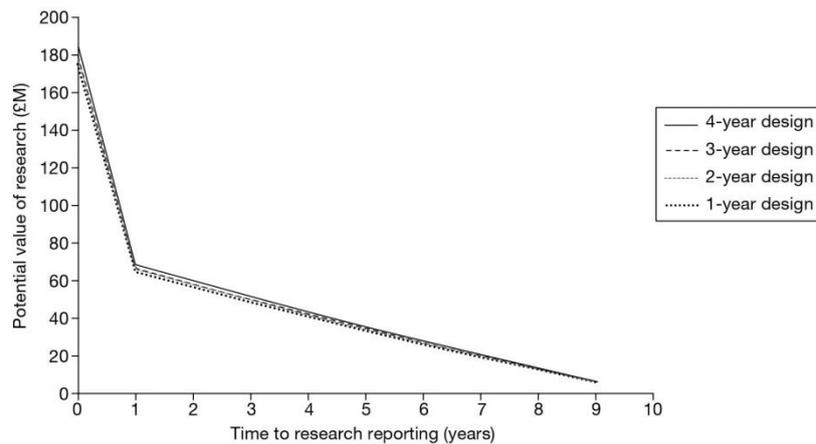
FIGURE 51 Potential benefits of research by time to report for different research designs.

the sustained duration of these improvements from 12 months to 2 years. Similarly, a trial with an even longer follow-up will provide information about the sustained duration in subsequent years. *Table 80* reports the potential benefits of alternative research designs of 1–4 years of follow-up. Although longer follow-up offers greater potential benefits, they are relatively small compared with the loss of potential value if longer follow-up delays the time until research findings are available, for example a 4-year design will require a minimum of 4 years to complete; however, as long as research reports before 8 years the potential benefits are likely to exceed the costs.

What is the impact of other sources of uncertainty?

In *Point 5: Will other sources of uncertainty resolve over time?*, no information was identified regarding other sources of uncertainty for EECP. A change in price for EECP is not anticipated and there is no evidence of new technologies emerging. However, given that the development and launch of new devices are more difficult to identify in advance, two alternative scenarios

TABLE 80 Potential benefits of research by time to report for different research design

Time until research reports, years	Potential benefits of research, QALYs (£M)			
	4-year design	3-year design	2-year design	1-year design
Immediately	9287 (186)	8987 (180)	8944 (179)	8756 (175)
1	3435 (69)	3324 (67)	3308 (66)	3238 (65)
2	2999 (60)	2902 (58)	2888 (58)	2827 (57)
3	2577 (52)	2494 (50)	2482 (50)	2430 (49)
4	2170 (43)	2100 (42)	2090 (42)	2046 (41)
5	1776 (36)	1719 (34)	1711 (34)	1675 (34)
6	1396 (28)	1351 (27)	1345 (27)	1316 (26)
7	1029 (21)	996 (20)	991 (20)	970 (19)
8	674 (14)	652 (13)	649 (13)	636 (13)
9	331 (7)	321 (6)	319 (6)	312 (6)
10	0 (0)	0 (0)	0 (0)	0 (0)

are considered in *Figure 52*. In scenario A the new technology enters at year 5 and makes EECP entirely obsolete, that is, not cost-effective and not uncertain – equivalent to a shorter technology horizon of 5 years. At this point there is no value in the evidence generated by research about EECP. The potential benefits of further research about EECP are likely to exceed the costs only if the research reports quickly. Nevertheless, even in this extreme scenario the benefits of research with only 1 year of follow-up are likely to exceed the costs as long as it reports before 4 years. In scenario B the new technology enters at year 5 and has similar NHEs to treatment with EECP and the uncertainty surrounding its expected cost-effectiveness is also similar. Now research about EECP has more potential value in the future because it will also help to resolve some of the uncertainty in the choice between EECP and the new technology.

The potential value of research presented in *Figure 52* should be regarded as an upper bound to the value that is likely to be realised by actual research for two reasons: (1) even well-designed research with large sample sizes will not fully resolve the uncertainty in the value that a parameter might take and (2) insofar as implementation of NICE guidance is not 'perfect' and all clinical practice might not immediately respond to the results of research, the full benefits will only be realised over time or with additional implementation efforts. For these reasons a judgement of whether or not benefits of research are likely to exceed the costs might be made conservatively, requiring evidence that, even in pessimistic scenarios, the research would still be worthwhile.

Types and categories of guidance resulting from points 5 and 6

The judgements made at points 5 and 6 of the checklist are critical because if the benefits of research are not judged to exceed the costs then no further assessments are required (unless there are significant irrecoverable costs, as in the case of EECP). For EECP, no other sources of uncertainty were identified at point 5 and the potential value of research is likely to exceed the costs. Therefore, the categories and types of guidance that could ultimately result from the sequence of assessments up to point 6 are AWR[4], OIR[4], Approve[9] and OIR[6] (*Table 81*).

Are the benefits of approval greater than the costs?

Point 7: Are the benefits of approval greater than the costs?

The seventh and final point on the checklist requires an assessment and comparison of the benefits and costs of early approval. The costs of approval include the potential value of any
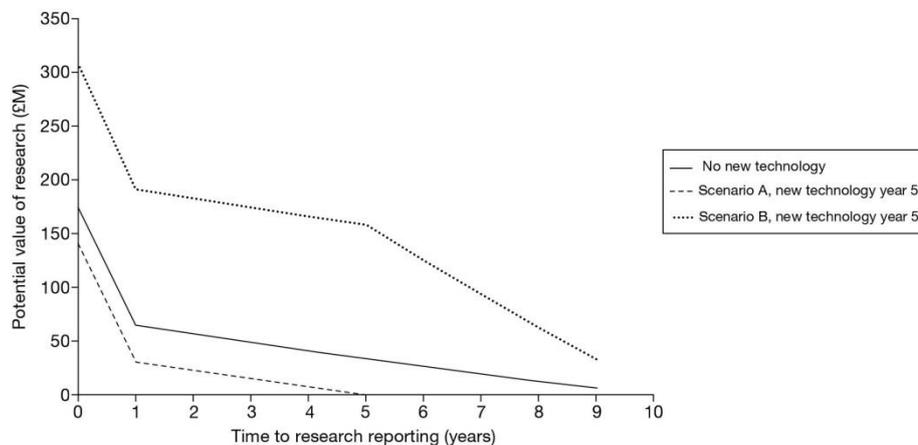
FIGURE 52 Potential value of research by time to report and other sources of uncertainty. These potential values are based on a 1-year follow-up design.

TABLE 81 Types and categories of guidance that could ultimately result from assessments 5 and 6

Assessment	1	2	3	4	5	6	7	Guidance
17	Yes	Yes	Yes	Yes	No	Yes	Yes	AWR[4]
18	Yes	Yes	Yes	Yes	No	Yes	No	OIR[4]
19	Yes	Yes	Yes	Yes	No	No	–	Approve[6]
24	Yes	Yes	Yes	No	No	Yes	Yes	Approve[9]
25	Yes	Yes	Yes	No	No	Yes	No	OIR[6]
26	Yes	Yes	Yes	No	No	No	–	Approve[10]

research that may be forgone as a consequence of approving EECP and any costs that are irrecoverably committed by approval. A judgement of whether or not the benefits of approval and early access for current patients are likely to exceed the opportunity costs for future patients is required, that is, at the following point in the algorithm:
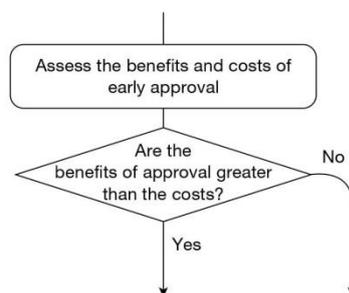
The decision at this point always leads directly to a particular category and type of guidance. For EECP, this could be AWR, OIR or approve (see *Table 81*).

Technologies with significant irrecoverable costs and research is possible with approval

Even when research is possible with approval, the impact of committing irrecoverable costs through AWR (or approval) must be considered, so OIR remains a possibility. EECP commits capital costs associated with long-lived equipment as well as initially negative per-patient NHEs. These irrecoverable opportunity costs at a patient level are significant because treatment choice for a presenting patient is not irreversible over relevant time frames (see *Point 2: Are there significant irrecoverable cost?*). As a consequence, even if research is possible with approval it is not clear that AWR would be appropriate, because OIR avoids the commitment of irrecoverable costs until research findings are available and a more informed decision can be made.

Figure 53 shows the expected NHEs of AWR and OIR over a range of possible time horizons for when the research recommended might report. OIR offers greater expected NHEs than AWR as long as research reports before 9 years. This is because the consequences (losses of population NHEs) of committing both aspects of irrecoverable costs through AWR are greater than the NHEs forgone by restricting access to EECP through OIR. The costs of research have not been included because they are incurred with both AWR and OIR guidance.[f]

However, there is no guarantee that the research recommended as part of OIR or AWR guidance will be conducted by manufacturers or commissioned by research funders. Even if it is, it is not certain that it will be successfully completed. Therefore, the probability that research will report at a particular time also needs to be considered. The implications of considering whether the recommended research will be conducted and when it might report are illustrated in *Figure 54*, which presents a boundary for when OIR rather than AWR might be appropriate. The boundary is illustrated for four research designs with differing lengths of follow-up. Designs with a shorter follow-up might be preferred if the additional benefits of a longer follow-up are less than the additional opportunity costs of waiting longer until the research reports. In this case it seems likely that the type of research required could report quickly enough for all research designs and with sufficient confidence that OIR would be appropriate, even though the research could be conducted while EECP is approved. Therefore, these assessments would support a judgement that the benefits of approval (through AWR) are unlikely to exceed the opportunity costs (the NHEs of OIR) and so OIR[4] (pathway 18, see *Table 81*) rather than AWR[4] (pathway 17, see *Table 81*) would be more appropriate.
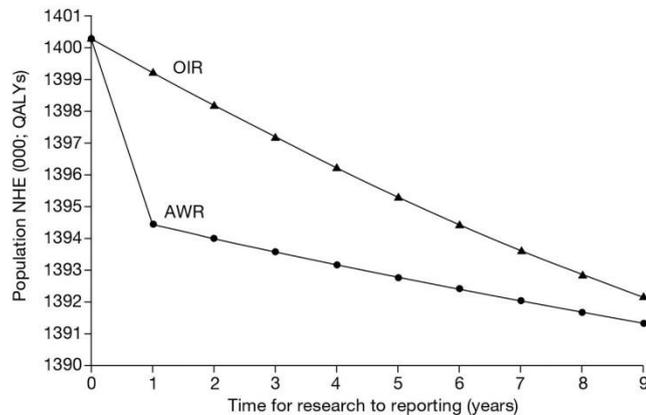

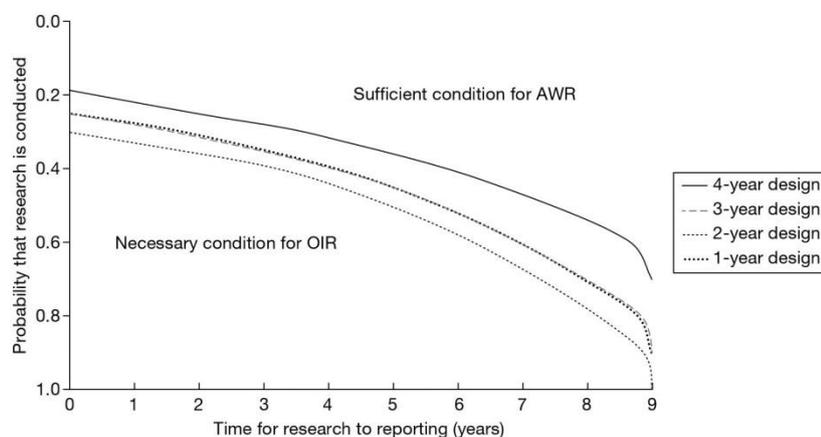FIGURE 53 Population NHEs of AWR and OIR by time to research reporting.

FIGURE 54 An OIR or AWR boundary.

TABLE 82 Population NHEs over the technology time horizon for different policies

NHE for time to research reporting	Approve	OIR	AWR	Reject	Value of AWR	Uncertainty resolved at launch	Value of evidence at launch
NHEs in QALYs							
T = 3	1,391,001	1,397,192	1,393,578	1,389,596	−3614	1,400,288	3096
T = 7	1,391,001	1,393,608	1,392,030	1,389,596	−1578	1,400,288	6680
NHEs in £M							
T = 3	27,820	27,944	27,872	27,792	−72	28,006	62
T = 7	27,820	27,872	27,841	27,792	−32	28,006	134

T, time to research reporting.

Table 82 summarises the population NHEs over the technology time horizon for different policies for a 4-year design. This can be used to help inform the policy questions: (1) what would be the value of being able to conduct research while EECP is approved? and (2) what would be the value of making the evidence that is needed by the NHS available at launch? In this case, because of the irrecoverable costs associated with EECP, there is no value to the NHS of being able to conduct research while EECP is approved for use. In fact, these figures are negative, indicating that even if AWR was possible it would not be appropriate. The difference in NHEs between OIR and AWR (the next best policy, if it were possible) represents the value to the NHS of withholding approval until the research is conducted (i.e. £72M if 3 years). The difference in population NHEs between all uncertainty resolved prior to appraisal (at launch) and the next best available policy represents the value to the NHS of having access to the evidence needed at launch. The value, expressed in the equivalent NHS resources, depends on how long it would otherwise have taken for an OIR recommendation to deliver the same evidence, for example £62M if 3 years and £134M if 7 years.

Technologies with significant irrecoverable costs and research is not possible with approval

For the reasons discussed earlier, the type of experimental research required to robustly estimate the effect of EECP on quality of life is unlikely to be possible once EECP is approved and in

widespread use; therefore, research may not be possible with approval. In this case, approve (rather than AWR) is an alternative option. However, approve not only commits the type of irrecoverable costs discussed above but also means that the potential value of evidence to future patients must also be forgone. This is reflected in *Figure 55* in which the difference between OIR and approve is always greater than that between OIR and AWR in *Figure 53*. It suggests that as long as the cost of the research exceeds the difference between OIR and approve, when it is expected to report, OIR rather than approve would be appropriate. This is also reflected in the boundaries for OIR and approve reported in *Figure 56*. The boundaries for the different research designs are always to the north-east of the OIR/AWR boundaries reported in *Figure 54*, again reflecting the fact that approval not only commits irrecoverable costs but also forgoes the potential value of evidence that might have been generated through an OIR recommendation. These assessments would support a judgement that the benefits of approval are unlikely to exceed the opportunity costs (the NHEs of OIR) and so 'OIR[6]' (pathway 25, see *Table 81*) rather than 'Approve[9]' (pathway 24, see *Table 81*) would be more appropriate.
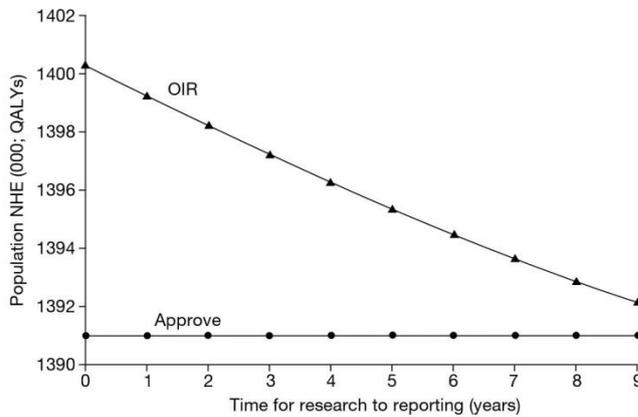
FIGURE 55 Population NHEs of approve and OIR by time to research reporting.
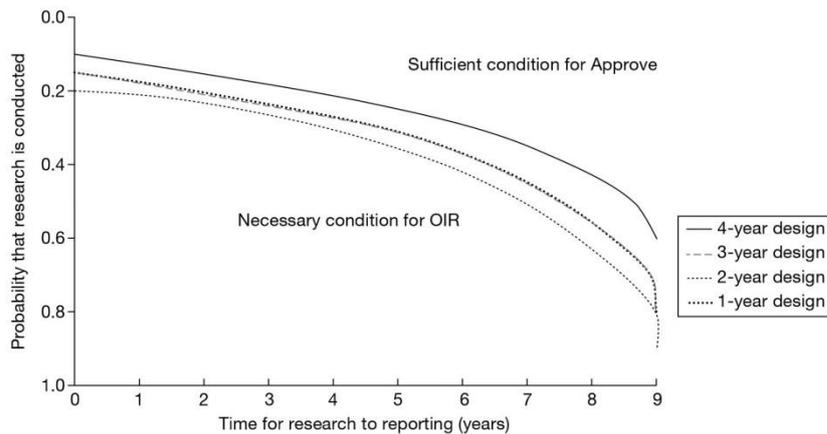
FIGURE 56 An OIR or approve boundary.

284

Types and categories of guidance resulting from point 7

The decision at this point always leads directly to guidance, allocating all remaining possible pathways to a particular type and category of guidance. The resulting guidance for EECP is summarised in *Table 83*. Guidance will depend on whether or not EECP can be approved for widespread use while research is conducted (AWR); however, as demonstrated above, even when research is possible with approval, OIR appears to offer greater expected NHEs than AWR or approve as long as research reports before 9 years. This is largely because the consequences of committing significant irrecoverable costs through AWR or approve are greater than the NHEs forgone by restricting access to EECP through OIR.

TABLE 83 Types and categories of guidance for EECP

Assessment	1	2	3	4	5	6	7	Guidance
17	Yes	Yes	Yes	Yes	No	Yes	Yes	AWR[4]
18	Yes	Yes	Yes	Yes	No	Yes	No	OIR[4]
24	Yes	Yes	Yes	No	No	Yes	Yes	Approve[9]
25	Yes	Yes	Yes	No	No	Yes	No	OIR[6]

Appendix 8

Clopidogrel for the management of patients with non-ST segment elevation acute coronary syndromes

Introduction

The use of CLOP (for up to 12 months) in combination with low-dose aspirin was recommended by NICE following a MTA appraisal for patients with NSTE-ACS presenting with a moderate to high risk of ischaemic events (TA80[122] in 2004). In TA80 the Appraisal Committee considered 12 months' or lifetime treatment with CLOP, but recommended research to inform optimal treatment duration. The original report had included an analysis of shorter treatment durations (< 12 months) and the NIHR HTA programme subsequently commissioned additional reanalysis based on the original work to inform the research recommendation in 2009. This case study is based on the reanalysis of TA80 undertaken in 2009, which included standard therapy compared with four alternative treatment durations of CLOP of 1, 3, 6 and 12 months. This analysis informed clinical guideline CG94[123] in 2010. Importantly, although this case study is based on the later reanalysis of TA80, the analysis considered here has been undertaken from the standpoint of the original TA80 appraisal and asks what assessments might have been made at that time when standard therapy was low-dose aspirin.

Background to the case study

An acute coronary syndrome is a set of symptoms of myocardial ischaemia occurring because of the presence of occlusion thrombi in the arteries because of the fissuring or rupturing of atheromatous plaques. NSTE-ACS can be classed as either unstable angina or non-ST elevation myocardial infarction (NSTEMI). NSTE-ACS are associated with a high risk of death or ischaemic complications, and antiplatelet agents such as CLOP are effective at reducing the risk of further ischaemic events.

In recommending CLOP in TA80,[122] NICE used a published report of analyses undertaken by the Centre for Reviews and Dissemination (CRD) and the Centre for Health Economics (CHE).[175] This evaluation assessed the clinical effectiveness and cost-effectiveness of CLOP in combination with aspirin for people with NSTE-ACS, the results of which were later updated by Rogowski et al.[148] in 2009. The optimal duration of CLOP treatment was assessed at this later stage and an analysis of the value of further research was undertaken. The update of the MTA appraisal was used to inform NICE's clinical guidelines (CG94[123]) in 2010.

The model and methods used in the evaluation by Rogowski et al.[148] were used here to illustrate the information and associated analyses that would have been required to inform a decision process linking adoption with further evidence collection. Despite using the model and methods of Rogowski et al.,[148] the analysis has been undertaken from the standpoint of the original TA80[122] appraisal. Given the objectives of exploring what *additional* information and analysis might be

required, de novo or substantial reanalysis of original assessments was not undertaken in the current work.

Interventions and population

The population of interest includes patients with NSTE-ACS presenting with a moderate to high risk of ischaemic events. Treatment should start immediately after the acute event; thus, prevalent cases are not assumed eligible for treatment, only incident cases. The incidence of NSTE-ACS in the UK was estimated to be 60,000 patients per year,[148] reflecting the size of the population eligible to receive treatment annually. This figure is assumed to be non-stochastic in further analyses.

The treatment strategies considered in this case study were:

- CLOP12: treatment with CLOP as an adjunct to standard therapy for 12 months
- CLOP6: treatment with CLOP as an adjunct to standard therapy for 6 months
- CLOP3: treatment with CLOP as an adjunct to standard therapy for 3 months
- CLOP1: treatment with CLOP as an adjunct to standard therapy for 1 month
- NHS: lifetime treatment with standard therapy alone (including aspirin).

Lifetime treatment with standard therapy is considered here as the treatment established in the NHS, as we aim to recreate the decision problem set out in 2004 (TA80[122]). A possible decision to use CLOP will thus represent a change in guidance. Given that the evaluation undertaken was used to inform NICE's decision-making, this study used the NHS perspective and a discount rate of 3.5% for future costs and QALYs.[1]

Evidence on clinical effectiveness

The Clopidogrel in Unstable Angina to Prevent Recurrent Events (CURE) trial established the effectiveness of CLOP in patients with NSTE-ACS.[176–181] This was a multicentre, double-blind RCT that recruited 12,562 patients who presented within 24 hours of experiencing an NSTE-ACS event. The trial compared CLOP (300 mg initially followed by 75 mg daily) in combination with aspirin (75–325 mg/day) with placebo plus aspirin (75–325 mg/day). Results show that CLOP was significantly more effective than placebo at reducing the risk of the composite outcome of cardiovascular death, NFMI or stroke. The proportion of patients experiencing the composite outcome was greater in the aspirin-only group at 30 days [relative risk (RR) 0.79; 95% confidence interval (CI) 0.67 to 0.92] and from 30 days to 12 months (RR 0.82; 95% CI 0.70 to 0.95). The results also suggested that the benefits of CLOP may be most apparent within the first 3 months of treatment. This was investigated further using an exploratory post hoc analysis of the CURE data (comprising non-randomised comparisons), presented in the SIGN guidelines (no. 93).[182] This analysis showed that, within the CURE trial, the event rate was statistically significantly lower in the CLOP group for the periods 0–1 month, 1–3 months and 0–12 months, but not for the periods 3–6 months, 6–9 months and 9–12 months.

Decision model

The model used by Rogowski *et al.*[148] combines a short-term decision tree and a Markov model representing the longer term. The short-term tree characterises the period up to 12 months following an acute coronary syndrome, tracking for possible myocardial infarctions (MIs) or death. The Markov model was defined to have four states (well, death, MI and post MI). The model was run for a period of 40 years, the 'patient time horizon', using yearly cycles. The patient time horizon is the time horizon over which costs and benefits are likely to differ for an individual patient and was here assumed to approximate the lifetime of the patients. A summary of the parameters used to define the decision model is shown in *Table 84*.

TABLE 84 Parameters of the decision model for CLOP

		Parameter	Description	Source	Distribution
Natural history	1	P_die_0.1	Short term: probability of death, 0–1 months	PRAIS-UK data[11]	Dirichlet
	2	P_NFMI_0.1	Short term: probability of a NFMI, 0–1 months		
	3	P_die_1.3	Short term: probability of death, 1–3 months		Dirichlet
	4	P_NFMI_1.3	Short term: probability of a NFMI, 1–3 months		
	5	P_die_3.6	Short term: probability of death, 3–6 months		Dirichlet
	6	P_NFMI_3.6	Short term: probability of a NFMI, 3–6 months		
	7	P_die_6.12	Short term: probability of death, 6–12 months		Dirichlet
	8	P_NFMI_6.12	Short term: probability of a NFMI, 6–12 months		
	9	TP_AC	Long term: annual TP – well(A) to MI(C)	Main *et al.* 2004[175]	Normal (log-hazard)
	10	TP_AD	Long term: annual TP – MI(C) to dead(D)		Normal (log-hazard)
	11	TP_CD	Long term: annual TP – MI(C) to dead(D)		Normal (log-hazard)
	12	TP_BD	Long term: annual TP – post MI(B) to dead(D)		Normal (log-hazard)
Relative effects	17	RR_death	Treatment effects (RR) – all-cause mortality	CURE trial[176–181]	Normal (log-RR)
	18	RR_NFMI	Treatment effects (RR) – NFMI		Normal (log-RR)
Utilities	13	U_Well	Utility weights per health state: IHD year 1	Karnon *et al.* 2006[183]	Beta
	14	U_Well1	Utility weights per health state: post IHD		Beta
	15	U_NFMI	Utility weights per health state: MI year 1		Beta
	16	U_POSTMI	Utility weights per health state: post MI		Beta
Costs	19	C_Well	Cost per health state: IHD year 1	a	a
	20	C_MI_LT	Cost per health state: MI year 1	a	a
	21	C_PostMI	Cost per health state: post MI	a	a
	22	TC_Well_Dead	Cost per health state: transitions to death	a	a
	23 to 27	C_clop12, …, C_NHS	Cost per treatment strategy	a	a
	28	Clop_cost	Cost per mg of CLOP (2100 mg, 2007)	BNF 2007[184]	Constant

C, cost; IHD, ischaemic heart disease; P, probability of events in the short-term decision tree; PRAIS-UK, Prospective Registry of Acute Ischaemic Syndrome in the UK; RR, relative risk; TP, annual transition probabilities between states [well (A), post MI (B), MI (C), dead (D)] of a Markov model characterising long-term progression of disease; U, utility.

a   To reduce the number of model parameters and allow tractability in the further analysis, resource use and cost parameters from the original model[148] were used to derive costs per health state and costs per treatment. The distribution assumed to describe uncertainty over these parameters was an empirical distribution, evaluated using the original Monte Carlo simulations.

The probabilities of death and NFMI applied in the first year were derived from the Prospective Registry of Acute Ischaemic Syndrome in the UK (PRAIS-UK), an observational UK cohort registry of 1046 patients with ACS.[185] Treatment with CLOP was assumed to prevent death and NFMI and this was parameterised using relative treatment effects. These were applied throughout the duration of the CLOP treatment period (in the short-term model), within which these were constant; hence, any benefits of treatment with CLOP were assumed to stop at the time of withdrawal, and patients were modelled to rebound to the same prognosis as an equivalent patient on aspirin alone (NHS standard care).

The long-term model was used to quantify the remaining QALYs and costs of patients once they exited the short-term model. The rate at which transitions happen within the long-term Markov model were parameterised using transition probabilities. These transitions were estimated using the Nottingham Heart Attack Register (NHAR) data ($n = 1279$) (for further details see Rogowski *et al.*[148]) and were defined as independent of treatment; hence, CLOP was assumed to not impact directly on the long-term outcomes of patients.

Utility parameters were derived using published evidence from Karnon *et al.*[183] Costs and resource use categories considered were treatment related, adverse event related, those associated with health states and those associated with possible revascularisation (tracked in the short-term model). The specificities of how these were considered are detailed in the main report.[99] In the current assessment, to allow tractability in the further analysis and clarity in the presentation of results, the number of model parameters was reduced by using the resource use and cost parameters in the original model to derive costs per health state and costs per treatment.

Inputs in the model were assumed to be uncertain and the model was run probabilistically using Monte Carlo simulation (5000 simulations). Expected cost-effectiveness was determined using these probabilistic results, as established within the NICE appraisal process.[a] The use of probabilistic analysis also allows uncertainty over model parameters to be translated into uncertainty in the overall results.

Key features and possible pathways

The CLOP case study focuses on evaluating alternative durations of treatment with CLOP alongside lifetime treatment with standard therapy as well as standard therapy alone (current NHS care) for the management of patients with NSTE-ACS. This is an acute condition and so only incident populations after research reports can benefit from the information.

A research recommendation was made in section 5 of the FAD in TA80[122] regarding uncertainty over the duration of treatment; therefore, CLOP is not an example of AWR at FAD but an example of AWR considered during appraisal. In an update of the assessments undertaken,[99] alternative assumptions about the way that treatment effects were modelled were also evaluated.

Using the CLOP case study we aim to demonstrate how the key principles and assessments could inform the development of guidance regarding the use of CLOP through application of the checklist developed in this project. We also aim to illustrate how existing methods of appraisal can be applied to this case study and also what additional information and analysis (and when) can be useful for the Appraisal Committee to undertake the proposed judgements.

The possible pathways through the algorithm that the CLOP case study illustrates are reported in *Figure 29* in *Appendix 4*, with the new technology expected to be cost-effective and with no significant irrecoverable costs. The CLOP case study also illustrates a number of other important characteristics, including (1) the impact that other sources of uncertainty (price change following patent expiry) can have on the value of further research, (2) the interpretation of analyses when there are multiple alternatives and (3) the use of scenarios to represent alternative but credible assumptions. The following sections examine each of the seven points on the checklist relating to the possible sequence of assessments and decisions, which lead to a particular category and type of guidance for CLOP.

Is it cost-effective and what are the risks?

Point 1: Is it expected to be cost-effective?

The sequence of assessments starts with cost-effectiveness and the expected impact on population NHEs, that is, at the following point in the algorithm:
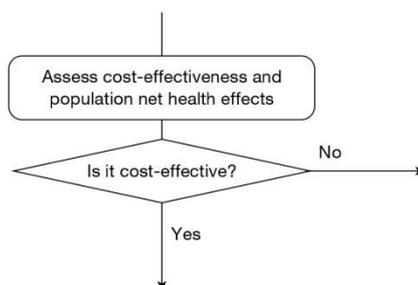
The assessment of expected cost-effectiveness is made based on the balance of the evidence and analysis currently available. Commonly, expected cost-effectiveness is summarised and presented using ICERs. Equivalently, but more usefully in this context, cost-effectiveness can be expressed in terms of expected NHEs, which can be expressed per patient treated or for a population of patients. All of the information required to express expected cost-effectiveness in these ways is already available during appraisal.

Cost-effectiveness at the patient level

Estimates of the expected NHS costs and QALYs for each patient treated over an appropriate time horizon – the 'patient time horizon' – can be summarised as the per patient NHEs of each intervention, that is, the difference between any health gained and health forgone elsewhere.
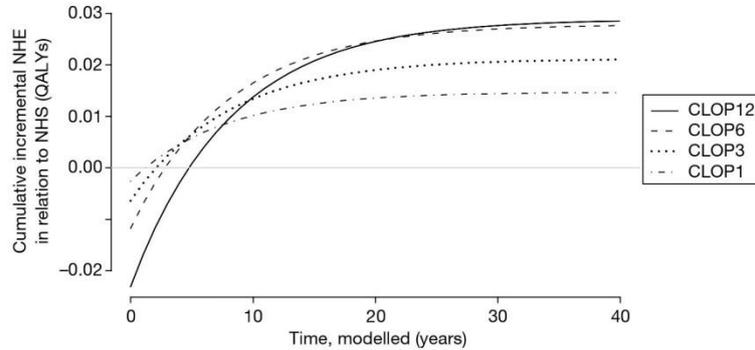
The results for CLOP are summarised in *Table 85*.[b] There are more than two alternatives: four treatment durations as well as current NHS treatment (aspirin alone) were considered. The results indicate that 12-month treatment with CLOP is expected to be cost-effective at a threshold of £20,000 per QALY,[c] although the difference in NHEs between 12 months and 6 months of treatment is small. Consequently, the NHEs of 12 months' treatment with CLOP are greater than the NHEs of 6 months' treatment with CLOP but the difference per patient treated (the incremental NHE) is small.

It is also important to consider how NHEs accumulate over time or the investment profile per patient treated with CLOP. *Figure 57* illustrates the cumulative incremental NHEs over the patient time horizon. The per patient costs of CLOP are in excess of the health benefits during the period of treatment. These negative NHEs are eventually offset by positive NHEs in later periods. In this case, it is only after 5 years that 12 months of treatment with CLOP breaks even against current NHS care and it is not until 21 years that it is better than a shorter treatment duration of 6 months. Notice that shorter treatment durations with CLOP offer a much less 'risky profile', for example the break-even point for 1 month of treatment is 2 years against current NHS care.

TABLE 85 Expected cost-effectiveness of CLOP per patient treated

Treatment	Cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY NHE, QALYs (£)	Cost-effectiveness threshold £30,000 per QALY NHE, QALYs (£)
CLOP12	20,127	8.122	18,663	7.115 (142,307)	7.451 (223,525)
CLOP6	19,860	8.107	10,477	7.114 (142,288)	7.445 (223,362)
CLOP3	19,712	8.093	9396	7.108 (142,154)	7.436 (223,087)
CLOP1	19,598	8.081	4961	7.101 (142,025)	7.428 (222,837)
NHS	19,502	8.062	–	7.087 (141,734)	7.412 (222,353)

FIGURE 57 Cumulative incremental NHEs of CLOP over the patient time horizon.

Cost-effectiveness at the population level

Per patient NHEs can also be expressed for the population of current and future patients. This requires information about prevalence and future incidence of the target population (already required in appraisal). It also requires a judgement about the time horizon over which the technology will be used. This 'technology time horizon' ought to reflect the period over which the technology is likely to be part of clinical practice and generate the expected NHEs.[d] An estimate of the scale of the total population NHEs and how they cumulate over time is important for subsequent assessments, including (1) when the NHEs for current patient populations must be compared with the benefits to future patients and (2) when the treatment decision can be changed so the irrecoverable costs of initially negative NHEs become significant.

Given the acute nature of the condition being assessed in the CLOP case study, it is assumed that the prevalent population is not significant and that only incident populations are eligible for treatment with CLOP. The total population NHEs, assuming that the technology will be used to treat incident patients over 10 years, are reported in *Table 86*. Expected cost-effectiveness is unchanged (ICERs are the same as in *Table 85*) but the incremental NHEs, although small per patient, are more significant at a population level.

The investment profile for CLOP when used to treat patients over 10 years is illustrated in *Figure 58*. At a population level it is not until 11 years (rather than 5 years at a patient level) that initial losses are compensated for by later gains and CLOP (12 months of treatment) breaks even against current NHS care. It is not until 27 years (rather than 21 years at a patient level) that 12 months of treatment is better than a shorter treatment duration of 6 months. In other words, CLOP appears a more risky investment when evaluated at a population rather than an individual level. This is because, although each patient treated with CLOP is expected to offer the same profile of NHEs shown in *Figure 57*, the negative NHEs associated with patients who are incident and treated in year 10 will not be offset by later gains.

Table 87 shows that the population-level investment profile would exhibit greater risk (break-even later) if the technology time horizon was longer. For example, the break-even point for 12 months of treatment with CLOP against 6 months of treatment extends to 33 years when the technology time horizon is increased to 20 years. Shorter durations of treatment still offer a 'risky profile', for example the break-even point for 1 month of treatment compared with current NHS care is 4 years.

TABLE 86 Expected cost-effectiveness of CLOP for the population

Treatment	Cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY NHE, QALYs (£M)	Cost-effectiveness threshold £30,000 per QALY NHE, QALYs (£M)
CLOP12	10,394,830,647	4,194,554	18,663	3,674,813 (73,496)	3,848,060 (115,442)
CLOP6	10,256,672,674	4,187,151	10,477	3,674,318 (73,486)	3,845,262 (115,358)
CLOP3	10,180,425,730	4,179,874	9396	3,670,853 (73,417)	3,840,526 (115,216)
CLOP1	10,121,529,942	4,173,605	4961	3,667,529 (73,351)	3,836,221 (115,087)
NHS	10,072,035,344	4,163,629	–	3,660,027 (73,201)	3,827,894 (114,837)
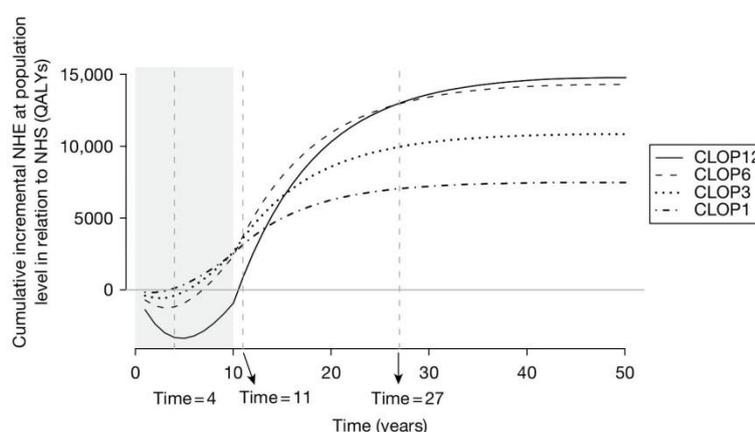
FIGURE 58 Cumulative incremental NHEs of CLOP for the population. Incremental NHEs are presented for the different treatment strategies in relation to standard NHS care. A 10-year technology time horizon is assumed, indicated by the grey shading.

Point 2: Are there significant irrecoverable costs?

The second point on the checklist requires (1) an assessment of whether there are irrecoverable costs and (2) a judgement of their potential significance, that is, at the following point in the algorithm:
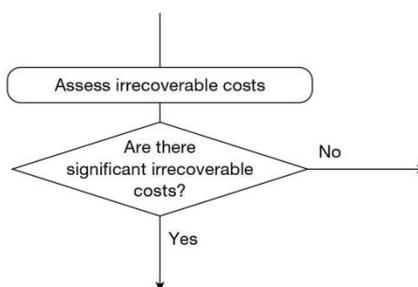
Irrecoverable costs are those that, once committed, cannot be recovered if guidance is changed at a later date. Irrecoverable costs are most commonly thought of as 'upfront' or capital costs of new facilities or equipment with long life expectancy (they might also include any practitioner training and the costs of implementation efforts). In the CLOP case study, no capital costs were included in the analysis.

TABLE 87 Expected cost-effectiveness of CLOP for the population for a range of technology time horizons

Technology time horizon	Treatment	Incremental NHE, QALYs (£M)	Break-even point (years)		
			12 months vs 6 months	12 months vs NHS	1 month vs NHS
5 years	CLOP12	269 (5.4)	24	8	4
	CLOP6	1881 (37.6)			
	CLOP3	1804 (36.1)			
	CLOP1	4073 (81.5)			
	NHS	–			
10 years	CLOP12	495 (9.9)	27	11	4
	CLOP6	3465 (69.3)			
	CLOP3	3324 (66.5)			
	CLOP1	7502 (150)			
	NHS	–			
15 years	CLOP12	686 (13.7)	30	12	4
	CLOP6	4799 (96)			
	CLOP3	4603 (92.1)			
	CLOP1	10,389 (207.8)			
	NHS	–			
20 years	CLOP12	846 (16.9)	33	12	4
	CLOP6	5921 (118.4)			
	CLOP3	5680 (113.6)			
	CLOP1	12,820 (256.4)			
	NHS	–			

However, even in the absence of capital costs of equipment and facilities, NHEs accumulate over time both at a patient and a population level. The analysis in *Point 1: Is it expected to be cost-effective?* (see *Figure 58*) indicates a common pattern of initially negative NHEs that are only gradually offset by positive NHEs in later periods. The investment profile of CLOP (at a patient and more so at a population level) thus exhibits irrecoverable costs. Therefore, in the case of an approval or AWR recommendation, opportunity costs of negative NHEs are committed that may be irrecoverable.

Are they likely to be significant?

Whether or not irrecoverable costs are significant (i.e. might influence guidance) depends critically on whether or not guidance is likely to change and whether that is more likely in the near or distant future. That will depend on whether or not research is likely to be undertaken and when it is likely to report, as well as on other events that might occur, for example a change in price following patent expiry. These are assessed later, at points 5 and 6 on the checklist. However, the *potential* significance of any irrecoverable costs can be assessed at this point.

Judging the potential significance of the investment profiles of NHEs is more nuanced. It depends whether treatment decisions for individual patients are irreversible, which in part depends on the nature of the disease. CLOP is a treatment for ACS and, although decisions about treatment and its duration are not irreversible in the short run, over the time scales more likely for research being conducted (and reporting) or other events occurring that could change guidance they can be regarded as such. Of course, it is possible that the later benefits are not realised but it is also possible that they will realise more (the profiles of NHEs in *Figure 57* are the average over these possibilities). Similarly, the possibility that guidance might change in the future (e.g. research suggests that the longer-term benefits will not offset initial losses) will not influence

the irreversible decision to treat a presenting patient with a technology that is expected to be cost-effective prior to the research reporting. Therefore, although the investment profile of CLOP exhibits irrecoverable costs these should not be judged significant in the sense that they have little potential to influence guidance.

Types and categories of guidance resulting from points 1 and 2

Points 1 and 2 of the checklist do not lead directly to a category or type of guidance. The sequence of assessments and decisions, which ultimately lead to guidance, starts with cost-effectiveness, expected impact on population NHEs and significance of irrecoverable costs. In the case of CLOP, the technology is expected to be cost-effective and it does not have significant irrecoverable costs. Although treatment with CLOP does commit initially negative NHEs that are irrecoverable, these should not be regarded as significant as the treatment decision for a presenting patient is irreversible in relevant time frames.

Table 88 summarises the categories and types of guidance that could ultimately result from the sequence of assessments up to point 2 (see Table 32 in Appendix 4).

Is further research required?

Point 3: Does more research seem worthwhile?

The third point on the checklist requires an assessment of the potential benefits of conducting further research, that is, at the following point in the algorithm:

This requires judgements about (1) how uncertain a decision to approve or reject a technology might be based on the estimates of expected cost-effectiveness and (2) whether the scale of the likely consequences of this uncertainty might justify further research. Some assessment of the potential consequences of uncertainty is important because it indicates the scale of the population NHEs, over the technology time horizon, that could be gained if the uncertainty

TABLE 88 Types and categories of guidance that could ultimately result from points 1 and 2

Assessment	1	2	3	4	5	6	7	Guidance
1	Yes	No	Yes	Yes	Yes/no	Yes	–	AWR[1]
2	Yes	No	Yes	Yes	Yes/no	No	–	Approve[1]
3	Yes	No	Yes	No	Yes/no	Yes	Yes	Approve[2]
4	Yes	No	Yes	No	Yes/no	Yes	No	OIR[1]
5	Yes	No	Yes	No	Yes/no	No	–	Approve[3]
6	Yes	No	No	–	–	–	–	Approve[4]

surrounding this decision could be resolved immediately, that is, it represents an expected upper bound on the benefits of more research.[e]

Assessing the consequences of uncertainty

In the CLOP case study 12 months of treatment is expected to be cost-effective (see *Tables 85 and 86*), but the estimates of costs and QALYs are uncertain so there is a chance that a decision to approve CLOP based on existing evidence will be incorrect, that is, other treatment options might offer greater NHEs. A judgement is required about the chance that 12 months of treatment is incorrect and if so which of the other four alternatives is likely to offer higher NHEs and how much higher. In other words, for decisions involving multiple alternatives, a judgement is required on the level of uncertainty surrounding the decision, how this uncertainty is distributed across the various alternatives and what the consequences are likely to be.

Some assessment of the likely consequences of approving CLOP under uncertainty could be based on the difference in expected NHEs, that is, the expected incremental population NHEs reported in *Table 87*. The simplest approach could be to take the difference in expected NHEs for CLOP and standard NHS care and weight it by a judgement of the probability that the decision is correct, that is, allocate the consequences of an incorrect decision to NHS standard care.[f] For example, if the decision was judged to be 100% certain (the probability that the decision is correct would be 1) then there are no consequences and so there would be nothing to be gained by more research; however, as the decision becomes more uncertain, the expected consequences (and hence potential value of more research) increases.

In the presence of multiple treatment alternatives, there are other options than to allocate all of the consequences of an erroneous decision to NHS standard care. We have examined four alternative ways:

1. The consequences are assigned to NHS standard care.
2. The consequences are assigned to CLOP6, the next best treatment.
3. Equal shares of the consequences are assigned to treatments other than CLOP12.
4. The consequences are assigned to treatments other than CLOP12 based on the probability of the alternative treatments being cost-effective from PSA. Note that this uses the PSA only to inform the magnitude of the consequences of a wrong decision; a judgement on the likelihood of adopting CLOP being the incorrect decision is still required.

The results of these alternative analyses are illustrated in *Table 89* in which a judgement about the probability that a decision based on expected cost-effectiveness is correct translates into expected consequences based on expected incremental population NHEs.

This judgement, of how uncertain a decision might be, can be informed by the PSA already used to estimate costs and QALYs and required as part of the NICE reference case. The probabilities that each of the five alternatives is cost-effective are reported in *Table 90*. The probability that 12 months of treatment is cost-effective is 0.524, which would translate into 1723 QALYs over the technology time horizon (assumed to be 10 years) based on the expected or average difference between NHEs (*Figure 59*). The time horizon over which evidence generated by research about a technology might be valuable may be longer (or shorter) than the period over which the technology is used; therefore, the technology time horizon and the time horizon for the benefits of research may differ. For simplicity, in this case study we have assumed these to be equal.

The previous analyses assumed that the difference in NHEs when CLOP is not the correct decision is the average; however, this is not necessarily true and estimates based on means may substantially under- or overestimate the expected consequences of uncertainty (see *Appendix 11,*

TABLE 89 Expected consequences of uncertainty based on a weighting of average NHEs

Probability decision is correct, p	Cost-effectiveness threshold £20,000 per QALY				Cost-effectiveness threshold £30,000 per QALY			
	Expected consequences, QALYs (£M)				Expected consequences, QALY (£M)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
0.99	148 (3)	5 (0.1)	66 (1.3)	37 (0.7)	202 (4.0)	28 (0.6)	106 (2.1)	56 (1.1)
0.95	739 (14.8)	25 (0.5)	332 (6.6)	185 (3.7)	1008 20.2)	140 (2.8)	529 (10.6)	281 (5.6)
0.90	1479 (29.6)	50 (1.0)	663 (13.3)	370 (7.4)	2017 40.3)	280 (5.6)	1058 (21.2)	561 (11.2)
0.75	3696 (73.9)	124 (2.5)	1658 (33.2)	925 (18.5)	5041 (100.8)	699 (14)	2646 (52.9)	1403 (28.1)
0.50	7393 (147.9)	248 (5.0)	3316 (66.3)	1850 (37)	10,083 (201.7)	1399 (28)	5292 (105.8)	2806 (56.1)
0.25	11,089 (221.8)	371 (7.4)	4973 (99.5)	2776 (55.5)	15,124 (302.5)	2098 (42)	7938 (158.8)	4209 (84.2)
0.10	13,307 (266.1)	446 (8.9)	5968 (119.4)	3331 (66.6)	18,149 (363)	2518 (50.4)	9525 (190.5)	5050 (101)
0.05	14,046 (280.9)	470 (9.4)	6299 (126)	3516 (70.3)	19,157 (383.1)	2658 (53.2)	10,055 (201.1)	5331 (106.6)
0.01	14,638 (292.8)	490 (9.8)	6565 (131.3)	3664 (73.3)	19,964 (399.3)	2770 (55.4)	10,478 (209.6)	5555 (111.1)
0	14,786 (295.7)	495 (9.9)	6631 (132.6)	3701 (74)	20,165 (403.3)	2798 (56)	10,584 (211.7)	5611 (112.2)

TABLE 90 Expected consequences of uncertainty for CLOP

Treatment	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY			Cost-effectiveness threshold £30,000 per QALY		
		Incremental NHE,[a] QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)	Incremental NHE,[a] QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
CLOP12	18,663	495 (9.9)	0.524	5194 (103.9)	2798 (56.0)	0.677	3657 (109.7)
CLOP6	10,477	3465 (69.3)	0.180		4736 (94.7)	0.092	
CLOP3	9396	3324 (66.5)	0.018		4305 (86.1)	0.009	
CLOP1	4961	7502 (150.0)	0.075		8327 (166.5)	0.052	
NHS	–	–	0.202		–	0.170	

a The mean additional population NHE of moving from the least to the most effective alternative, that is, the incremental NHE of 12 months' treatment compared with NHS care is the sum of these increments (14,786 QALY or £295.7M at £20,000 per QALY).

Why the consequences of uncertainty differ from mean incremental effects). To accurately record the difference between the NHEs of CLOP and alternative treatments, and the frequency of such errors, the probabilistic analysis can be used. This distribution of consequences from PSA is illustrated in *Figure 60*. Most commonly (52.4%) there are no consequences because 12 months of treatment with CLOP is the correct decision. When it is not, there is a greater chance of relatively small consequences (30% are < 10,000 QALYs), which occur predominantly when 6 months of treatment offers the highest NHEs. However, there is a small chance of larger consequences (< 5% chance that they are > 30,000 QALYs) when standard NHS treatment offers the highest NHEs, that is, there remains important uncertainty about the cost-effectiveness of treatment itself, not just its duration.

The average over this distribution provides the expected consequences of uncertainty from PSA, which in this case is 5194 QALYs (see *Table 90*). This is substantially greater than the estimate of 1723 QALYs based on mean incremental population NHEs, demonstrating that such
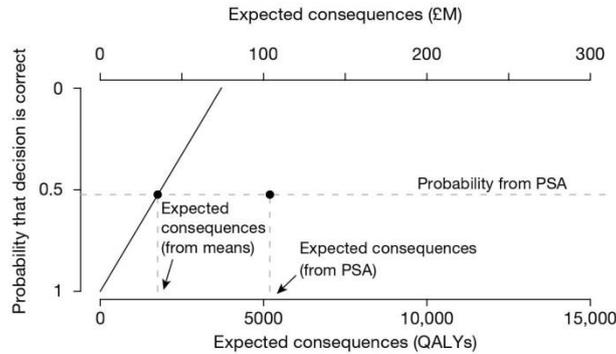
FIGURE 59 Probability that CLOP is cost-effective and the consequences of uncertainty. The expected consequences of uncertainty (from means) are from the results of alternative analysis (4) in *Table 89*.
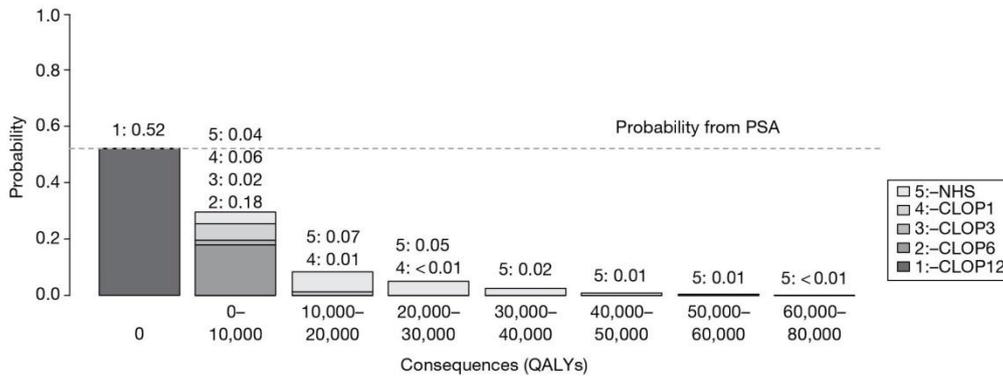
FIGURE 60 Distribution of the consequences of uncertainty for CLOP.

simple estimates may be misleading (see *Figure 59*). The consequences can also be expressed as the equivalent NHS resources required to generate the same population NHEs: £103.9M (see *Table 90*). At a higher cost-effectiveness threshold (£30,000 per QALY) the consequences fall because a decision to approve CLOP will be less uncertain (see *Table 90*).

The value of the expected consequences of uncertainty can be interpreted as an estimate of the population NHEs that could be gained, over the time horizon of this technology, if the uncertainty about treatment and its duration could be immediately resolved; it indicates an expected upper bound on the benefits of more research.[g] A judgement at this point that more research might be worthwhile seems reasonable as the potential benefits exceed the likely costs.

The expected consequences increase with the technology time horizon and size of the patient population. *Table 91* evaluates the consequences of uncertainty (using PSA) for alternative technology time horizons.

Analysis of subgroups
There were no relevant subgroups in the CLOP case study.

TABLE 91 Expected consequences of uncertainty for different technology time horizons

Technology time horizon, years	Expected consequences, QALYs (£M)
5	2819 (56.4)
10	5194 (103.9)
15	7193 (143.9)
20	8876 (117.5)

Alternative scenarios

The probabilistic analysis reported above reflects mainly the uncertainty over input parameters, given a set of assumptions made when estimating expected costs and QALYs. However, there are often alternative views not considered initially about, for example, the quality and relevance of evidence as well as other assumptions that might have been made. These are commonly presented as separate scenarios, with estimates of costs and QALYs presented for each. In the CLOP case study, base-case analyses assume a constant relative treatment effect for different durations of treatment. An alternative assumption (scenario B) was that the relative treatment effect also differed by duration based on the data reported in the SIGN guidelines. *Tables 92* and *93* present a summary of the cost-effectiveness, decision uncertainty and expected consequences of uncertainty for the alternative scenarios. The alternative assumption that relative treatment effects differ by duration made longer durations less cost-effective and reduced the expected consequences of uncertainty from 5195 to 3969 QALYs.

The uncertainty within each scenario will be sufficient to indicate the potential benefits of research only when this scenario is the only one regarded as credible. However, when more than one scenario might be credible and carry some 'weight', there will be uncertainty *between* as well as *within* scenarios. Much of the deliberation by the Appraisal Committee often surrounds the scientific value judgements required to judge the credibility of the alternative assumptions represented by the scenarios. The 'weighting' of scenarios can be made explicit by assigning probabilities to represent how credible each is believed to be. The weighted average of costs and QALYs across scenarios can easily be calculated. It is also tempting to take a simple weighted average of the expected consequences of uncertainty across these scenarios as well. However, a simple weighted average may under- or overestimate the combined consequences of the uncertainty within and between scenarios. The correct estimate requires the probabilities (weights) to be applied in a way that correctly identifies the consequences of uncertainty (for further details refer to *Appendix 11, Why averaging scenarios may be misleading*). Although this does not require additional simulation and is quick and easy to implement, it does require that either the probabilities are made explicit in advance or that estimates be presented for a range of probabilities that might represent the judgement of the Appraisal Committee following deliberation.

In the CLOP case study, although scenario A was regarded as more credible by the Appraisal Committee, scenario B might nevertheless carry some weight or have some probability associated with it. *Figure 61* shows that, in this case study, the simple weighted average of expected consequences (linear combination of mean estimates) approximates reasonably the correct estimate based on weighting the output of PSA. This figure also shows how these estimates can be presented for a range of probabilities.

Closed-form solutions can also correctly estimate the consequences of within- and between-scenario uncertainty (*Figure 62*) and can be used instead of weighting the output of PSA.

TABLE 92 Expected cost-effectiveness of CLOP with alternative scenarios

Treatment	Cost (£)	Health effects, QALYs	ICER (£/ QALY)	Cost-effectiveness threshold £20,000 per QALY NHE, QALYs (£M)	Cost-effectiveness threshold £30,000 per QALY NHE, QALYs (£M)
Scenario A (base case)					
CLOP12	10,394,830,647	4,194,554	18,663	3,674,813 (73,496)	3,848,060 (115,442)
CLOP6	10,256,672,674	4,187,151	10,477	3 674,318 (73,486)	3,845,262 (115,358)
CLOP3	10,180,425,730	4,179,874	9396	3,670,853 (73,417)	3,840,526 (115,216)
CLOP1	10,121,529,942	4,173,605	4961	3,667,529 (73,351)	3,836,221 (115,087)
NHS	10,072,035,344	4,163,629	–	3,660,027 (73,201)	3,827,894 (114,837)
Scenario B					
CLOP12	10,377,895,363	4,236,359	20,494	3,717,464 (74,349)	3,890,429 (116,713)
CLOP6	10,236,295,027	4,229,449	11,963	3,717,635 (74,353)	3,888,240 (116,647)
CLOP3	10,154,505,201	4,222,613	4087	3,714,887 (74,298)	3,884,129 (116,524)
CLOP1	10,044,958,670	4,195,806	3601	3,693,558 (73,871)	3,860,974 (115,829)
NHS	9,941,658,953	4,167,119	–	3,670,036 (73,401)	3,835,730 (115,072)

TABLE 93 Expected consequences of uncertainty for CLOP with alternative scenarios

Treatment	ICER (£/ QALY)	Cost-effectiveness threshold £20,000 per QALY Incremental NHE,[a] QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)	Cost-effectiveness threshold £30,000 per QALY Incremental NHE,[a] QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
Scenario A (base case)							
CLOP12	18,663	495 (9.9)	0.524	5194 (103.9)	2798 (56.0)	0.677	3657 (109.7)
CLOP6	10,477	3465 (69.3)	0.180		4736 (94.7)	0.092	
CLOP3	9396	3324 (66.5)	0.018		4305 (86.1)	0.009	
CLOP1	4961	7502 (150.0)	0.075		8327 (166.5)	0.052	
NHS	–	–	0.202		–	0.170	
Scenario B							
CLOP12	20,494	–171 (–3.4)	0.435	3969 (79.4)	2189 (43.8)	0.564	2871 (86.1)
CLOP6	11,963	2747 (54.9)	0.327		4110 (82.2)	0.268	
CLOP3	4087	21,329 (426.6)	0.237		23,155 (463.1)	0.168	
CLOP1	3601	23,522 (470.4)	0.001		25,244 (504.9)	0.000	
NHS	–	–	0.000		–	0.000	

a   For scenario A, the incremental NHE represents the mean additional population NHE of moving from the least to the most effective alternative. For simplicity, in scenario B the order of treatments was maintained the same as that in scenario A.

Point 4: Is research possible with approval?

The fourth point on the checklist requires an assessment of the type of evidence that is needed and a judgement of whether or not the research required to generate it can be conducted while the technology is approved. Although the decision at this point does not lead directly to guidance, it does determine whether AWR or OIR recommendations are possibilities. This judgement will depend, in part, on whether or not the type of evidence that is needed will require an experimental research design. For example, more precise estimates of relative treatment effect are likely to require a RCT if the dangers of selection bias are to be avoided. However,
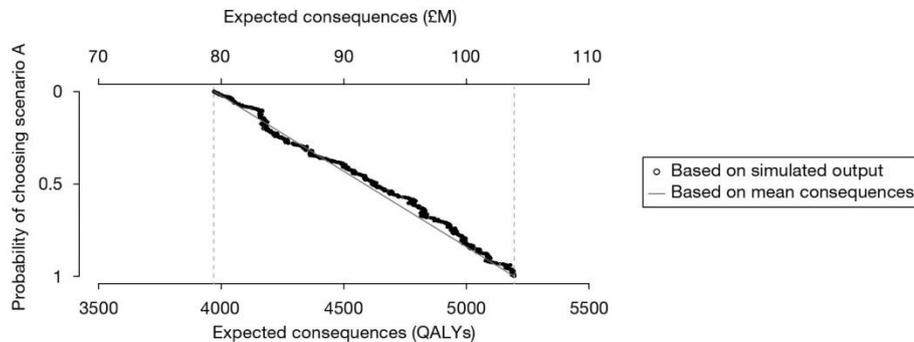
FIGURE 61 Expected consequences of uncertainty with alternative scenarios: CLOP case study.
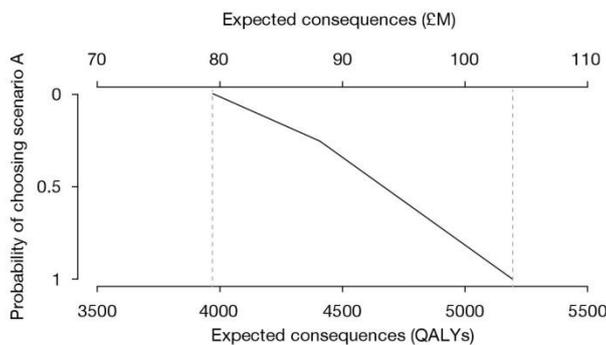
FIGURE 62 Expected consequences of uncertainty with alternative scenarios: CLOP case study.

further RCTs for this particular indication and patient group are unlikely to be possible once a technology is approved for widespread NHS use.

This assessment requires judgements about (1) how important particular types of parameters (inputs to the economic model) are to estimates of cost and QALYs, (2) what values these parameters would have to take to change a decision based on expected cost-effectiveness, (3) how likely it is that parameters might take such values and (4) what would be the consequences if they did, that is, what might be gained in terms of population NHEs if the uncertainty in the values of these parameters could be immediately resolved.

Assessing the importance of parameters
The type of economic model used to estimate expected cost-effectiveness in NICE appraisal specifies the relationship between the inputs (the parameters) and the outputs (costs and QALYs). A simple summary of the direction and strength of these relationships can be provided by calculating elasticities for each, that is, the proportionate change in the NHEs of each alternative, and differences in NHEs, owing to a 1% change in the value of the parameter. The elasticities regarding parameters of the CLOP case study are presented in *Table 94*. Those parameters with high elasticities in absolute value (especially with respect to differences in NHEs) might be regarded as more 'important'; see as an example the elasticity for RR_death.

TABLE 94 Elasticity for small changes (1%) in mean parameter values: CLOP case study

Parameter			Elasticity over the NHEs (QALYs) of					Elasticity over the incremental NHEs (QALYs) of		
			CLOP12	CLOP6	CLOP3	CLOP1	NHS	CLOP12 vs NHS	CLOP12 vs CLOP6	CLOP12 vs all[a]
Natural history	1	P_die_0.1	−0.208	−0.207	−0.207	−0.207	−0.222	0.014	–	0.003
	2	P_NFMI_0.1	−0.012	−0.012	−0.011	−0.011	−0.015	0.004	–	–
	3	P_die_1.3	−0.137	−0.137	−0.137	−0.147	−0.145	0.008	–	0.004
	4	P_NFMI_1.3	−0.002	−0.002	−0.002	−0.002	−0.002	0.001	–	–
	5	P_die_3.6	−0.146	−0.146	−0.157	−0.157	−0.154	0.008	–	0.007
	6	P_NFMI_3.6	−0.005	−0.005	−0.007	−0.007	−0.007	0.002	–	0.001
	7	P_die_6.12	−0.148	−0.159	−0.158	−0.157	−0.155	0.007	0.011	0.010
	8	P_NFMI_6.12	−0.005	−0.007	−0.007	−0.007	−0.007	0.002	0.002	0.002
	9	TP_AC	−0.121	−0.120	−0.120	−0.120	−0.118	−0.003	−0.001	−0.002
	10	TP_AD	−3.637	−3.622	−3.604	−3.594	−3.541	−0.096	−0.016	−0.047
	11	TP_CD	−0.233	−0.235	−0.239	−0.240	−0.253	0.020	0.002	0.009
	12	TP_BD	−0.586	−0.593	−0.602	−0.605	−0.641	0.055	0.007	0.024
Utilities	13	U_Well	0.746	0.745	0.743	0.742	0.737	0.009	0.001	0.004
	14	U_Well1	6.090	6.064	6.034	6.017	5.929	0.160	0.026	0.079
	15	U_NFMI	0.133	0.134	0.136	0.136	0.144	−0.011	−0.001	−0.005
	16	U_PostMI	1.138	1.150	1.165	1.171	1.236	−0.099	−0.012	−0.043
Relative effect	17	RR_death	−0.639	−0.491	−0.344	−0.207	–	−0.641	−0.150	−0.380
	18	RR_NFMI	−0.024	−0.018	−0.013	−0.011	–	−0.025	−0.006	−0.014
Costs	19	C_Well	−0.740	−0.737	−0.733	−0.731	−0.720	−0.019	−0.003	−0.009
	20	C_MI_LT	−0.051	−0.052	−0.053	−0.053	−0.056	0.004	0.001	0.002
	21	C_PostMI	−0.142	−0.143	−0.145	−0.146	−0.154	0.012	0.002	0.005
	22	TC_Well_Dead	−0.027	−0.027	−0.027	−0.027	−0.027	–	–	–
	23	C_t1	−0.045	–	–	–	–	−0.045	−0.045	−0.045
	24	C_t2	–	−0.033	–	–	–	–	0.033	0.008
	25	C_t3	–	–	−0.026	–	–	–	–	0.007
	26	C_t4	–	–	–	−0.022	–	–	–	0.005
	27	C_t5	–	–	–	–	−0.016	0.016	–	0.004

C, cost; P, probability of events in the short-term decision tree; RR, relative risk; TP, annual transition probabilities between states [well (A), post MI (B), MI (C), dead (D)] of a Markov model characterising long-term progression of disease; U, utility.

a   A weighted average of the NHEs of the remaining treatments was used. The weights reflect the probabilities of each treatment being cost-effective.

Although these measures of importance are more instructive than a series of arbitrary one-way sensitivity analyses, it does not directly help the assessment of what values parameters must take to change decisions and how likely such values might be. A simple summary of the values that particular parameters must take to make each of the alternatives cost-effective can also be provided. For CLOP this information is in *Table 95*. It shows, for example, that if P_die_0.1 is > 0.10, treatments other than 12 months of treatment with CLOP are deemed cost-effective. However, although instructive, such 'threshold values' do not indicate how likely it is that a threshold will be crossed or the combined effect of groups of related parameters.

Assessment of uncertainty
The judgement about how likely it is that parameters might take values that will change the technology expected to be cost-effective can be informed by the results of probabilistic analysis.

TABLE 95 Range of parameter values to make each alternative cost-effective

Parameter			Mean value	CLOP12	CLOP6	CLOP3	CLOP1	NHS
Natural history	1	P_die_0.1	0.032	0 to 0.10	0.11 to 0.54	0.54 to 0.63	0.63 to 1	–
	2	P_NFMI_0.1	0.040	0 to 0.14	0.14 to 0.71	0.71 to 0.82	0.82 to 1	–
	3	P_die_1.3	0.022	0 to 0.10	0.10 to 0.55	0.55 to 1	–	–
	4	P_NFMI_1.3	0.004	0 to 0.10	0.10 to 0.7	0.7 to 1	–	–
	5	P_die_3.6	0.023	0.01 to 0.10	0.10 to 1	0 to 0.01	–	–
	6	P_NFMI_3.6	0.011	0 to 0.11	0.11 to 1	–	–	–
	7	P_die_6.12	0.024	0.02 to 1	0 to 0.02	–	–	–
	8	P_NFMI_6.12	0.009	0.005 to 1	0 to 0.005	–	–	–
	9	TP_AC	0.018	0 to 0.06	0.06 to 1	–	–	–
	10	TP_AD	0.072	0 to 0.08	0.08 to 0.10	–	–	0.10 to 1
	11	TP_CD	0.188	0.12 to 1	0 to 0.12	–	–	–
	12	TP_BD	0.070	0.06 to 1	0.04 to 0.06	–	–	0 to 0.04
Utilities	13	U_Well	0.798	0.29 to 1	0 to 0.29	–	–	–
	14	U_Well1	0.930	0.90 to 1	0.74 to 0.90	–	–	0 to 0.74
	15	U_NFMI	0.801	0 to 1	–	–	–	–
	16	U_PostMI	0.931	0 to 1	–	–	–	–
Relative effect	17	RR_death	0.931	0 to 0.93	0.94 to 0.97	0.97 to 0.98	0.98 to 0.99	1.00 to max.[a]
	18	RR_NFMI	0.710	0 to 0.82	0.83 to 1.55	1.56 to 1.83	–	1.84 to max.[a]
Costs	19	C_Well	2061.5	0 to 2690	2690 to 5611	–	–	5611 to max.[a]
	20	C_MI_LT	6050	0 to max.[a]	–	–	–	–
	21	C_PostMI	2309.7	870 to max.[a]	0 to 870	–	–	–
	22	TC_Well_Dead	871.5	0 to 20,474	20,474 to max.[a]	–	–	–
	23	C_t1	895.1	0 to 910	910 to max.[a]	–	–	–
	24	C_t2	651.6	630 to max.[a]	0 to 630	–	–	–
	25	C_t3	524.2	370 to max.[a]	–	0 to 370	–	–
	26	C_t4	434.8	150 to max.[a]	–	–	0 to 150	–
	27	C_t5	329.8	0 to max.	–	–	–	–

C, cost; max., maximum; P, probability of events in the short-term decision tree; RR, relative risk; TP, annual transition probabilities between states [well (A), post MI (B), MI (C), dead (D)] of a Markov model characterising long-term progression of disease; U, utility.
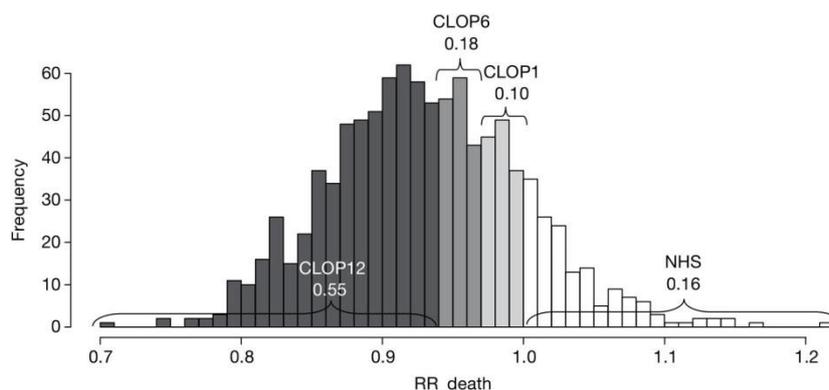
a An upper bound was used in evaluating unbounded parameters: relative risk parameters were evaluated to a maximum value of 50 and cost parameters were evaluated to a maximum value of £100,000.

This is because the distributions assigned to parameters in PSA describe how uncertain the parameter estimates are, such that they ought to reflect the amount and quality of exiting evidence. The probabilities that each parameter might take values that would lead to each of the alternatives being cost-effective are reported for the CLOP case study in *Table 96*. This essentially decomposes the overall probabilities reported in *Table 91* into the contribution that each parameter makes.[h] Interestingly, it indicates that it is uncertainty in the estimates of relative effect (RR_death, parameter 17) that contributes most to the probability of error associated with 12 months of treatment. It is the only parameter that might take values which could make any of the other alternatives cost-effective. To better visualise this, the histogram in *Figure 63*, showing the distribution of values that this parameter may take, is shaded to highlight the ranges of values that lead to different adoption decisions. Despite contributing to decision uncertainty, we are still unclear on whether or not there are significant consequences from these shifts in the adoption decision.

TABLE 96 Probabilities associated with parameter values: CLOP case study

Parameter			CLOP12	CLOP6	CLOP3	CLOP1	NHS
Natural history	1	P_die_0.1	1	–	–	–	–
	2	P_NFMI_0.1	1	–	–	–	–
	3	P_die_1.3	1	–	–	–	–
	4	P_NFMI_1.3	1	–	–	–	–
	5	P_die_3.6	1	–	–	–	–
	6	P_NFMI_3.6	1	–	–	–	–
	7	P_die_6.12	0.65	0.35	–	–	–
	8	P_NFMI_6.12	0.91	0.09	–	–	–
	9	TP_AC	1	–	–	–	–
	10	TP_AD	0.83	0.17	–	–	–
	11	TP_CD	1	–	–	–	–
	12	TP_BD	0.85	0.15	–	–	–
Utilities	13	U_Well	1	–	–	–	–
	14	U_Well1	0.94	0.06	–	–	–
	15	U_NFMI	1	–	–	–	–
	16	U_PostMI	1	–	–	–	–
Relative effect	17	RR_death	0.55	0.18	0.01	0.10	0.16
	18	RR_NFMI	0.97	0.03	–	–	–
Costs	19	C_Well	0.78	0.19	–	–	0.03
	20	C_MI_LT	1	–	–	–	–
	21	C_PostMI	0.89	0.11	–	–	–
	22	TC_Well_Dead	1	–	–	–	–
	23	C_t1	0.95	0.05	–	–	–
	24	C_t2	0.99	0.01	–	–	–
	25	C_t3	1	–	–	–	–
	26	C_t4	1	–	–	–	–
	27	C_t5	1	–	–	–	–

C, cost; P, probability of events in the short-term decision tree; RR, relative risk; TP, annual transition probabilities between states [well (A), post MI (B), MI (C), dead (D)] of a Markov model characterising long-term progression of disease; U, utility.

FIGURE 63 Histogram of the values of parameter RR_death: CLOP case study.

What type of evidence is needed?

Although an understanding of uncertainty and the importance of parameters separately is helpful, an assessment of the likely consequences of this uncertainty, and therefore what might potentially be gained in terms of population NHEs if uncertainty could be immediately resolved, is required. This assessment can directly inform the judgement of what evidence is needed and whether the type of research required to generate it will be possible with approval. As in the section on *Point 3: Does more research seem worthwhile?*, the results of PSA can inform this judgement as estimates of the expected consequences of uncertainty associated with each parameter combine both uncertainty in their potential values and their importance in terms of changing decisions and differences in NHEs. The distribution of consequences owing to uncertainty in specific parameters is illustrated for the parameter RR_death in *Figure 64*. There is a likelihood of 45% of values of RR_death leading to 12 months of treatment with CLOP not being the correct decision (100% – 55%; see *Table 97*). The average over this distribution provides the expected consequences of uncertainty for the parameter RR_death, which in this case is 4433 QALYs (or £88.7M).

The expected consequences of uncertainty associated with each parameter in the CLOP case study are reported in *Table 97*. This decomposes the overall expected consequences reported in *Table 93* into the contribution that each parameter makes and which other alternatives might offer higher NHEs than 12 months of treatment.[i] It confirms that it is uncertainty in the estimates of relative effect (RR_death) that contributes most and for which there is potentially the most to be gained by resolving this uncertainty through additional research (4433 QALYs or £88.7M).

Because more precise estimates of relative effects are likely to require a RCT, a judgement that the type of research needed will not be possible if 12 months of treatment with CLOP is approved may be reasonable. However, the potential benefits of resolving the uncertainty associated with other groups of parameters, for example costs and natural history (*Table 98*), might mean that other types of cheaper, non-experimental research could be worthwhile as well.[j]

Implications of between-scenario uncertainty

In *Point 3: Does more research seem worthwhile?*, *Alternative scenarios* the contribution that alternative scenarios might make to the overall expected consequences of uncertainty and therefore the potential gains from further evidence was considered and discussed. In situations in which more than one scenario might be regarded as credible there will be uncertainty *between* as well as *within* each of the scenarios. It was demonstrated that an assessment of the combined
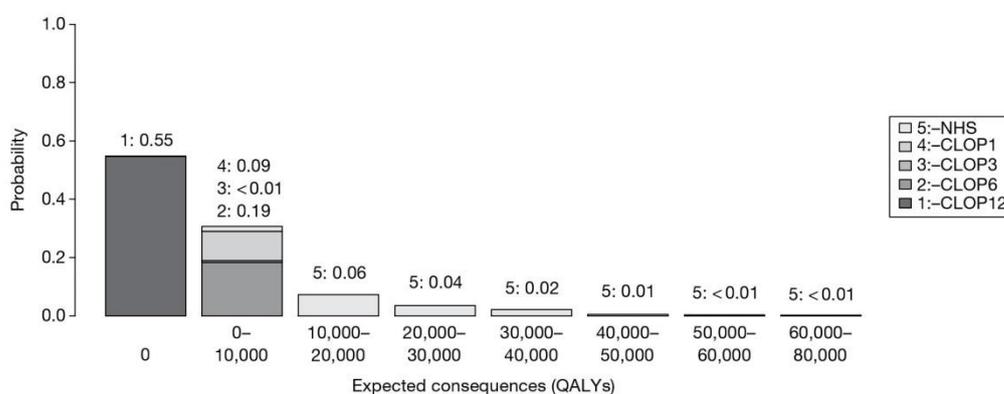
FIGURE 64 Distribution of consequences due to uncertainty in RR_death in the CLOP case study.

TABLE 97 Consequences of uncertainty associated with parameter values: CLOP case study

Parameter			Expected consequences (QALYs)					
			Decomposed by treatment choice					
			CLOP12	CLOP6	CLOP3	CLOP1	NHS	Overall
Natural history[a]	1	P_die_0.1	0	–	–	–	–	0
	2	P_NFMI_0.1	0	–	–	–	–	0
	3	P_die_1.3	0	–	–	–	–	0
	4	P_NFMI_1.3	0	–	–	–	–	0
	5	P_die_3.6	0	–	–	–	–	0
	6	P_NFMI_3.6	0	–	–	–	–	0
	7	P_die_6.12	0	250	–	–	–	250
	8	P_NFMI_6.12	0	9	–	–	–	9
	9	TP_AC	0	–	–	–	–	0
	10	TP_AD	0	47	–	–	–	47
	11	TP_CD	0	–	–	–	–	0
	12	TP_BD	0	35	–	–	–	35
Utilities[a]	13	U_Well	0	–	–	–	–	0
	14	U_Well1	0	10	–	–	–	10
	15	U_NFMI	0	–	–	–	–	0
	16	U_PostMI	0	–	–	–	–	0
Relative effect[a]	17	RR_death	0	284	16	518	3614	4433
	18	RR_NFMI	0	3	–	–	–	3
Costs[a]	19	C_Well	0	153	–	–	321	474
	20	C_MI_LT	0	–	–	–	–	0
	21	C_PostMI	0	8	–	–	–	8
	22	TC_Well_Dead	0	–	–	–	–	0
	23	C_t1	0	8	–	–	–	8
	24	C_t2	0	0	–	–	–	0
	25	C_t3	0	–	–	–	–	0
	26	C_t4	0	–	–	–	–	0
	27	C_t5	0	–	–	–	–	0

C, cost; P, probability of events in the short-term decision tree; RR, relative risk; TP, annual transition probabilities between states [well (A), post MI (B), MI (C), dead (D)] of a Markov model characterising long-term progression of disease; U, utility.

a Expected consequences for groups of parameters are (1) natural history: 369 QALYs (£7.4M); (2) utilities: 15 QALYs (£0.3M; (3) relative effect: 4504 QALYs (£90.1M); (4) costs: 547 QALYs (£10.9M). These are not equal to the sum of expected consequences for component parameters.

TABLE 98 Consequences of uncertainty associated with groups of parameters: CLOP case study

Group of parameters	Expected consequences, QALYs (£M)
Natural history	369 (7.4)
Relative effects	4504 (90.1)
Utilities	15 (0.3)
Costs	547 (10.9)

consequences of both sources of uncertainty requires 'weights' (probabilities) to be assigned to represent their credibility, which can then be applied directly to the simulated output from PSA (see *Appendix 11, Why averaging scenarios may be misleading*). However, the same analysis can also be used to identify the expected consequences of uncertainty associated with the alternative scenarios themselves, that is, what might be gained if evidence could immediately distinguish which scenario was 'true'. This can help to inform the assessment of what type of evidence might be needed and whether the research required to generate it is likely to be possible once a technology is approved for widespread NHS use.

In the CLOP analysis, scenarios A and B (treatment effect was constant or differed by treatment duration respectively) were associated with expected consequences of uncertainty of 5195 and 3969 QALYs respectively (*Table 99* and see *Point 3: Does more research seem worthwhile?, Alternative scenarios*). If both scenarios were regarded as equally likely, the overall expected consequences of uncertainty (combining consequences within and between scenarios) would be 4667 QALYs. However, the expected consequences of uncertainty associated with the two alternative scenarios themselves and what might be potentially gained if the uncertainty between them could be immediately resolved is relatively modest at 85 QALYs, that is, most of what might be gained from further evidence is associated with the parameters in *Table 97* rather than the alternative scenarios. This suggests that more evidence about the overall relative effect on mortality is more important than resolving uncertainty about whether such an effect differs by treatment duration.

Types and categories of guidance resulting from points 3 and 4

Points 3 and 4 of the checklist are critical because if research is not judged to be worthwhile then no further assessments are required. In the case of CLOP, more research appears to be worthwhile and the expected consequences of uncertainty are high.

In CLOP, research was not considered possible *with* approval. It is research that would provide more precise estimates of the relative effect of CLOP and of shorter treatment durations that is potentially valuable (see *Point 4: Is research possible with approval?*). As a consequence, the type of experimental design that is likely to be needed is unlikely to be possible if 12 months of treatment with CLOP is already approved for widespread NHS use.

Table 100 summarises the categories and types of guidance that could ultimately result from the sequence of assessments up to point 4.

TABLE 99 Expected consequences of uncertainty when scenarios are equally likely

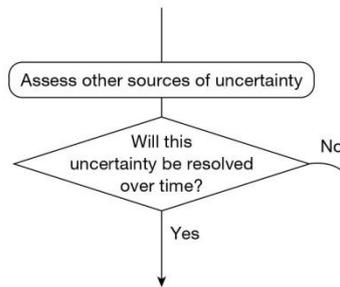| Consequences of uncertainty on	QALYs (£M)
Parameters, within scenario A (base case)	5194 (103.9)
Parameters, within scenario B	3969 (79.4)
Parameters, considering between-scenario uncertainty	2356 (47.1)
Between-scenario uncertainty, considering uncertainty within scenarios	85 (1.7)
Parameters and scenario uncertainty	4667 (13.3)

TABLE 100 Types and categories of guidance that could ultimately result from assessments 3 and 4

Assessment	1	2	3	4	5	6	7	Guidance
3	Yes	No	Yes	No	Yes/no	Yes	Yes	Approve[2]
4	Yes	No	Yes	No	Yes/no	Yes	No	OIR[1]
5	Yes	No	Yes	No	Yes/no	No	–	Approve[3]

Do the benefits of research exceed the costs?

Point 5: Will other sources of uncertainty resolve over time?

The fifth point on the checklist requires an assessment of whether or not changes are likely to occur in the future that will influence the cost-effectiveness of the alternative technologies and the potential benefits of research, that is, at the following point in the algorithm:

This assessment requires information about (1) changes in the prices of the technology and its comparators, (2) the emergence of new technologies that might make existing ones obsolete or change their cost-effectiveness and (3) other relevant research reporting. A number of potential sources of information and evidence were examined to inform this assessment for each case study (the full details of the sources and searches conducted are reported in *Appendix 4*).

Changes in the prices of the technology and its comparators

Changes in prices influence not only expected cost-effectiveness but also uncertainty and the potential benefits of research to future patients, for example if the price of a technology expected to be cost-effective is likely to fall significantly just before research reports the potential benefits will not be realised because approval of the technology will be less uncertain and there may be much less or little to gain from the results of the research. This assessment requires information about when major changes in prices are likely and some evidence about the likely extent of any changes. A major event in the life cycle of a pharmaceutical technology is the date at which the patient expires and cheaper generic versions of the brand become available. At the time of TA80[122] the patent for CLOP was expected to expire 7 years later and subsequent analysis assumes that at that time equivalent generic prices will be 25% of the original price of CLOP at the time of TA80.[k]

Entry of new technologies

The entry of a new technology may make the existing technology that is expected to be cost-effective obsolete (no longer the most cost-effective alternative). Even when it does not, it will tend to change the relative cost-effectiveness of the alternatives, influencing how uncertain a decision to approve the original technology will be for future patients and the potential gains from research. The information that was available indicated that one new technology relevant to CLOP might have been expected to enter. Information about this technology was limited so scenarios are used in *Point 6: Are the benefits of research greater than the costs?* to explore the implications for CLOP.

Other research reporting the technology and its comparators

Research that is already under way, commissioned or likely to be undertaken, whether in the UK or elsewhere, is relevant for the reasons discussed in this section. Despite an assiduous search no records relevant to CLOP were identified.

Point 6: Are the benefits of research greater than the costs?

The sixth point on the checklist requires a judgement of whether or not the potential benefits of conducting further research (initially considered at point 3) are likely to exceed the costs, that is, at the following point in the algorithm:
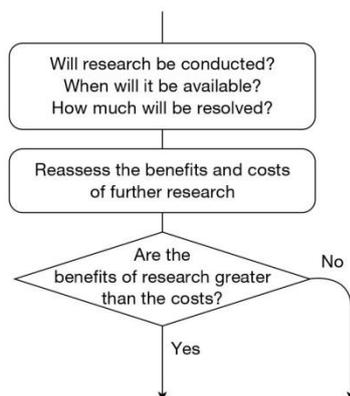
This requires an assessment of (1) whether or not the type of research that is required is likely to be conducted, (2) if conducted, when the results are likely to be available, (3) how much uncertainty is likely to be resolved and (4) the likely impact of other sources of uncertainty (identified in *Point 5: Will other sources of uncertainty resolve over time?*) on the longer-term benefits of research.

Will the research be conducted?

Even if research is recommended in OIR or AWR, it might not be undertaken by manufacturers or commissioned by research funders. Even if undertaken or commissioned, there is no guarantee that research will be able to recruit or it may not complete for other reasons. The expected consequences of uncertainty for CLOP reported in *Point 3: Does more research seem worthwhile?* are illustrated in *Figure 65* for a range of probabilities that research will be successfully undertaken. This indicates that the potential gains depend on a judgement of whether the research recommended as part of OIR will be successfully completed. It also illustrates that the cost of research (in this case considered to be either £1.5M or £10M) can be compared directly with the potential benefits by either expressing the potential gains in population NHEs as the equivalent NHS resources (i.e. the resources that would be required to generate the same NHEs) or expressing the cost of research in terms of the QALYs that could be gained elsewhere in the NHS by using the same resources to provide access to health care.

When will it be available?

Research, even if commissioned and successfully completed, will take time to complete and report; therefore, any assessment of the potential benefits should account for the fact that patient populations will not benefit from the results of research until they are available. Whether or not those patients who are prevalent while research is under way will be able to benefit from the results will depend on whether or not treatment decisions for presenting patients are irreversible (see *Point 2: Are there significant irrecoverable costs?*). In the CLOP case study, because the indication is acute, it is only those patients who are incident after the research reports who will realise any of the potential benefits. Treatment decisions are thus irreversible in the CLOP case study.

The potential value of research in CLOP over a range of possible time horizons for research to report is reported in *Figure 66*. How long research might take to report will depend in part on the design (follow-up, sample size and end points), recruitment rates and size of the eligible patient population, as well as on how efficient the organisation and data collection might be. The potential value of further research declines with the time to research reporting. This relationship gives some indication of the value of improving the timeliness of research through, for example, investment in research infrastructure or by adopting a research design that, although offering less potential benefits, can be conducted more quickly.

How much will be resolved?

Most research will not inform all of the parameters that determine expected costs and QALYs but usually a subset of them; therefore, the potential benefits of research that might be conducted will not be the total expected cost of uncertainty surrounding expected cost-effectiveness but some part of it. Earlier the potential benefits of different types of evidence were assessed. In the CLOP case study it was additional evidence about relative treatment effects that was most valuable and therefore experimental research may be required to provide more precise estimates of RR_death. The potential value of research that resolved uncertainty only about this relative treatment effect over a range of times to report is also represented in *Figure 66*. Although the potential value of research is lower at every time point, unless research is likely to take >8 years the potential value is still likely to exceed the costs. *Table 101* reports the potential benefits of research for groups of parameters by time to report.
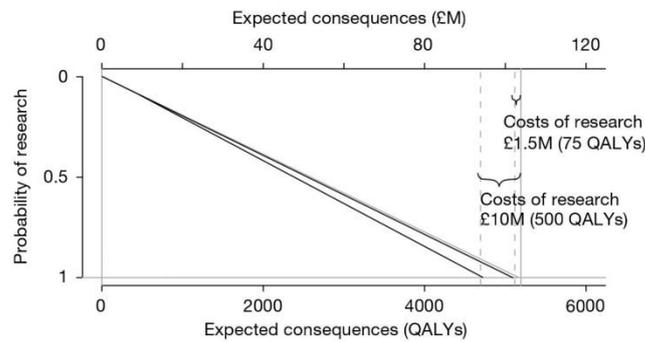
FIGURE 65 Expected potential benefits of research: CLOP case study.
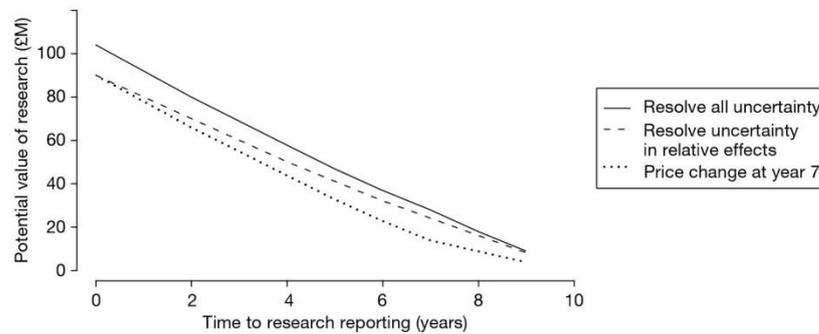
FIGURE 66 Potential value of research by time to report: CLOP case study.

TABLE 101 Potential value of research by time to report: CLOP case study

Time until research reports, years	Potential value of research, QALYs (£M)				
	Natural history	Relative effects	Utilities	Costs	Scenario uncertainty
Immediately	369 (7.4)	4504 (90.1)	15 (0.3)	547 (10.9)	85 (1.7)
1	326 (6.5)	3981 (79.6)	13 (0.3)	483 (9.7)	75 (1.5)
2	285 (5.7)	3475 (69.5)	12 (0.2)	422 (8.4)	65 (1.3)
3	245 (4.9)	2987 (59.7)	10 (0.2)	363 (7.2)	56 (1.1)
4	206 (4.1)	2515 (50.3)	8 (0.2)	305 (6.1)	47 (0.9)
5	169 (3.4)	2059 (41.1)	7 (0.1)	250 (5.0)	39 (0.8)
6	133 (2.7)	1618 (32.4)	5 (0.1)	197 (3.9)	30 (0.6)
7	98 (1.9)	1193 (23.8)	4 (0.1)	145 (2.9)	22 (0.5)
8	64 (1.3)	781 (15.6)	3 (0.1)	95 (1.9)	15 (0.3)
9	31 (0.6)	384 (8)	1 (0)	47 (0.9)	7 (0.1)

## What is the impact of other sources of uncertainty?

In *Point 5: Will other sources of uncertainty resolve over time?* the information that was publically available identified that the patent for CLOP was due to expire 7 years after the appraisal. Based on the OFT estimate that generic prices tend to be 25% of the original brand price,[12] this other source of uncertainty that will resolve over time can be integrated quantitatively when estimating the potential value of research over the lifetime of the technology. In this case a significant fall in price in year 7 will substantially reduce the uncertainty surrounding 12 months of treatment with CLOP. Therefore, after year 7 there is less to be gained from resolving uncertainty and so the potential and value of research findings for patients who are incident after year 7 are thereby reduced. The effect of a price change on research that could potentially resolve all uncertainty (costs, natural history and utilities as well as relative effect) is also illustrated in *Figure 66*. The potential value of the research is lower whenever the research reports. This is because after the price reduction the value of research is lower, and calculations include the value to future as well as current patient populations. Nevertheless, even if research did not report until 7 years the potential value (676 QALY or £14M) is likely to exceed the costs. The impact of the time of introduction of the discount is analysed in *Table 102*. This shows that if both the price change and research reporting are immediate the value of research is 2552 QALY (or 51M).

There was some evidence of the possible entry of a new technology (comparator) in the indication described in the CLOP case study; however, there was limited information on its characteristics. Therefore, two alternative but somewhat extreme scenarios are illustrated in *Figure 67*. In scenario A the new technology enters at year 5 and makes CLOP entirely obsolete, that is, not cost-effective and not uncertain – equivalent to a shorter technology horizon of 5 years. At this point there is no value in the evidence generated by research about CLOP.[l] In these circumstances the potential value of research is likely to exceed its costs only if it reports quickly. In scenario B the new technology has similar NHEs to 12 months of treatment with CLOP[m] and the uncertainty surrounding its expected cost-effectiveness is also similar. Now research about CLOP has more potential value in the future because it will also help resolve some of the uncertainty in the choice between CLOP and the new technology for patients who become incident after that time.

Tables 103 and *104* evaluate alternative times of introduction of the new technology in scenarios A and B, given a change in the price of CLOP at year 7 because of patent expiry.

TABLE 102 Potential value of research by time to report for a range of patent expiry times: CLOP case study

Time until research reports, years	Value of research for possible times of patent expiry, QALYs (£M)										
	No price change	9 years	8 years	7 years	6 years	5 years	4 years	3 years	2 years	1 year	Immediate price change
Immediately	5194 (104)	4969 (99)	4736 (95)	4495 (90)	4245 (85)	3986 (80)	3719 (74)	3442 (69)	3156 (63)	2859 (57)	2552 (51)
1	4591 (92)	4366 (87)	4132 (83)	3891 (78)	3641 (73)	3383 (68)	3116 (62)	2839 (57)	2552 (51)	2256 (45)	2256 (45)
2	4008 (80)	3783 (76)	3549 (71)	3308 (66)	3058 (61)	2800 (56)	2533 (51)	2256 (45)	1969 (39)	1969 (39)	1969 (39)
3	3444 (69)	3219 (64)	2986 (60)	2745 (55)	2495 (50)	2237 (45)	1969 (39)	1692 (34)	1692 (34)	1692 (34)	1692 (34)
4	2900 (58)	2675 (53)	2442 (49)	2201 (44)	1951 (39)	1692 (34)	1425 (28)	1425 (28)	1425 (28)	1425 (28)	1425 (28)
5	2374 (47)	2149 (43)	1916 (38)	1675 (33)	1425 (29)	1167 (23)	1167 (23)	1167 (23)	1167 (23)	1167 (23)	1167 (23)
6	1866 (37)	1641 (33)	1408 (28)	1167 (23)	917 (18)	917 (18)	917 (18)	917 (18)	917 (18)	917 (18)	917 (18)
7	1375 (28)	1150 (23)	917 (18)	676 (14)	676 (14)	676 (14)	676 (14)	676 (14)	676 (14)	676 (14)	676 (14)
8	901 (18)	676 (14)	443 (9)	443 (9)	443 (9)	443 (9)	443 (9)	443 (9)	443 (9)	443 (9)	443 (9)
9	443 (9)	218 (4)	218 (4)	218 (4)	218 (4)	218 (4)	218 (4)	218 (4)	218 (4)	218 (4)	218 (4)
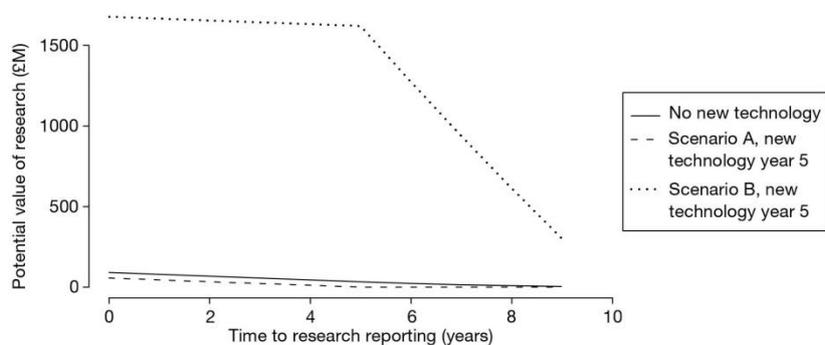
FIGURE 67 Potential value of research and other sources of uncertainty: CLOP case study. Assumes that a prince change will occur at year 7 because of patent expiry.

TABLE 103 Potential value of research by time to report for a range of times of entry of the new technology under scenario A: CLOP case study

Time until research reports, years	Value of research for possible times of introduction of a new technology, QALYs (£M)										
	Not introduced	9 years	8 years	7 years	6 years	5 years	4 years	3 years	2 years	1 year	Immediate introduction
Immediately	4495 (90)	4277 (86)	4052 (81)	3819 (76)	3328 (67)	2820 (56)	2294 (46)	1750 (35)	1186 (24)	603 (12)	0 (0)
1	3891 (78)	3674 (73)	3448 (69)	3215 (64)	2725 (54)	2216 (44)	1691 (34)	1146 (23)	583 (12)	0 (0)	0 (0)
2	3308 (66)	3091 (62)	2865 (57)	2632 (53)	2142 (43)	1633 (33)	1108 (22)	563 (11)	0 (0)	0 (0)	0 (0)
3	2745 (55)	2527 (51)	2302 (46)	2069 (41)	1578 (32)	1070 (21)	544 (11)	0 (0)	0 (0)	0 (0)	0 (0)
4	2201 (44)	1983 (40)	1758 (35)	1525 (30)	1034 (21)	526 (11)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
5	1675 (33)	1457 (29)	1232 (25)	999 (20)	508 (10)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
6	1167 (23)	949 (19)	724 (14)	491 (10)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
7	676 (14)	458 (9)	233 (5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
8	443 (9)	225 (5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
9	218 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

TABLE 104 Potential value of research by time to report for a range of times of entry of the new technology under scenario B: CLOP case study

Time until research reports, years	Value of research for possible times of introduction of a new technology, QALYs (£M)										
	Not introduced	9 years	8 years	7 years	6 years	5 years	4 years	3 years	2 years	1 year	Immediate introduction
Immediately	4495 (90)	18,020 (360)	32,018 (640)	46,507 (930)	61,253 (1225)	76,515 (1530)	92,311 (1846)	108,660 (2173)	127,024 (2541)	146,030 (2921)	165,701 (3314)
1	3891 (78)	17,416 (348)	31,415 (628)	45,904 (918)	60,649 (1213)	75,912 (1518)	91,708 (1834)	108,057 (2161)	126,420 (2528)	145,426 (2908)	145,426 (2908)
2	3308 (66)	16,833 (337)	30,832 (617)	45,321 (906)	60,066 (1201)	75,328 (1507)	91,125 (1822)	107,474 (2149)	125,837 (2517)	125,837 (2517)	125,837 (2517)
3	2745 (55)	16,270 (325)	30,269 (605)	44,757 (895)	59,503 (1190)	74,765 (1495)	90,561 (1811)	106,910 (2138)	106,910 (2138)	106,910 (2138)	106,910 (2138)
4	2201 (44)	15,726 (314)	29,724 (594)	44,213 (884)	58,959 (1179)	74,221 (1484)	90,017 (1800)	90,017 (1800)	90,017 (1800)	90,017 (1800)	90,017 (1800)
5	1675 (33)	15,200 (304)	29,199 (584)	43,687 (874)	58,433 (1169)	73,695 (1474)	73,695 (1474)	73,695 (1474)	73,695 (1474)	73,695 (1474)	73,695 (1474)
6	1167 (23)	14,692 (294)	28,690 (574)	43,179 (864)	57,925 (1159)	57,925 (1159)	57,925 (1159)	57,925 (1159)	57,925 (1159)	57,925 (1159)	57,925 (1159)
7	676 (14)	14,201 (284)	28,200 (564)	42,688 (854)	42,688 (854)	42,688 (854)	42,688 (854)	42,688 (854)	42,688 (854)	42,688 (854)	42,688 (854)
8	443 (9)	13,968 (279)	27,966 (5,59)	27,966 (560)	27,966 (560)	27,966 (560)	27,966 (560)	27,966 (560)	27,966 (560)	27,966 (560)	27,966 (560)
9	218 (4)	13,743 (275)	13,743 (275)	13,743 (275)	13,743 (275)	13,743 (275)	13,743 (275)	13,743 (275)	13,743 (275)	13,743 (275)	13,743 (275)

If the expected price reduction does not occur the potential value of research at each time point for both scenarios increases, for example for scenario B from 165,701 with a price change at year 7 to 174,519 QALYs without a price change when research is immediately available.[n]

The potential value of research presented in these figures, even after accounting for the type of evidence, follow-up and time until research reports, should still be regarded as an upper bound to the value that is likely to be realised by actual research for two reasons: (1) even well-designed research with large sample sizes will not fully resolve the uncertainty in the value that a parameter might take, especially in specific target populations and in a particular (future) context and (2) insofar as implementation of NICE guidance is not 'perfect' and all clinical practice might not immediately respond to the results of research, the full benefits will be realised only over time or with additional implementation efforts. For these reasons a judgement of whether benefits of research are likely to exceed the costs might be made conservatively, requiring evidence that, even in pessimistic scenarios, the research would still be worthwhile.

Types and categories of guidance resulting from points 5 and 6

The judgements made at points 5 and 6 of the checklist are critical because if the benefits of research are not judged to exceed the costs then no further assessment is required. In the CLOP case study, other sources of uncertainty will resolve over time, but the benefits of research still seem likely to be greater than the costs. *Table 105* summarises the categories and types of guidance that could ultimately result from the sequence of assessments up to point 6.

Are the benefits of approval greater than the costs?

Point 7: Are the benefits of approval greater than the costs?

The seventh and final point on the checklist requires an assessment and comparison of the benefits and costs of early approval. The costs of approval are not financial ones but opportunity costs and will include the potential value of any research that may be forgone as a consequence, for example if the research needed cannot be conducted once the technology is approved for use. They will also include any costs that are irrecoverably committed by approval. As well as the capital costs of equipment and facilities (or training and learning), they will also include the irrecoverable opportunity costs of initially negative NHEs (if treatment decisions are not irreversible – see the discussion in *Point 1: Is it expected to be cost-effective?*). A judgement of whether or not the benefits of approval and early access for current patients are likely to exceed the opportunity costs for future patients is required, that is, at the following point in the algorithm:
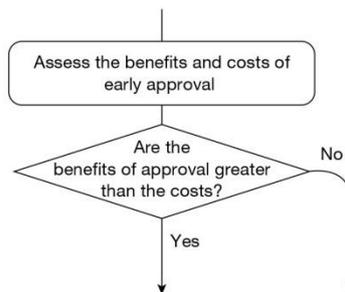
The decision at this point always leads directly to guidance, allocating all remaining possible pathways to a particular type and category of guidance.

An OIR recommendation is more likely to offer greater expected NHEs than approve if the research can be conducted quickly and report sooner, as fewer patients forgo access to CLOP and more can have treatment choice informed by the research findings. This is illustrated in *Figure 68*, which reports the difference between approve and OIR[o] in population NHEs over a range of times for when the research recommended in OIR might report. (Under OIR access to CLOP is not granted until research reports. In calculations we assumed that, before research reports, patients accrue an average NHE estimated using treatments other than CLOP. In averaging weights derived from the PSA results were used.) This takes account of both the expected changes in price at year 7 and research costs of £10M. It shows that OIR will be appropriate only if the research reports within 3 years of appraisal ($T^* = 3$) because beyond this time the NHEs forgone by withholding access to CLOP will exceed the potential gains to future patients.[p]

The trade-off between NHEs for current and future patients that lies behind *Figure 68* is illustrated in *Figure 69* using undiscounted values for ease of exposition. It illustrates the (per-period) population NHEs of approval and OIR if the research recommended as part of OIR reports at year 3. At this point, the initial losses of NHEs, caused by restricting access to CLOP (area A), start to be offset by the potential gains from the research findings (area B). The price change at year 7 increases the NHEs of approval (i.e. CLOP is more cost-effective) but on balance reduces the NHEs of OIR, that is, although CLOP is more cost-effective and offers greater NHEs the evidence generated by the research reporting is less valuable because the choice of treatment and duration is less uncertain. With research reporting at 3 years the initial losses of OIR (area A) are just offset by the later gains (area B) so $T^* = 3$. If research reported earlier than 3 years (area A < area B) OIR would be appropriate but if research reported later than 3 years (area A > area B) approve would be more appropriate.

The investment profile of alternative policies on the use of CLOP to treat patients over 10 years is illustrated in *Figure 70* in relation to rejecting CLOP. As expected, all regimens start by having initial losses (in OIR policies this is because of the initial investment in research); however, in
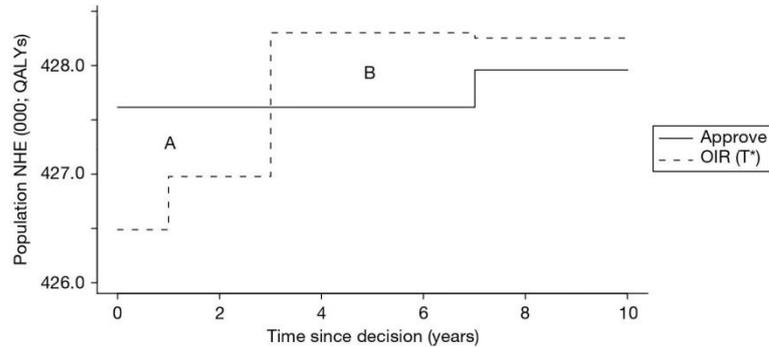
TABLE 105 Types and categories of guidance that could ultimately result from assessments 5 and 6

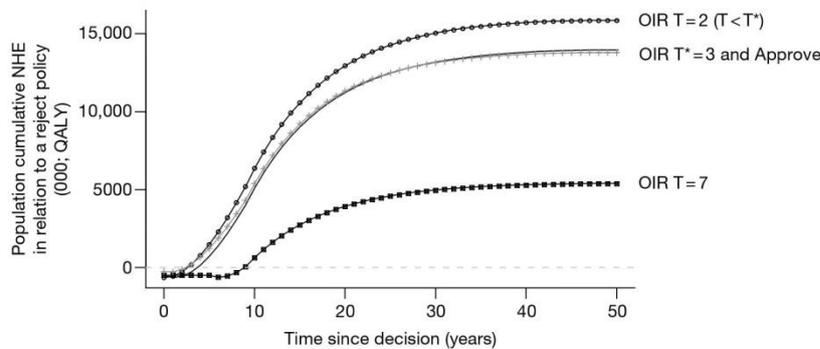| Assessment	1	2	3	4	5	6	7	Guidance
3	Yes	No	Yes	No	Yes	Yes	Yes	Approve[2]
4	Yes	No	Yes	No	Yes	Yes	No	OIR[1]
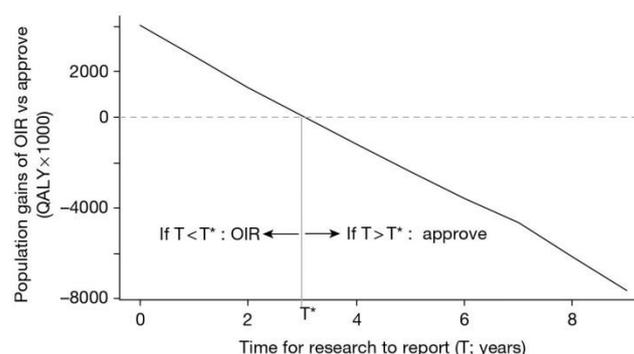
FIGURE 68 Population NHEs of approve and OIR by time to research reporting: CLOP case study. Assumes that a price change will occur at year 7 as a result of expiry of the patent.

FIGURE 69   Population NHEs of approve and OIR at T*: CLOP case study.

FIGURE 70   Time profile of approve and OIR: CLOP case study.

policies in which research reports earlier these initial losses are quickly compensated for by gains obtained from better treatment decisions after research reports.

There is no guarantee that the research recommended as part of OIR guidance will be conducted by manufacturers or commissioned by research funders. Even if it is, it is not certain that it will be successfully completed (see discussion in *Point 6: Are the benefits of research greater than the costs?*). Therefore, the probability that research will report at a particular time also needs to be considered. The implications of considering whether the recommended research will be conducted and when it might report are illustrated in *Figure 71*, which presents a boundary for when OIR might be appropriate or when approval should be granted. For example, if research is certain to report but will take 4 years, or when it will only take 1 year but has only a 50% chance of reporting, then OIR would not be appropriate and 12-month treatment with CLOP should be approved, that is, these points fall to the north-east of this boundary. Points to the south-west of the boundary indicate that OIR might be appropriate.

However, the estimates of the potential value of research on which these boundaries are based are still likely to overestimate the value that will be realised by research (see discussion in *Point 6: Are the benefits of research greater than the costs?*); they represent a necessary condition for OIR. Therefore, OIR guidance should require a conservative judgement that the point is almost certain to be below the boundary, rather than on balance close to it. For the same reason, points anywhere above the boundary represent a sufficient condition for approval. The boundary

when the change in price is included is to the south-west, reflecting the lower potential value of research, and OIR guidance, once CLOP becomes more cost-effective (see *Figure 71*). In this case it seems unlikely that the type of research required could report quickly enough and with sufficient confidence that OIR would be appropriate. Therefore, these assessments would support a judgement that the benefits of approval are likely to exceed the opportunity costs, and 'Approve[2]' (pathway 3 for CLOP) would be more appropriate.

The assessments that have been undertaken for CLOP can be brought together to consider (1) what would be the value of being able to conduct research while CLOP is approved and (2) what would be the value of making evidence that is needed by the NHS available at launch. These questions can be informed by the results already presented elsewhere but also reported in *Table 106*.

The difference in population NHEs between AWR (if it had been possible) and the next best feasible policy (i.e. OIR for $T < T^* = £30M$ and approve $T > T^* = £4M$) represents the value to the NHS of being able to conduct research while CLOP is approved for use, for example informing whether investment in better data collection, registries or information systems might make this possible. The difference in population NHEs between the case when all uncertainty had been resolved before appraisal (at launch) and the next best available policy (i.e. OIR for $T < T^* = £54M$ and approve for $T > T^* = £80M$) represents the value to the NHS of having access to the evidence needed at launch. This can inform policies that might make better and more relevant evidence available.
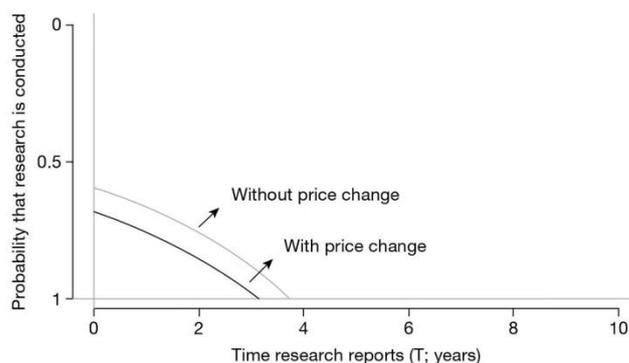
FIGURE 71 An OIR or approve boundary: CLOP case study.

TABLE 106 Population NHEs over the technology time horizon for different policies: CLOP case study

	Approve	OIR	AWR*	Reject	Value of AWR	Uncertainty resolved at launch	Value of evidence at launch
NHEs in QALYs							
$T < T^*$ (T = 2)	3,680,187	3,681,480	3,682,995	3,671,660	1515	3,684,181	2701
$T > T^*$ (T = 7)	3,680,187	3,675,487	3,680,362	3,671,660	175	3,684,181	3994
NHEs in £M							
$T < T^*$ (T = 2)	73,604	73,630	73,660	73,433	30	73,684	54
$T > T^*$ (T = 7)	73,604	73,510	73,607	73,433	4	73,684	80

AWR*, the expected population NHEs if AWR was to become a possibility; T, time to research reporting; T*, time to research reporting at which the NHEs of OIR and approve are equal.

Types and categories of guidance resulting from point 7

The decision at this point always leads directly to guidance, allocating all remaining possible pathways to a particular type and category of guidance. In CLOP, research was not considered possible *with* approval (i.e. 'Approve²' or 'OIR¹'). It is research that would provide more precise estimates of the relative effects of CLOP and of shorter treatment durations that is potentially valuable (see *Point 4: Is research possible with approval?*). As a consequence, the type of experimental design that is likely to be needed is unlikely to be possible if 12 months of treatment with CLOP is already approved for widespread NHS use. Although treatment with CLOP does commit initially negative NHEs that are irrecoverable, these should not be regarded as significant as the treatment decision for a presenting patient is irreversible in relevant time frames (see *Point 2: Are there significant irrecoverable costs?*). Therefore, AWR is not considered possible and so the benefits of early access to 12 months of treatment with CLOP (approval) must be compared with the potential value of OIR. The analysis of point 7 indicates that if research reports earlier than 3 years OIR would be appropriate; otherwise, approve would be more appropriate.

Appendix 9

Omalizumab for the treatment of severe persistent allergic asthma in children aged 6–11 years

Introduction

The use of OMAL for the treatment of severe persistent allergic asthma in children aged 6–11 years was not recommended by NICE following a STA appraisal (TA201[124] in 2010). The analysis compared OMAL as an add-on to standard therapy with standard therapy alone. The primary analysis was based on a prespecified severe asthma population within an international, multicentre, placebo-controlled RCT;[186] however, a high-risk subgroup within this population (recent hospitalisation for an asthma exacerbation) was also identified post hoc.

Omalizumab was not found to be cost-effective in either the severe or the severe/high-risk population. However, a RCT was recommended in section 6 of TA201[124] comparing OMAL with OCS in children to establish reduction in OCS use. The OMAL case study is used to demonstrate how the key principles and assessments in section 5 could inform the development of guidance for the use of OMAL. The following sections include a background to OMAL and a detailed examination of the sequence of assessments and decisions that lead to a particular category and type of guidance for OMAL. The analysis is from the standpoint of TA201; therefore, the case study can be used to see what the decision would have been if the checklist had been used.

Background to the case study

Asthma affects approximately 1.1 million children in the UK[187] and within this group there is a small but very significant number of children with severe symptoms whose asthma control remains poor despite best available therapy.[188] These children receive frequent or maintenance doses of OCS together with other controller medications. Clinical guidelines specify that treatment should aim to control asthma using the lowest possible OCS dose and, if possible, stop OCS treatment completely.[189]

The scope for the STA appraisal was the clinical effectiveness and cost-effectiveness of OMAL, within its licensed indication, for the treatment of severe persistent allergic asthma in children aged 6–11 years. OMAL is licensed as an add-on to existing therapy in patients aged 6 to < 12 years with severe, persistent allergic immunoglobulin E (IgE)-mediated asthma whose condition remains uncontrolled despite best standard care with high-dose inhaled corticosteroids (ICS) and long-acting beta agonists. The independent ERG[190] critiqued the evidence submitted by the manufacturer on the clinical effectiveness and cost-effectiveness of OMAL.

Intervention and population

The population of interest is children aged 6–11 years with severe persistent allergic asthma. Within this population, a high-risk subgroup is also identified, defined by a recent (within the previous year) hospitalisation for an asthma exacerbation. A post hoc subgroup analysis was

conducted as part of the STA submission by using a subpopulation of patients in the main IA-05 trial.[186] Both populations are considered in the case study analysis below. It is estimated that there is an annual incidence of 307 children in the UK with severe persistent allergic asthma who remain uncontrolled despite best available therapy and who would meet the criteria for therapy with OMAL.[188]

Omalizumab is administered by subcutaneous injection every 4 or 2 weeks. The permitted dosage depends on body weight and baseline levels of IgE. Evidence of treatment effectiveness must be observed at 16 weeks for treatment to be continued. Treatment duration on OMAL is assumed to be 10 years.[188] OMAL is compared as an add-on to standard therapy with standard therapy alone. Lifetime treatment with standard therapy is considered. The analysis is conducted from the perspective of the NHS and PSS and costs and outcomes are discounted at a rate of 3.5% per annum.[1]

Evidence on clinical effectiveness

The evidence on clinical effectiveness is primarily based on the results of an international, multicentre, placebo-controlled RCT in children aged 6–11 years with allergic asthma – study IA-05.[186] A subpopulation with more severe asthma was specified prospectively – referred to as IA-05 EUP. The primary analysis was conducted in the IA-05 EUP population. A high-risk subgroup (defined by a recent hospitalisation for an asthma exacerbation) was also specified and a post hoc analysis of efficacy used – referred to as the EUP hospitalisation subgroup. OMAL treatment was associated with a statistically significant reduction in the rate of clinically significant exacerbations but not clinically significant severe exacerbations.[188] OMAL use has been demonstrated to have only numerically small but statistically insignificant reductions in ICS use. There is limited evidence of a reduction in OCS use with OMAL.

Decision model

The manufacturer's STA submission included a Markov decision-analytic model. The model comprised six health states: (1) day-to-day symptoms with OMAL, (2) day-to-day symptoms with standard therapy, (3) clinically significant exacerbation, (4) clinically significant severe exacerbation, (5) asthma-related death, and (6) death from all causes. The model has been described in detail elsewhere.[188,191] Patients start in the appropriate day-to-day symptoms state based on treatment received. Treatment effectiveness was based on reduction in the rate of exacerbations and quality of life as a result of day-to-day symptoms. Movement between states was derived from the IA-05 EUP population[186] and three other studies.[192–194] Adverse events were not included in the model. *Table 107* summarises the key parameters and sources used in the decision model.

Lifetime costs and QALYs were determined by comparing lifetime standard therapy costs with costs of 10 years of OMAL add-on therapy followed by standard therapy. Only patients who were judged by physicians to have responded to OMAL after 16 weeks of therapy continued with treatment; non-responders to OMAL reverted to standard therapy at 16 weeks. The 28-week exacerbation rates (clinically significant and clinically significant severe) from IA-05 EUP were annualised for standard therapy and OMAL responders.[186] Death because of asthma was based on the probability reported by Watson *et al.*[193] and varied by age.

Most inputs in the model were assumed to be uncertain and the model was run probabilistically using Monte Carlo simulation (1000 simulations). Expected cost-effectiveness was determined using these probabilistic results as established within the NICE appraisal process.[a] The use of probabilistic analysis also allows uncertainty over model parameters to be translated into uncertainty in the overall results.

Key features and possible pathways

Omalizumab was not found to be cost-effective in either the severe or the severe/high-risk population. However, a RCT was recommended comparing OMAL with OCS in children to establish reduction in OCS use. This was made in section 6 of TA201;[124] therefore, OMAL is not an example of OIR at FAD but an example of OIR considered during appraisal. OMAL provides some interesting characteristics, which can be used to explore the implications of the principles and assessments outlined in *Chapter 5*:

- small patient population (approximately 300 children per annum) – rare condition
- consideration of subgroup analysis (inclusion of a high-risk hospitalisation subgroup)
- NICE decision to reject
- RCT recommended comparing OMAL with OCS in children to establish reduction in use of OCS in the FAD (but not an OIR decision).

The possible pathways through the algorithm that OMAL illustrates are reported in *Figure 31* in *Appendix 4*. OMAL is not expected to be cost-effective and there are no significant irrecoverable costs. The following sections examine the points on the checklist relating to the possible sequence of assessments and decisions that lead to a particular category and type of guidance for OMAL.

TABLE 107 Parameters of the decision model for OMAL

Parameter	Mean value	Distribution	Source
Treatment effectiveness			
Exacerbation rate per person for initial 24 weeks			
Standard therapy	1.939	Log-normal	IA-05 EUP data[186]
OMAL	1.363	Log-normal	IA-05 EUP data[186]
Exacerbation rates per year per person for OMAL	0.519	Log-normal	IA-05 EUP data[186]
Percentage of exacerbations that are severe, OMAL	27.30%	Beta	IA-05 EUP data[186]
Exacerbation rates per year per person for standard therapy	2.028	Log-normal	IA-05 EUP data[186]
Percentage of exacerbations that are severe, standard therapy	22.90%	Beta	IA-05 EUP data[186]
Proportion of OMAL responders	74.20%	Beta	IA-05 EUP data[186]
Relative risk for exacerbations, OMAL responders	0.256	Log-normal	IA-05 EUP data[186]
Relative risk for exacerbations, non-responders (same as standard therapy)	1	Fixed	IA-05 EUP data[186]
Exacerbation-related deaths by age (same for OMAL and standard therapy)			
0–11 years	0.09%	Beta	Watson *et al.* 2007[193]
12–16 years	0.32%	Beta	Watson *et al.* 2007[193]
17–44 years	0.38%	Beta	Watson *et al.* 2007[193]
45+ years	2.47%	Beta	Watson *et al.* 2007[193]
HRQoL			
Utility for standard therapy and non-responders all ages	0.669	Beta	Humbert *et al.* 2005[192]
Utility for OMAL responders aged < 12 years	0.669	Beta	Humbert *et al.* 2005[192]
Utility for OMAL responders aged 12+ years	0.779	Beta	Humbert *et al.* 2005[192]
Utility for clinically significant exacerbation	0.572	Beta	Lloyd *et al.* 2007[194]
Utility for clinically significant severe exacerbation	0.326	Beta	Lloyd *et al.* 2007[194]
Resource use and unit costs			
OMAL and standard therapy	See STA submission[188]	Fixed	–

Is it cost-effective and what are the risks?

Point 1: Is it expected to be cost-effective?

The sequence of assessments starts with cost-effectiveness and the expected impact on population NHEs, that is, at the following point in the algorithm:

Assess cost-effectiveness and
population net health effects

Is it cost-effective? — No →

Yes ↓
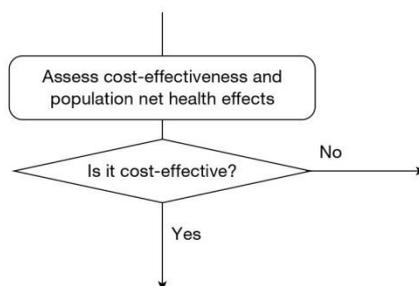
Cost-effectiveness at the patient level

Table 108 summarises the expected cost-effectiveness of OMAL per patient treated. OMAL as an adjunct to standard therapy is compared with standard therapy alone. The expected incremental differences in lifetime costs and QALYs of OMAL relative to standard therapy are £55,682 and 0.5933 QALYs, respectively, giving an ICER of £93,844. The ICER for OMAL is greater than the threshold of £20,000 per QALY;[b] therefore, OMAL is not expected to be cost-effective compared with standard therapy alone. Consequently, the incremental NHE of OMAL is negative.

The per-patient investment profile for OMAL is illustrated in *Figure 72* and shows that it is always expected to offer negative NHEs compared with standard care over the entire patient time horizon of 90 years, that is, the high costs of treatment are never compensated by future health gains. In this example, the initial treatment costs with OMAL continue for 10 years (10 years is assumed to represent the duration for which a patient would continue to receive treatment with OMAL) with health effects predominately obtained while on treatment. OMAL does not represent a 'risky purchase' but one that is simply not cost-effective at its current price.

Cost-effectiveness at the population level

Per-patient NHEs can also be expressed for the population of current and future patients. This requires information about the prevalence and future incidence of the target population. It also requires a judgement about the time horizon over which the technology will be used. It

TABLE 108 Expected cost-effectiveness of OMAL per patient treated

Treatment	Cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£)	Incremental NHE, QALYs (£)	NHE, QALYs (£)	Incremental NHE, QALYs (£)
OMAL + standard care	94,992	16.64	93,844	11.8861 (237,721)	−2.1908 (−43,815)	13.4693 (404,078)	−1.2627 (−37,882)
Standard care	39,310	16.04	–	14.0768 (281,536)	–	14.7320 (441,960)	–

Results are presented as the mean of the PSA. See *Appendix 11, Model linearity and correlation between parameters* for assessments of linearity and correlation.
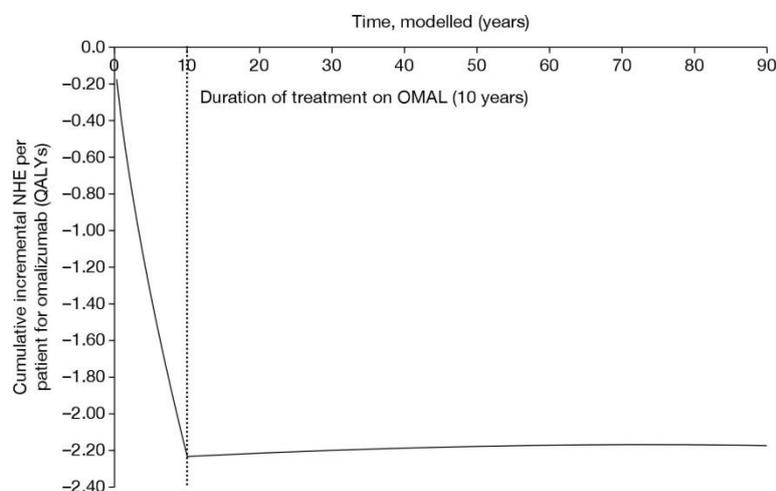
FIGURE 72 Cumulative incremental NHEs of OMAL over the patient time horizon.

is estimated that there are only 307 children in the UK per year with severe persistent allergic asthma who remain uncontrolled despite best available therapy and who would meet the criteria for therapy with OMAL.[188] As part of the budget impact assessment included with the manufacturer's STA submission, the expected uptake rate of OMAL in the first 5 years is assumed to be < 307 patients per year (*Table 109*). Market uptake is assumed to be 3.9% in the first year then 11.7%, 23.1%, 30.6% and 32.6% for years 2–5, respectively. This results in a patient pool of only 12 patients in year 1 rising to 100 patients in year 5. *Table 110* shows the total population NHE assuming that OMAL will be used over 10 years for (1) the estimated patient population in *Table 109* for the first 5 years and an incident population of 307 patients per year thereafter and (2) an incident population of 307 patients per year over the 10-year period. The expected cost-effectiveness is unchanged (ICER is the same as in *Table 108*) but the incremental NHE is more negative and significant at a population level. Clearly, as more patients are treated OMAL appears even less cost-effective as the high costs of treatment are never compensated by future health gains.

The investment profile for OMAL when used to treat patients over 10 years is illustrated in *Figure 73*. At a population level there are large losses that are never compensated for by later health gains. Unlike the other case studies, OMAL does not appear a more risky investment when evaluated at a population level; it is simply not cost-effective and never breaks even with standard therapy.

The time horizon of 10 years was chosen as a proxy for a complex and uncertain process of future changes in new technologies, prices and evidence.[71] The impact of different technology time horizons on NHEs is illustrated in *Table 111*. The incremental NHE becomes more negative when the technology time horizon is increased to 20 years. For a shorter duration of 5 years the losses are smaller but still amount to 3143 QALYs for a population of 307 patients per annum.

TABLE 109 Expected uptake rate of OMAL in the first 5 years

Year	Expected uptake rate (%)	Expected population starting on OMAL	Expected population on standard therapy
2010	3.9	12	295
2011	11.7	36	271
2012	23.1	71	236
2013	30.6	94	213
2014	32.6	100	207

TABLE 110 Expected cost-effectiveness of OMAL for the population

| Treatment	Costs (£M)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£M)	Incremental NHE, QALYs (£M)	NHE, QALYs (£M)	Incremental NHE, QALYs (£M)
Based on expected uptake rate							
OMAL + standard care	142	24,836	93,844	17,745 (355)	−3271 (−65)	20,108 (603)	−1885 (−57)
Standard care	59	23,950	–	21,015 (420)	–	21,994 (660)	–
Based on incidence of 307 patients per year							
OMAL + standard care	251	43,961	93,844	31,410 (628)	−5789 (−116)	35,593 (1,068)	−3337 (−100)
Standard care	104	42,393	–	37,199 (744)	–	38,930 (1,168)	–
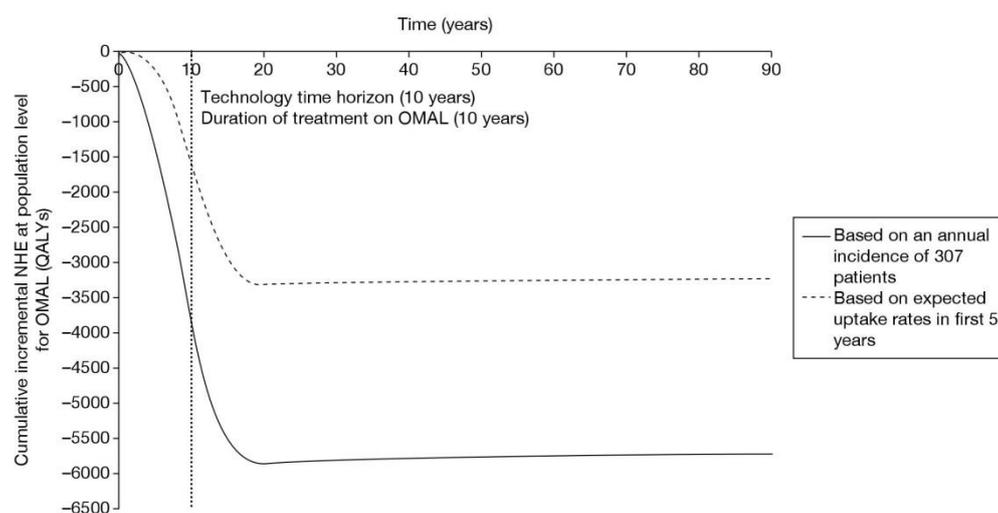
FIGURE 73 Cumulative incremental NHEs of OMAL for the population.

TABLE 111 Incremental NHEs for OMAL over a range of technology time horizons

Technology time horizon, years	Incremental NHE, QALYs (£M)	
	Annual incidence of 307 patients	Expected uptake rate in first 5 years
5	−3143 (−63)	−624 (−12)
10	−5789 (−116)	−3271 (−65)
15	−8017 (−160)	−5499 (−110)
20	−9893 (−198)	−7375 (−147)

Point 2: Are there significant irrecoverable costs?

The second point on the checklist requires (1) an assessment of whether or not there are irrecoverable costs and (2) a judgement of their potential significance, that is, at the following point in the algorithm:
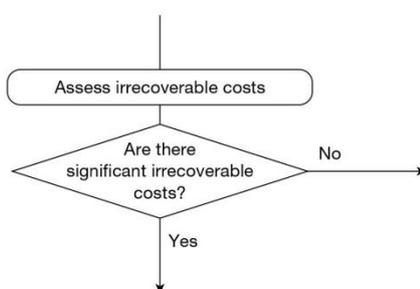
Identifying any significant irrecoverable costs

Irrecoverable costs are those that, once committed, cannot be recovered if guidance is changed at a later date. In the case of OMAL there are no capital investment costs associated with equipment or facilities. Drug costs associated with treatment occur annually per patient treated. In the absence of capital investment costs, NHEs accumulate over time at both a patient and a population level. For OMAL the profile of NHEs at a patient level (see *Figure 72*) exhibits negative NHEs in all time periods. Assessment at a population level and for longer technology time horizons simply increases the magnitude of the expected negative NHEs. Therefore, there are no irrecoverable costs associated with OMAL.

Types and categories of guidance resulting from points 1 and 2

Points 1 and 2 of the checklist do not lead directly to a category or type of guidance. The sequence of assessments and decisions, which ultimately lead to guidance, starts with cost-effectiveness, expected impact on population NHEs and significance of irrecoverable costs. In the case of OMAL, the technology is not expected to be cost-effective and there are no significant irrecoverable costs. *Table 112* summarises the categories and types of guidance that could ultimately result from the sequence of assessments up to point 2 (see *Table 32* in *Appendix 4*).

TABLE 112 Types and categories of guidance that could ultimately result from assessments 1 and 2

Assessment	1	2	3	4	5	6	7	Guidance
7	No	No	Yes	Yes	Yes/no	Yes	–	OIR[2]
8	No	No	Yes	Yes	Yes/no	No	–	Reject[1]
9	No	No	Yes	No	Yes/no	Yes	Yes	AWR[2]
10	No	No	Yes	No	Yes/no	Yes	No	Reject[2]
11	No	No	Yes	No	Yes/no	No	–	Reject[3]
12	No	No	No	–	–	–	–	Reject[4]

Is further research required?

Point 3: Does more research seem worthwhile?

The third point on the checklist requires an assessment of the potential benefits of conducting further research, that is, at the following point in the algorithm:
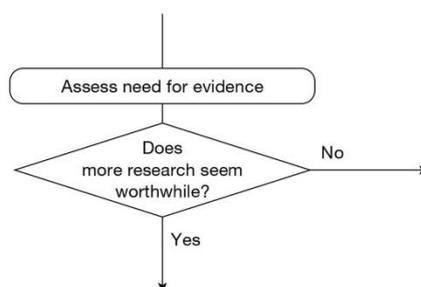
This requires judgements about (1) how uncertain a decision to approve or reject OMAL might be based on the estimates of expected cost-effectiveness and (2) whether or not the scale of the likely consequences of this uncertainty might justify further research. If the potential benefits of further research are unlikely to justify the costs, a judgement that more research does not seem worthwhile will lead directly to guidance.

Assessing the consequences of uncertainty

Omalizumab is not expected to be cost-effective compared with standard therapy but the estimates of costs and QALYs are uncertain so there is a chance that a decision to approve standard therapy rather than OMAL based on existing evidence will be incorrect, that is, OMAL might offer greater NHEs. Some assessment of the likely consequences of approving standard therapy when OMAL might be better could be based on the difference in expected NHEs, that is, the expected incremental population NHEs reported in *Table 110*. The simplest approach would be to weight the average NHEs for OMAL and standard therapy (reported in *Table 110*) by a judgement of the probability of an incorrect decision. For example, if the decision was judged to be 100% certain then there are no consequences and so there would be nothing to be gained by more research; however, as the probability that the decision is correct becomes less certain, the expected consequences (and hence potential value of more research) increase. *Table 113* shows the expected consequences of uncertainty based on a weighting of average NHEs.[c]

A judgement of how uncertain a decision might be can be informed by the PSA already used to estimate costs and QALYs. The probability that standard therapy is cost-effective rather

TABLE 113 Expected consequences of uncertainty based on a weighting of average NHEs

Probability decision is correct, p	Cost-effectiveness threshold £20,000 per QALY Expected consequences, QALYs (£M)	Cost-effectiveness threshold £30,000 per QALY Expected consequences, QALYs (£M)
1.00	0 (0)	0 (0)
0.99	58 (1)	33 (1)
0.95	289 (6)	167 (5)
0.90	579 (12)	334 (10)
0.75	1447 (29)	834 (25)
0.50	2895 (58)	1668 (50)
0.25	4342 (87)	2503 (75)
0.10	5210 (104)	3003 (90)
0.05	5500 (110)	3170 (95)
0.01	5731 (115)	3303 (99)
0.00	5789 (116)	3337 (100)

than OMAL is 1.00 at both £20,000 and £30,000 per QALY (*Table 114*). An estimate of the expected consequences of uncertainty should be based on the distribution of uncertainty from the PSA rather than on a simple estimate of average NHEs. *Table 114* shows that the expected consequences of uncertainty are 0.00; therefore, at the upper and lower bounds for the range that NICE has adopted for the threshold there is insufficient uncertainty to suggest that OMAL might be cost-effective.

The expected consequences provide an upper bound on the benefits of more research. Because the frequency of error is zero, there is no uncertainty. Therefore, a decision to reject OMAL is not uncertain; there are no consequences of uncertainty and nothing to be gained by more research.

Alternative scenarios

The uncertainty described above reflects uncertainty within the set of assumptions used to estimate expected costs and QALYs. However, there are often alternative views about the assumptions, which are usually presented as separate scenarios. For OMAL, the STA submission considered scenarios that would generate the most favourable ICERs but still plausibly reflect the benefits seen by paediatric asthma specialists in clinical practice. The scenarios included (1) a longer treatment duration, increased from 10 years to 20 years, (2) utility benefits in day-to-day symptoms for all age groups (the previous base-case analysis assumed a utility benefit for ages > 12 years only) and (3) an increase in mortality rate owing to clinically significant severe exacerbations of 100%. These scenarios were incorporated additively, that is, each scenario was considered in addition to the previous one. *Table 115* presents the cost-effectiveness of OMAL and the expected consequences of uncertainty for the alternative scenarios. Despite a fall in the ICER for OMAL from £93,844 to £62,049 with the addition of the three favourable assumptions, a decision to reject OMAL does not appear uncertain; the probability that OMAL is cost-effective is zero and there are no consequences of uncertainty or value to be gained by more research.

Types and categories of guidance resulting from point 3

Point 3 of the checklist is critical because if research is not judged to be worthwhile no further assessments are required. For OMAL, the ICER is substantially greater than the threshold and the NHE is always negative. A decision to reject OMAL is not uncertain and there are no consequences of uncertainty and nothing to be gained by additional research. *Table 116* summarises the category and type of guidance for OMAL.

TABLE 114 Expected consequences of uncertainty for OMAL

Treatment	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY			Cost-effectiveness threshold £30,000 per QALY		
		Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)	Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
OMAL + standard care	93,844	−5789 (−116)	0.000	0.000 (0)	−3337 (−100)	0.000	0.000 (0)
Standard care		–	1.000		–	1.000	

TABLE 115 Expected consequences of uncertainty with alternative scenarios

Additive scenario	ICER (£/QALY)	Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
Base case	93,844	−5789 (−116)	0.0000	0.0000 (0)
+ 20-year treatment duration	78,324	−9198 (−184)	0.0000	0.0000 (0)
+ utility difference in day-to-day symptoms for all ages	67,525	−8721 (−174)	0.0000	0.0000 (0)
+ mortality rate owing to clinically significant severe exacerbations increased by 100%	62,049	−8377 (−168)	0.0000	0.0000 (0)

TABLE 116 Type and category of guidance for OMAL

Assessment	1	2	3	4	5	6	7	Guidance
12	No	No	No	–	–	–	–	Reject[4]

Analysis of subgroups

Subgroups, once credibly defined, need to be considered in the same way as the base-case population. This involves starting at point 1 of the checklist, that is, entering at the top of the algorithm, and working through the sequence of assessments from point 1 to point 7 in the same way as for the primary base-case analysis. This can be used to identify whether or not different guidance should be arrived at for the subgroup population. It will also identify whether or not additional research is required in the specific patient subgroup that may not otherwise have been conducted for the base-case population.

For OMAL a high-risk subgroup was identified from within the base-case population. This was defined by a recent (within the previous year) hospitalisation for an asthma exacerbation. A post hoc subgroup analysis was conducted as part of the STA submission by using a subpopulation of patients in the main IA-05 trial.[186] Guidance for this subgroup is considered using the same assessments as for the base-case population. Consideration is given below to each of the checklist points.

Point 1: Is it expected to be cost-effective?
Cost-effectiveness at the patient level
Table 117 summarises the expected cost-effectiveness of OMAL per patient treated for the subgroup population. The expected incremental differences in lifetime costs and QALYs of OMAL relative to standard therapy are £40,927 and 0.5892 QALYs, respectively, giving an ICER of £69,463. The ICER for OMAL in the subgroup population is much lower than that in the base-case population but it is still greater than the threshold of £20,000 per QALY; therefore, OMAL is not expected to be cost-effective compared with standard therapy alone.

The per-patient investment profile for OMAL is illustrated in *Figure 74* for the subgroup population. It shows that OMAL is always expected to offer negative NHEs compared with standard therapy over the entire patient time horizon of 90 years, that is, the high costs of treatment are never compensated for by future health gains in this subpopulation either. OMAL is not expected to be cost-effective at its current price.

Cost-effectiveness at the population level

Table 118 shows the total population NHEs over 10 years, assuming that OMAL will be used to treat the estimated patient population in *Table 109* (based on uptake rates for the first 5 years and then 307 patients per year thereafter) rather than an incident population of 307 patients per year. The lower estimate of population size is used to reflect the fact that the high-risk subgroup would be expected to include fewer patients. The incremental NHE is less negative for the subgroup population than in the base case because of fewer patients and a lower differential effect between OMAL and standard therapy. *Figure 75* shows the investment profile for OMAL over 10 years. As seen above, the large losses are never compensated for by future health gains. OMAL is not cost-effective and never breaks even with standard therapy.

TABLE 117 Expected cost-effectiveness of OMAL per patient treated for the subgroup

Treatment	Cost (£)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£)	Incremental NHE, QALYs (£)	NHE, QALYs (£)	Incremental NHE, QALYs (£)
OMAL + standard care	82,681	15.11	69,463	10.9793 (219,585)	−1.4571 (−29,143)	12.3573 (370,718)	−0.7750 (−23,251)
Standard care	41,754	14.52	–	12.4364 (248,728)	–	13.1323 (393,969)	–
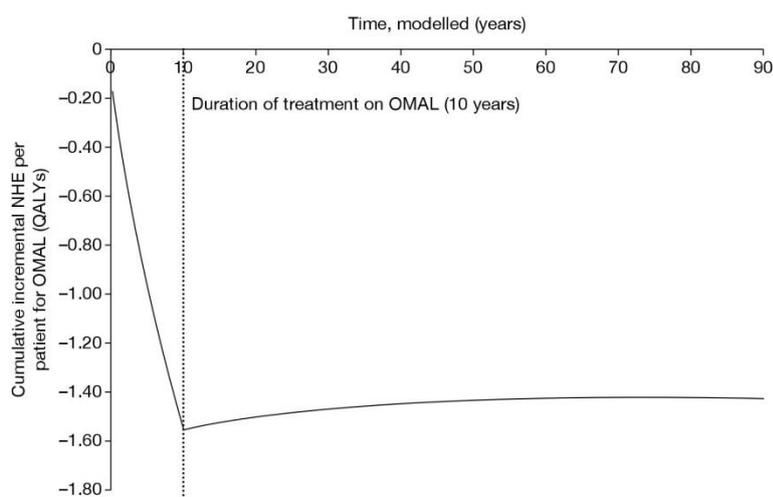
FIGURE 74 Cumulative incremental NHEs of OMAL over the patient time horizon for the subgroup.

TABLE 118 Expected cost-effectiveness of OMAL for the subgroup population

Treatment	Costs (£M)	QALYs	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£M)	Incremental NHE, QALYs (£M)	NHE, QALYs (£M)	Incremental NHE, QALYs (£M)
OMAL + standard care	123	22,563	69,463	16,391 (328)	−2175 (−44)	18,448 (553)	−1157 (−35)
Standard care	62	21,683	−	18,566 (371)	−	19,605 (588)	−
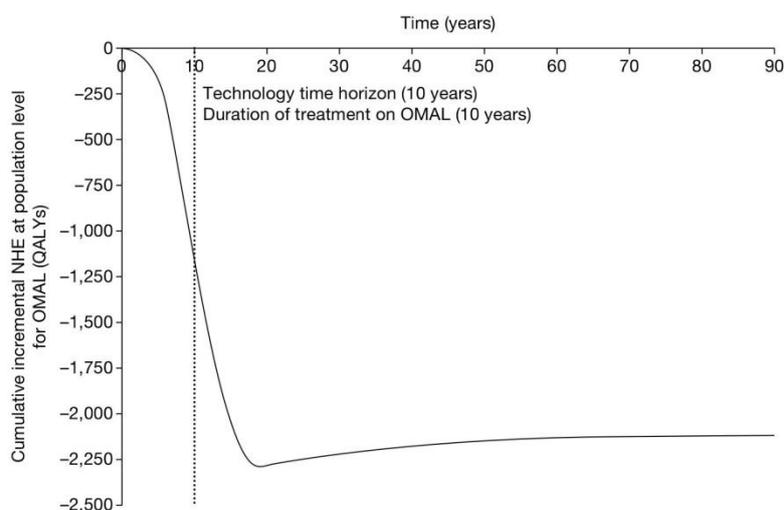
FIGURE 75 Cumulative incremental NHEs of OMAL for the subgroup population.

Point 2: Are there significant irrecoverable costs?

Omalizumub does not exhibit significant irrecoverable opportunity costs (see *Point 2: Are there significant irrecoverable costs?*). There are no additional irrecoverable costs associated with the subgroup population.

Point 3: Does more research seem worthwhile?

Omalizumub is not expected to be cost-effective based on existing evidence in the high-risk subgroup. Although the ICER is lower than in the base-case population, it is still significantly higher than the threshold. An assessment of the likely consequences of approving standard therapy when OMAL might be better is shown in *Table 119*. The results of PSA suggest that even at a threshold of £30,000 per QALY the probability that OMAL is cost-effective is very small (0.013) and the upper bound on the gains from more research is very limited (5.99 QALYs). Therefore, even after an analysis of subgroups OMAL is not expected to be cost-effective and more research does not seem worthwhile.

Table 120 shows the expected consequences of uncertainty in the subgroup with the alternative scenarios (see *Point 3: Does more research seem worthwhile?*, *Alternative scenarios*). Even under the most favourable assumptions, a decision to reject OMAL does not appear uncertain. There are only very small consequences of uncertainty or value to be gained by more research (0.624 QALYs) under the most favourable assumption for OMAL.

Types and categories of guidance resulting from subgroup analysis

The subgroup analysis was considered in the same way as for the base-case population; it started at point 1 of the checklist and worked through the same sequence of assessments as in the primary analysis. For the high-risk subgroup, OMAL is not expected to be cost-effective and there is no uncertainty. OMAL can be rejected at this point and no further assessment is required. *Table 121* summarises the category and type of guidance for OMAL in the high-risk subgroup.

TABLE 119 Expected consequences of uncertainty for OMAL in the subgroup

| Treatment | ICER (£/QALY) | Cost-effectiveness threshold £20,000 per QALY | | | Cost-effectiveness threshold £30,000 per QALY | | |
		Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)	Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
OMAL + standard care	69,463	−2175 (−44)	0.000	0.000 (0)	−1157 (−35)	0.013	5.99 (0.12)
Standard care		–	1.000		–	0.987	

TABLE 120 Expected consequences of uncertainty in the subgroup with the alternative scenarios

Additive scenario	ICER (£/QALY)	Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
Base-case subgroup	69,463	−2175 (−44)	0.000	0.000 (0)
+ 20-year treatment duration	59,096	−3399 (−68)	0.000	0.000 (0)
+ utility difference in day-to-day symptoms for all ages	53,269	−3171 (−63)	0.000	0.000 (0)
+ mortality rates owing to clinically significant severe exacerbations increased by 100%	44,921	−2861 (−57)	0.002	0.624 (0.01)

TABLE 121 Type and category of guidance for the OMAL high-risk subgroup

Assessment	1	2	3	4	5	6	7	Guidance
12	No	No	No	–	–	–	–	Reject[4]

Appendix 10

Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis

Introduction

Following a MTA appraisal (TA199[125] in 2010), the use of biologic treatment with etanercept, infliximab and adalimumab was recommended by NICE for patients with active and progressive PsA. However, the guidance also recommended that treatment should start with the least expensive biologic, taking account of dose, route of administration and price. This guidance updated an earlier MTA appraisal in 2006 (TA104[126]), which had recommended etanercept and restricted guidance on the use of infliximab to only those patients shown to be either intolerant or contraindicated to etanercept.[a]

In considering the results of the cost-effectiveness analysis, the NICE Appraisal Committee expressed concern about any differential effectiveness between treatments (specifically adalimumab and etanercept, which are biologically similar). They felt that a 'class effect' may be a more reasonable assumption. A scenario making this assumption (applying the same initial response and HAQ gain to adalimumab and etanercept) was therefore presented at the second appraisal meeting. In addition, in section 6 of TA199 the importance of data on long-term outcomes and adverse events from patient registries was highlighted. Therefore, PsA is not an example of AWR at FAD but an example of AWR considered during appraisal.

The analysis in this case study is from the standpoint of TA199 using the updated model that included new evidence and adalimumab as an additional comparator. At this point NICE guidance recommended etanercept and so the first question posed in the checklist can be interpreted as whether or not the other technologies available (infliximab, adalimumab or palliative care) are expected to be cost-effective compared with etanercept.

Background to the case study

Interventions and population

The three biologics (etanercept, infliximab and adalimumab) are compared with palliative care. No sequencing of treatments is considered; instead, if patients withdraw from any of the active treatments they will instead receive palliative care for the remainder of their lifetime. If patients do not withdraw from treatment (because of adverse events or lack of efficacy) they will remain on active treatment for the remainder of their lifetime, that is, PsA is a chronic disease.

The population of interest includes those patients with active and progressive PsA who have an inadequate response to standard treatment, including two conventional DMARDs (BSR criteria).[195] There is little information about the incidence of PsA in the UK; therefore, in the absence of a better alternative, based on clinical expert opinion the annual incidence of PsA was estimated at one-third of that for rheumatoid arthritis,[196] that is, 12,000 per year. As patients are unlikely to switch treatments once well controlled, a prevalent population was not considered relevant.

Evidence on clinical effectiveness

The review of clinical effectiveness identified six RCTs comparing one of the three biologics with placebo.[197–231] To synthesise all of these trials a Bayesian mixed-treatment comparison[232] was used to generate the following parameters for the decision model (for each biologic and for palliative care): probability of PsARC response at 12 weeks, probability of achieving Psoriasis Area Severity Index (PASI) 50, 75 and 90 responses at 12 weeks, and the associated HAQ for PSARC responder/non-responder.[233]

The trials, however, are limited in follow-up and therefore the longer-term effectiveness of the biologics and the impact on HAQ scores of withdrawing from biologics are not available. Expert elicitation was used in the model to generate estimates of the progression of HAQ for patients continuing on biologics and the progression of HAQ following withdrawal from treatment (rebound).[233] *Table 122* shows the probabilistic input parameters used in the model and their distributions.

Decision model

A probabilistic cohort model was developed to estimate costs and QALYs and explore any uncertainties.

These processes are inherently non-linear in terms of the relationship between model inputs and outputs (costs and QALYs) (see *Appendix 11, Model linearity and correlation between parameters*). The model was run for a period of 40 cycles (time horizon of the model, 3-monthly cycles) and probabilistic analysis was used (Monte Carlo simulation was run for 5000 iterations). The discount rate was 3.5%.

The initial response of the drug is defined in the model using PsARC for joints and PASI 75 for psoriasis; these parameters were estimated using a Bayesian evidence synthesis. Given a response (or no response) there is an associated impact on functional status, defined using the HAQ, for the arthritis aspect of the disease. The expected change in PASI is modelled as a constant. The model was implemented in R.

Patients can withdraw from biologics at any point after the initial response phase (first 12 weeks). If they withdraw during the initial response phase this is regarded as a 'no response'. Patients will receive palliative care after withdrawal at any stage. On withdrawing from treatment, it is assumed that mean PASI returns to its initial score at baseline (rebound equal to initial gain). There is considerable uncertainty about change in HAQ associated with withdrawal (rebound). Previous modelling work assumed that rebound of HAQ follows either of two alternative scenarios, with no data to inform which scenario is the more likely: rebound equal to initial gain and rebound equal to natural history. In TA199[125] this rebound assumption is informed by an expert opinion elicitation exercise conducted with five experts.

Key features and possible pathways

Biologics are not associated with any large upfront costs of treatment as treatment costs are incurred on a monthly basis according to whether or not the patient is still receiving treatment. The gains in quality-of-life benefits are also apparent throughout, with an initial gain in quality of life (through HAQ and PASI) associated with those who are treatment responders, as assessed at 3 months. The technology does, however, have a number of interesting characteristics, which can be used to explore the implications of the principles and assessments outlined in *Chapter 5*:

■ Etanercept, infliximab and adalimumab all have potentially significant irrecoverable costs because of the high per-patient treatment costs combined with a chronic condition in which treatment decisions are not irreversible. PsA is a chronic condition in which patients

TABLE 122 Parameters of the decision model for PsA

Description	Variable name	Mean	SE	Distribution	Source
Change in utility for 1 unit change in HAQ	U_HAQ	−0.298	0.006	Normal	Rodgers *et al.* 2001[233]
Change in utility for 1 unit change in PASI	U_PASI	−0.004	0.0003	Normal	Rodgers *et al.* 2001[233]
Change in cost for 1 unit change in HAQ	C_HAQ	187	21	Normal	Kobelt *et al.* 2002[234]
3-month cost for mild-to-moderate psoriasis if uncontrolled by biologics	C_psoriasis	198	9	Normal	Department of Health 2009[235]
Change in HAQ while not on treatment per 3-month period	LT_NH_HAQ	0.018	0.007	Gamma[236]	Norfolk Arthritis Register (NOAR) database[236]
Log-withdrawal rate from biologics per year	Withdraw	−1.823	0.2044	Normal	Registry data[233]
Probability of PsARC response on placebo	PSARC_NH	0.249	0.0384	Beta	Evidence synthesis
Change in HAQ given a PsARC response on placebo	HAQ_NH	−0.218	0.0465	Normal	
Probability of PASI 50 response on placebo	PASI_50_NH	0.130	0.021	Beta	
Probability of PASI 75 response on placebo	PASI_75_NH	0.044	0.009	Beta	
Probability of PASI 90 response on placebo	PASI_90_NH	0.016	0.004	Beta	
Probability of PsARC response on biologic_etanercept	PSARC_treat_E	0.713	0.071	Beta	
Probability of PsARC response on biologic_infliximab	PSARC_treat_I	0.795	0.058	Beta	
Probability of PsARC response on biologic_adalimumab	PSARC_treat_A	0.587	0.072	Beta	
Change in HAQ in first 3 months given no PsARC response of biologic_etanercept	HAQ_noPSARC_treat_E	−0.185	0.102	Beta	
Change in HAQ in first 3 months given no PsARC response of biologic_infliximab	HAQ_noPSARC_treat_I	−0.190	0.073	Beta	
Change in HAQ in first 3 months given no PsARC response of biologic_adalimumab	HAQ_noPSARC_treat_A	−0.064	0.064	Beta	
Change in HAQ in first 3 months given PsARC response of biologic_etanercept	HAQ_treat_E	−0.623	0.095	Beta	
Change in HAQ in first 3 months given PsARC response of biologic_infliximab	HAQ_treat_I	−0.652	0.072	Beta	
Change in HAQ in first 3 months given PsARC response of biologic_adalimumab	HAQ_treat_A	−0.423	0.061	Beta	
Probability of PASI 50 response on biologic_etanercept	PASI_50_treat_E	0.4026	0.0916	Beta	
Probability of PASI 50 response on biologic_infliximab	PASI_50_treat_I	0.9128	0.0374	Beta	
Probability of PASI 50 response on biologic_adalimumab	PASI_50_treat_A	0.7383	0.0853	Beta	
Probability of PASI 75 response on biologic_etanercept	PASI_75_treat_E	0.1768	0.0586	Beta	
Probability of PASI 75 response on biologic_infliximab	PASI_75_treat_I	0.7687	0.0795	Beta	
Probability of PASI 75 response on biologic_adalimumab	PASI_75_treat_A	0.4772	0.1085	Beta	
Probability of PASI 90 response on biologic_etanercept	PASI_90_treat_E	0.0737	0.0292	Beta	
Probability of PASI 90 response on biologic_infliximab	PASI_90_treat_I	0.5571	0.1088	Beta	
Probability of PASI 90 response on biologic_adalimumab	PASI_90_treat_A	0.2571	0.0863	Beta	

SE, standard error.

are unlikely to switch treatment once maintained. Switching will happen only if patients withdraw from treatment because of adverse events or lack of efficacy.

■ There is uncertainty about the effectiveness of biologics. A separate scenario was assessed that assumed a class effect for the biologics etanercept and adalimumab. This provides an opportunity to examine the role of exploring separate extreme scenarios to characterise uncertainty.

The possible pathways through the algorithm are reported in *Figure 32* in *Appendix 4*. The alternatives to etanercept (adalimumab, infliximab and palliative care) are not expected to be cost-effective. There are, however, a number of uncertainties. These uncertainties might influence the category of guidance, for example OIR or AWR rather than reject.
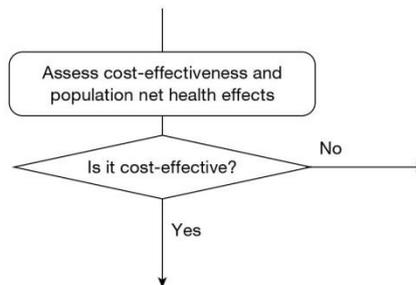
The following sections examine each of the seven points on the checklist relating to the possible sequence of assessments and decisions that lead to a particular category and type of guidance for PSA.

Is it cost-effective and what are the risks?

The judgements made at points 1 and 2 of the checklist are critical because, although neither leads directly to a particular category of guidance, they determine the subsequent path that might be taken, sometimes avoiding further and potentially complex assessments. For example, the absence of significant irrecoverable costs means that only four out of the 12 possible pathways require all seven assessments to be made (see *Appendix 4*).

Point 1: Is it expected to be cost-effective?

The sequence of assessments starts with cost-effectiveness and the expected impact on population NHEs, that is, at the following point in the algorithm:

This requires an assessment of expected cost-effectiveness based on the balance of the evidence and analysis currently available. Methods to estimate expected cost-effectiveness are well established within the NICE appraisal process and are extensively described in the *Guide to Methods of Technology Appraisal*.[1,b] Commonly, expected cost-effectiveness is summarised and presented using ICERs. Equivalently, but more usefully in this context, cost-effectiveness can be expressed in terms of expected NHEs, which can be expressed per patient treated or for a population of patients. This is especially important when later assessments require a comparison of benefits to current or future patient populations and when assessing the significance of irrecoverable costs (see *Cost-effectiveness at the population level*). All of the information required to express expected cost-effectiveness in these ways is already available during appraisal.
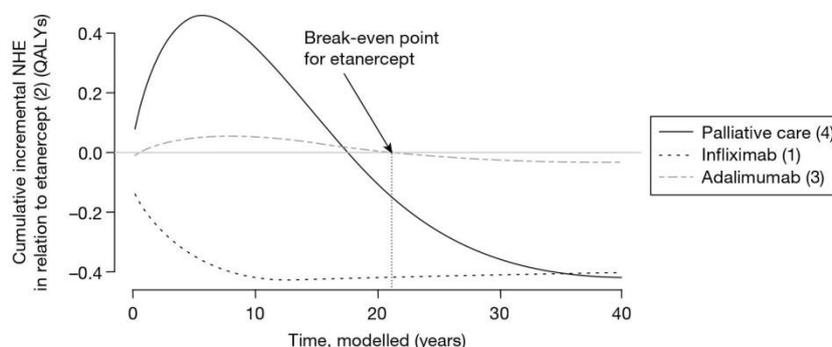
Cost-effectiveness at the patient level

Estimates of the expected NHS costs and QALYs for each patient treated over an appropriate time horizon – the 'patient time horizon'[t] – can be summarised as an ICER, which must be compared with a cost-effectiveness threshold to judge cost-effectiveness. Equivalently, this can be expressed as the per-patient NHE of each intervention, that is, the difference between any health gained and health forgone elsewhere.

For the PsA case study, the treatment already recommended by NICE (etanercept at the time of TA199[125]) is expected to be cost-effective (*Table 123*).[d] Other alternatives to etanercept (adalimumab, infliximab and palliative care) are not expected to be cost-effective. Although adalimumab is less effective than etanercept, it is also cheaper; however, the resource savings it offers do not compensate for the reduction in health benefits (for an ICER of around £20,000 per QALY). Infliximab is more effective but it is also much more expensive. These increased costs are, however, not justified by the QALY gains.

The investment profiles of the alternatives to etanercept, illustrated in *Figure 76* relative to etanercept, differ in appearance. All of the biologic treatments for PsA (etanercept, infliximab and adalimumab) have high initial costs, which are only gradually compensated for by later health benefits. Palliative care offers a positive NHE initially compared with etanercept. As the time horizon of the model increases the NHE of palliative care compared with etanercept becomes negative; this happens at 19 years. Adalimumab also offers positive NHEs in the short run compared with etanercept (lower costs) so it is only at 21.25 years that etanercept is expected to offer the highest NHE (all others have negative NHEs relative to this). Infliximab is always a risky strategy and does not break-even within a reasonable modelled time horizon (40 years).

TABLE 123 Expected cost-effectiveness in PsA per patient treated

Treatment	Cost (£)	QALYs	ICER (£/ QALY)	Cost-effectiveness threshold £20,000 per QALY NHE, QALYs (£)	Cost-effectiveness threshold £30,000 per QALY NHE, QALYs (£)
1: Infliximab	90,343	7.269	60,965	2.752 (5504)	4.258 (8516)
2: Etanercept	78,150	7.069	17,733	3.161 (6322)	4.464 (8928)
3: Adalimumab	72,972	6.777	14,622	3.129 (6258)	4.345 (8690)
4: Palliative care	51,800	5.329	–	2.739 (5478)	3.602 (7204)

FIGURE 76 Cumulative incremental NHEs in PsA over the patient time horizon.

Etanercept, infliximab and adalimumab all therefore have potentially significant irrecoverable costs because of the high per-patient treatment costs combined with a chronic condition in which treatment decisions are not irreversible.

Cost-effectiveness at the population level

Per-patient NHEs can also be expressed for the population of current and future patients. This requires information about prevalence and future incidence of the target population (already required in appraisal). It also requires a judgement about the time horizon over which the technology will be used. This 'technology time horizon' ought to reflect the period over which the technology is likely to be part of clinical practice and generate the expected NHEs.[e] An estimate of the scale of the total population NHEs and how they cumulate over time is important for subsequent assessments, including (1) when the NHEs for current patient populations must be compared with the benefits to future patients and (2) when the treatment decision can be changed so the irrecoverable costs of initially negative NHEs become significant.

In the PsA case study there is no prevalent population eligible for adalimumab, infliximab or palliative care if it is approved. This is because the prevalent population will not switch treatments whilst maintained on their current therapy. They will switch only if they relapse (withdraw from treatment because of adverse events or lack of efficacy). The total population NHEs, assuming that the technology will be used to treat only incident patients (12,000 per year) over 10 years, are reported in *Table 124*. The expected cost-effectiveness is unchanged (the ICER is the same as that in *Table 121*) but the incremental NHEs, although small per patient, are more significant at a population level.

The investment profile for PsA when patients are treated over 10 years is illustrated in *Figure 77*. The NHEs for infliximab, adalimumab and palliative care are calculated relative to etanercept (standard NHS care). At a population level it is not until 26 years (rather than 21.25 years at a patient level) that initial losses are compensated for by later gains and etanercept breaks even (all other NHEs are negative relative to etanercept[213]). In other words, etanercept appears a more risky investment when evaluated at a population rather than at an individual level. This is because, although each patient treated with etanercept is expected to offer the same profile of NHEs shown in *Figure 76*, the negative NHEs associated with patients who are incident and treated in year 10 will not be offset by later gains until year 26. The incremental NHEs and break-even points when evaluated at a population level are reported in *Table 125* over different time horizons. These are reported for adalimumab compared with palliative care, palliative care compared with etanercept, palliative care compared with infliximab and adalimumab compared with etanercept. The table shows that as the technology time horizon increases treatment with biologics becomes a more risky prospect and the NHE of palliative care remains

TABLE 124 Expected cost-effectiveness of PsA for the population

| Treatment	Cost (£M)	QALYs	ICER (£/ QALY)	Cost-effectiveness threshold £20,000 per QALY		Cost-effectiveness threshold £30,000 per QALY	
				NHE, QALYs (£M)	Incremental NHE, QALYs (£M)	NHE, QALYs (£M)	Incremental NHE, QALYs (£M)
1: Infliximab	9332	750,874	60,965	284,288 (5686)	−42,227 (−845)	439,816 (13,194)	−21,237 (−637)
2: Etanercept	8072	730,128	17,733	326,515 (6530)	–	461,053 (13,832)	–
3: Adalimumab	7537	700,037	14,622	323,162 (6463)	−3353 (−67)	448,787 (13,464)	−12,266 (−368)
4: Palliative care	5351	550,422	–	282,894 (5658)	−43,621 (−872)	372,070 (11,162)	−88,983 (−2669)
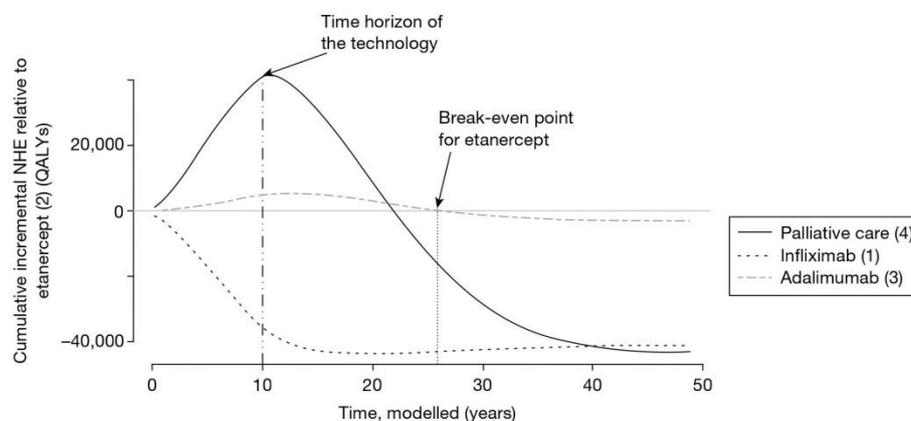
FIGURE 77 Cumulative incremental NHEs in PsA for the population (in relation to standard care, etanercept).

TABLE 125 Incremental NHEs in PsA over a range of technology time horizons

Technology time horizon	Treatment	Incremental NHE (compared with etanercept), QALYs	Break-even point (years)		
			vs 4: Palliative care	vs 2: Etanercept	vs 1: Infliximab
5 years	1: Infliximab	−22,925 (−459)			
	2: Etanercept	–			
	3: Adalimumab	−1820 (−25)	17	22	
	4: Palliative care	−23,681 (−710)		18.25	36.25
10 years	1: Infliximab	−42,227 (−1267)			
	2: Etanercept	–			
	3: Adalimumab	−3353 (−101)	18.5	22	
	4: Palliative care	−43,624 (−1309)		18.75	36.25
15 years	1: Infliximab	−58,480 (−1754)			
	2: Etanercept	–			
	3: Adalimumab	−4643 (−139)	19	22.5	
	4: Palliative care	−60,410 (−1812)		19.5	36.5
20 years	1: Infliximab	−72,163 (−2165)			
	2: Etanercept	–			
	3: Adalimumab	−5730 (−172)	19.5	23.75	
	4: Palliative care	−77,545 (−2326)		20	37

positive for longer. At a technology time horizon of 10 years it is only at 18.75 years that current NHS care (etanercept) breaks even against palliative care and not until 22 years that it is better than adalimumab.

Point 2: Are there significant irrecoverable costs?

The second point on the checklist requires (1) an assessment of whether or not there are irrecoverable costs and (2) a judgement of their potential significance, that is, at the following point in the algorithm:
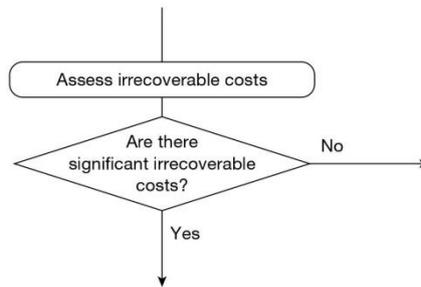
Identifying the irrecoverable costs

Irrecoverable costs are those that once committed cannot be recovered if guidance is changed at a later date (see *Chapter 5* and *Appendix 11, Accounting for the investment profile of displaced interventions*). The impact of irrecoverable costs will tend to be greater if they represent a greater proportion of the total costs and if guidance is more likely to change and to change in the near future.

Even in the absence of capital costs of equipment and facilities, NHEs accumulate over time at both a patient and a population level. The analysis in *Point 1: Is it expected to be cost-effective?* indicates a common pattern of initially negative NHEs that are only gradually offset by positive NHE in later periods. Therefore, AWR may commit opportunity costs of negative NHEs that are irrecoverable.

Are they likely to be significant?

Whether or not irrecoverable costs are significant (i.e. might influence guidance) depends critically on whether guidance is likely to change and whether that is more likely in the near or in the distant future. That will depend on whether research is likely to be undertaken and when it is likely to report, as well as on other events that might occur, for example a change in price following patent expiry. These are assessed later, in points 5 and 6; however, the *potential* significance of any irrecoverable costs can be assessed at this point.

The potential significance of the investment profiles of NHEs depends on if treatment decisions for individual patients are irreversible, which in part depends on the nature of the disease. For example, in an acute condition the decision to treat a particular presenting patient with a technology cannot be reconsidered at a later date – it is irreversible with associated reversal costs. Of course, it is possible that the later benefits are not realised but it is also possible that they will realise more (the profiles of NHEs in *Figures 76* and *77* are the average over these possibilities). Similarly, the possibility that guidance might change in the future (e.g. research suggests that the longer-term benefits will not offset initial losses) will not influence the irreversible decision to treat a presenting patient with a technology that is expected to be cost-effective prior to the research reporting.

Psoriatic arthritis is a chronic condition in which the decision to treat a particular patient cannot be changed at some later date (decisions are assumed irreversible). Therefore, the NHS will commit to irrecoverable costs by deciding not to use technologies not expected to be cost-effective now [if they have less irrecoverable costs than standard care (etanercept)]. This

is the case for adalimumab and palliative care for which the NHEs are initially higher than for etanercept (*Figure 77* shows less irrecoverable costs for adalimumab and palliative care initially). For infliximab, however, the NHEs are lower than for etanercept; therefore, deciding not to use infliximab will reduce the loss in NHEs (as the initial loss in NHE is lower for etanercept). In deciding if these costs are likely to be significant, however, it is important to note that, in the example of PsA, decisions cannot be reversed (patients remain on treatment if they are responding). Therefore, patients will still experience the positive NHEs that occur later in the modelled time horizon (resulting from the recommended treatment, etanercept). In conclusion, therefore, there are no significant investment or reversal costs for this case study.

Types and categories of guidance resulting from points 1 and 2

Points 1 and 2 of the checklist do not lead directly to a category or type of guidance. The sequence of assessments and decisions, which ultimately lead to guidance, starts with cost-effectiveness, expected impact on population NHEs and significance of irrecoverable costs. In the case of PsA, alternatives to etanercept are not expected to be cost-effective and there are no significant irrecoverable costs. *Table 126* summarises the categories and types of guidance that could ultimately result from the sequence of assessments up to point 2.

Is further research required?

Point 3: Does more research seem worthwhile?

The third point on the checklist requires an assessment of the potential benefits of conducting further research, that is, at the following point in the algorithm:
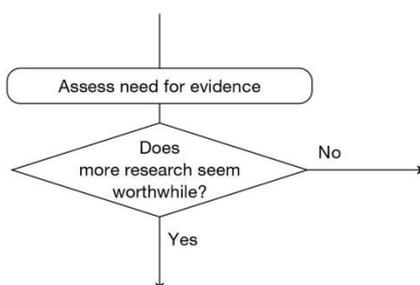
This requires judgements about (1) how uncertain a decision to approve or reject a technology might be based on the estimates of expected cost-effectiveness and (2) whether or not the scale of the likely consequences of this uncertainty might justify further research.

The judgements made at points 3 and 4 of the checklist are critical because if more research is not judged to be worthwhile then no further assessments are required (unless there are significant

TABLE 126 Types and categories of guidance that could ultimately result from assessments 1 and 2

Assessment	1	2	3	4	5	6	7	Guidance
7	No	No	Yes	Yes	Yes/no	Yes	–	OIR[2]
8	No	No	Yes	Yes	Yes/no	No	–	Reject[1]
9	No	No	Yes	No	Yes/no	Yes	Yes	AWR[2]
10	No	No	Yes	No	Yes/no	Yes	No	Reject[2]
11	No	No	Yes	No	Yes/no	No	–	Reject[3]
12	No	No	No	–	–	–	–	Reject[4]

irrecoverable costs; see *Appendix 4*). If research is worthwhile then what type of evidence is needed and whether the research required to generate it can be conducted while the technology is approved will determine whether AWR or OIR is a possibility.

Assessing the consequences of uncertainty

Infliximab, palliative care and adalimumab are not expected to be cost-effective compared with standard care (see *Tables 123* and *124*) but the estimates of costs and QALYs are uncertain so there is a chance that a decision to reject these treatments based on existing evidence will be incorrect, that is, infliximab, palliative care or adalimumab might offer greater NHEs. Some assessment of the likely consequences of rejecting infliximab, palliative care and adalimumab might be based on the difference in expected NHEs, that is, the expected incremental population NHEs reported in *Table 124*. However, as there are more than two alternatives, a judgement is required about the chance that using standard care (etanercept) is incorrect and if so which of the other three alternatives are likely to offer higher NHEs and how much higher. In other words, for decisions involving multiple alternatives, a judgement is required on the level of uncertainty surrounding the decision, how this uncertainty is distributed across the various alternatives and what the consequences are likely to be.

The simplest approach could be to take the difference in expected NHEs between adalimumab, palliative care or infliximab and standard NHS care and weight it by a judgement of the probability that the decision is correct, that is, allocate the consequences of an incorrect decision to NHS standard care. For example, if the decision was judged to be 100% certain (the probability that the decision is correct would be 1) then there are no consequences and so there would be nothing to be gained by more research; however, as the decision becomes more uncertain, the expected consequences (and hence the potential value of more research) increase.

In the presence of multiple treatment alternatives there are other options than to allocate all of the consequences of an erroneous decision to NHS standard care. We have examined four alternative ways:

1. The consequences are assigned to NHS standard care (etanercept).
2. The consequences are assigned to adalimumab, the next best treatment.
3. Equal shares of the consequences are assigned to treatments other than NHS standard care (etanercept).
4. The consequences are assigned to treatments other than NHS standard care (etanercept) based on the probability of the alternative treatments being cost-effective from PSA. Note that this uses the PSA only to inform the magnitude of the consequences of a wrong decision; a judgement on the likelihood of adopting etanercept being the incorrect decision is still required.

The results of these alternative analyses are illustrated in *Table 127* in which a judgement about the probability that a decision based on expected cost-effectiveness is correct translates into expected consequences based on expected incremental population NHEs.

The expected consequences of a wrong decision are high in scenarios (2), (3) and (4). There are no consequences of a wrong decision if all of the error gets assigned to treatment 2 (NHS standard care) because treatment 2 is the most cost-effective treatment (NHEs calculated relative to this).

The judgement regarding how uncertain a decision might be can be informed by the PSA already used to estimate costs and QALYs. The probabilities that each of the three alternatives is cost-effective are reported in *Table 128*. This shows that although infliximab, palliative care and adalimumab are not considered to be cost-effective there is a non-zero probability that this

TABLE 127 Expected consequences of uncertainty based on a weighting of average NHEs

Error probability, p	Cost-effectiveness threshold £20,000 per QALY				Cost-effectiveness threshold £20,000 per QALY			
	NHEs for the population, QALYs				NHEs for the population, £			
	1: Infliximab	2: Etanercept	3: Adalimumab	4: Palliative care	1: Infliximab	2: Etanercept	3: Adalimumab	4: Palliative care
0	0	0	0	0	0	0	0	0
0.01	0	34	297	133	0	670,500	5,946,720	2,669,666
0.05	0	168	1487	667	0	3,352,540	29,733,600	13,348,331
0.1	0	335	2973	1335	0	6,705,080	59,467,220	26,696,661
0.25	0	838	7433	3337	0	16,762,700	148,668,040	66,741,653
0.5	0	1676	14,867	6674	0	33,525,420	297,336,080	133,483,307
0.75	0	2514	22,300	10,011	0	50,288,120	446,004,120	200,224,960
0.99	0	3319	29,436	13,215	0	66,380,320	588,725,440	264,296,947
1	0	3353	29,734	13,348	0	67,050,820	594,672,160	266,966,613

TABLE 128 Expected consequences of uncertainty in PsA

Treatment	ICER (£/ QALY)	Cost-effectiveness threshold £20,000 per QALY			Cost-effectiveness threshold £30,000 per QALY		
		Incremental net health effect,[a] QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)	Incremental net health effect,[a] QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
1: Infliximab	60,965	−42,240 (−844)	0.012	35,341 (707)	−21,279 (−425)	0.113	18,079 (542)
2: Etanercept	17,733	3306 (66)	0.443		12,292 (245)	0.502	
3: Adalimumab	14,622	40,284 (805)	0.146		76,746 (1534)	0.103	
4: Palliative care	–	–	0.399		–	0.280	

a The mean additional population NHE of moving from the least to the most effective alternative, i.e. the incremental NHE of infliximab compared with palliative care, is the sum of these.

judgement is incorrect (they are indeed cost-effective). At a threshold of £20,000, the probabilities that the three treatments are cost-effective are 0.012, 0.146 and 0.399 for infliximab, adalimumab and palliative care, respectively (*Figure 78*). At a threshold of £30,000, the probabilities that the three treatments are cost-effective are 0.113, 0.103 and 0.280 for infliximab, adalimumab and palliative care, respectively.

The previous analysis assumed that the differences in NHEs when etanercept is not cost-effective are the averages; however, this is not necessarily true and estimates based on means may substantially under- or overestimate the expected consequences of uncertainty (see *Appendix 11, Why the consequences of uncertainty differ from mean incremental effects*). To accurately record the differences between the NHEs of the alternative treatments and the probability that these errors will occur, the results from the PSA can be used. The distribution of consequences is illustrated in *Figure 79*. Most commonly (44.3%) there are no consequences because etanercept is the correct decision. When it is not, there is a greater chance of relatively small consequences (19% are < 27,250 QALYs), which occur predominantly when adalimumab offers the highest NHEs. There is a small chance of larger consequences (< 0.1% chance that they are > 190,750 QALYs), when palliative care offers the highest NHEs, that is, there remains important uncertainty about the cost-effectiveness of treatment. The expected consequence of uncertainty (35,341 QALYs or £706M) is simply the average over this distribution. Again, this can be interpreted as an estimate of the population NHEs that could be gained over the time horizon of this technology if the uncertainties could be immediately resolved. A judgement at this point
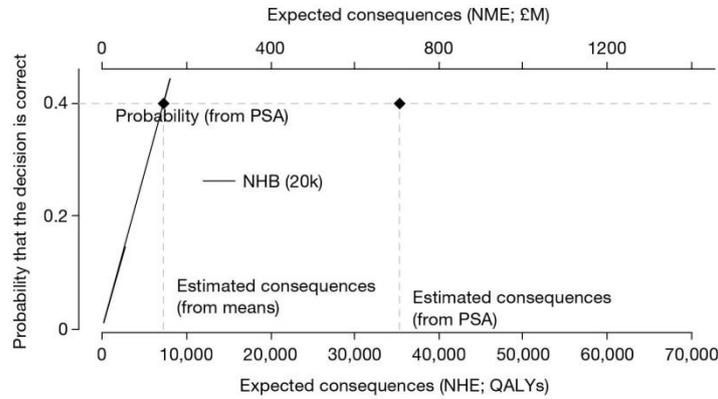
FIGURE 78 Probability that etanercept is cost-effective and the consequences of uncertainty.
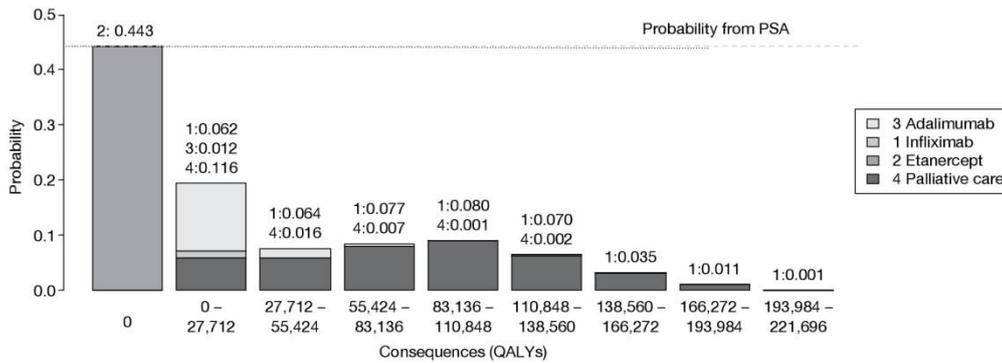

FIGURE 79 Distribution of the consequences of uncertainty in PsA.

that more research might be worthwhile seems reasonable, as the potential benefits exceed the likely costs.

The time horizon over which evidence generated by research about a technology might be valuable may be longer (or shorter) than the period over which the technology is used. In this case the technology time horizon is assumed to be equal to the time horizon for the benefits of research, that is, 10 years. *Table 129* shows the expected consequences for different technology time horizons. The consequences increase with the technology time horizon and will also increase with the size of the patient population.

Alternative scenarios

There are often alternative views about the quality and relevance of evidence as well as other assumptions that might be made when estimating expected costs and QALYs. These are commonly presented as separate scenarios, with estimates of costs and QALYs presented for each. Much of the deliberation by the NICE Appraisal Committee often surrounds the scientific value judgements required to judge the credibility of the alternative assumptions represented by the scenarios. The type of probabilistic analysis reported represents the uncertainty within

each scenario and will be sufficient to indicate the potential benefits of research when only one scenario is regarded as credible. However, when more than one scenario might be credible and carry some 'weight', there will be uncertainty *between* as well as *within* scenarios. The 'weighting' of scenarios can be made explicit by assigning probabilities to represent how credible each is believed to be. The weighted average of costs and QALYs across scenarios can easily be calculated. It is also tempting to take a simple weighted average of the expected consequences of uncertainty across these scenarios as well; however, a simple weighted average may under- or overestimate the combined consequences of uncertainty within and between scenarios. The correct estimate requires the probabilities (weights) to be applied directly to the simulated output from PSA rather than to the mean values. Although this does not require additional simulation and is quick and easy to implement, it does require either that the probabilities are made explicit in advance or that estimates are presented for a range of probabilities that might represent the judgement of the Appraisal Committee following deliberation.

An alternative assumption of a common class effect across the three biologics was considered in the PsA case study (scenario B) but this was judged less credible than the analysis that allowed differential effects (scenario A). The alternative scenario made etanercept less likely to be cost-effective (*Table 130*) and increased the expected consequences of uncertainty from 34,930 to 38,521 QALYs (*Figure 80*). In this case a simple weighted average of expected consequences based on the probability assigned to each scenario is, in general, lower than the correct estimate of expected consequences based on the output from PSA.

TABLE 129 Expected consequences of uncertainty for different technology time horizons

Technology time horizon, years	Expected consequences, QALYs (£M)
5	19,186 (3837)
10	35,341 (7068)
15	48,943 (9788)
20	60,396 (12,079)

TABLE 130 Expected consequences of alternative scenarios

| Treatment	ICER (£/QALY)	Cost-effectiveness threshold £20,000 per QALY			Cost-effectiveness threshold £30,000 per QALY		
		Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)	Incremental NHE, QALYs (£M)	Probability cost-effective	Expected consequences, QALYs (£M)
Scenario A							
1: Infliximab	60,965	−42,240 (−844)	0.012	35,341 (707)	−21,279 (−425)	0.113	18,079 (542)
2: Etanercept	17,733	3306 (66)	0.443		12,292 (245)	0.502	
3: Adalimumab	14,622	40,284 (805)	0.146		76,746 (1534)	0.103	
4: Palliative care	–	–	0.399		–	0.280	
Scenario B							
1: Infliximab	59,208	−42,349 (−846)	0.003	38,521 (770)	−21,071 (−421)	0.065	22,008 (660)
2: Etanercept	32,630	−309 (−6.1)	0.306		0	0.332	
3: Adalimumab	15,036	4451 (8.9)	0.293		89,451 (1789)	0.329	
4: Palliative care	–	–	0.398		–	0.274	
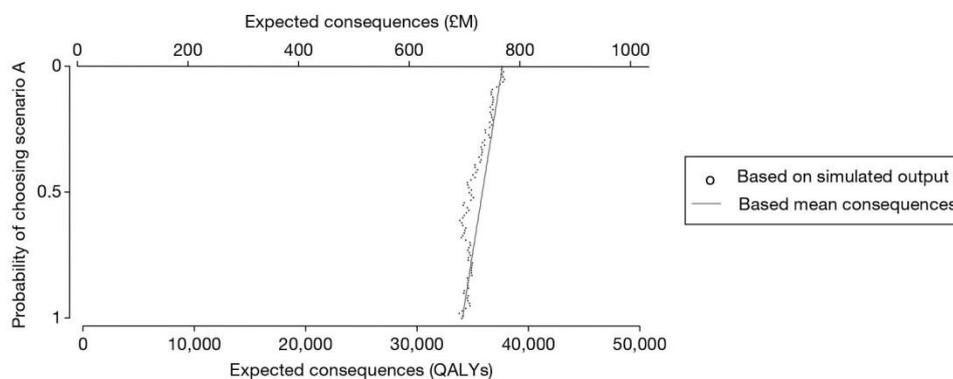
FIGURE 80 Expected consequences of uncertainty with alternative scenarios: PsA case study.

Point 4: Is research possible with approval?

The fourth point on the checklist requires an assessment of what type of evidence is needed and a judgement of whether or not the research required to generate it can be conducted while the technology is approved, that is, at the following point in the algorithm:
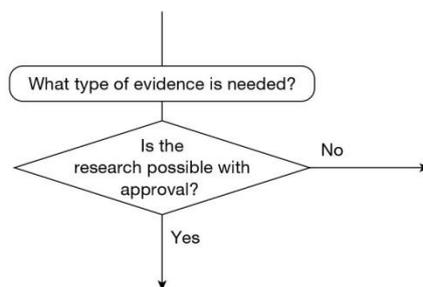
This requires judgements about (1) how important particular types of parameters are to estimates of costs and QALYs, (2) what values these parameters would have to take to change a decision based on expected cost-effectiveness, (3) how likely is it that parameters might take such values and (4) what would be the consequences if they did, that is, what might be gained in terms of population NHEs if the uncertainty in the values of these parameters could be immediately resolved.

Assessing the importance of parameters

The expected cost-effectiveness of infliximab, adalimumab and palliative care is based on the relationship between the input parameters (see *Table 122*) and outputs of costs and QALYs. A simple summary of the direction and strength of these relationships can be provided by the elasticity, that is, the proportionate change in NHEs owing to a 1% change in the value of a parameter (*Table 131*). Parameters with high elasticities (especially with respect to differences in NHEs) might be regarded as more 'important'.

Although these measures of importance are more instructive than a series of arbitrary one-way sensitivity analyses, they do not directly help the assessment of what values parameters must take to change decisions and how likely such values might be. A simple summary of the values that particular parameters must take to make each of the alternatives cost-effective can also be

TABLE 131 Elasticity for small changes (1%) in mean parameter values: PsA case study

		Parameter	Elasticity over the NHEs (QALYs) of				Elasticity over the incremental NHEs (QALYs) of	
			1: Infliximab	2: Etanercept	3: Adalimumab	4: Palliative care	2: Etanercept vs 4: Palliative care	2: Etanercept vs all
Natural history	1	LT_NH_HAQ	−0.15445	−0.08421	−0.2139	−0.04515	−0.0167	−0.24492
	2	PsARC_NH	–	–	–	−0.004	–	–
	3	HAQ_NH	–	–	–	−0.004	–	–
	4	PASI_50_NH	–	–	–	−0.003	–	–
	5	PASI_75_NH	–	–	–	−0.004	–	–
	6	PASI_90_NH	–	–	–	−0.004	–	–
	7	U_HAQ	−0.22571	−0.16120	−0.10134	−0.13878	−0.01850	−0.22297
	8	U_PASI	−0.15774	−0.09167	−0.02711	−0.05061	−0.01766	−0.25942
Treatment effect	9	Withdrawal	−0.19082	−0.12534	−0.06514	−0.09596	−0.01781	−0.23562
	10	Prob_PSARC	−0.15254	−0.07236	−0.01781	–	−0.01478	−0.27423
	11	HAQ_noPSARC	−0.15338	−0.08665	−0.02275	–	−0.01750	−0.38879
	12	HAQ_resp	−0.15469	−0.08569	−0.02213	–	−0.01698	−0.37380
	13	PASI_50	−0.15314	−0.08650	−0.02213	–	−0.01752	−0.39059
	14	PASI_75	−0.15236	−0.08627	−0.02165	–	−0.01766	−0.39315
	15	PASI_90	−0.15330	−0.08665	−0.02235	–	−0.01752	−0.39095
Cost	16	C_psoriasis	−0.16242	−0.09599	−0.03224	−0.05698	−0.01757	−0.25331
	17	C_HAQ	−0.15838	−0.09263	−0.02796	−0.05167	−0.01774	−0.26011

provided (*Table 132*). It shows, for example, that if HAQ_resp is valued between −0.371 and −0.661 for adalimumab (treatment 3) or at −0.838 for infliximab (treatment 1), treatments other than etanercept (treatment 2) are deemed cost-effective. However, although instructive, such 'threshold values' do not indicate how likely it is that the threshold will be crossed.

The judgement about how likely it is that parameters might take values that will change the technology expected to be cost-effective can be informed by the results of the PSA. The distributions assigned to the parameters in PSA describe how uncertain the parameter estimates are, such that they ought to reflect the amount and quality of existing evidence. The probability that each parameter might take values that would lead to each of the alternatives being cost-effective is reported in *Table 133* for the uncertain parameters. This essentially decomposes the overall probabilities reported in *Table 130* into the contribution that each parameter makes. It indicates that it is uncertainty in the estimates of natural history of HAQ progression that contributes most to the probability of error associated with etanercept. At certain values of this parameter treatment 4 (palliative care) would be considered cost-effective. In addition, uncertainty regarding the treatment effect (measured using HAQ) also contributes to the probability of error. At certain values treatment 3 (adalimumab) would be considered cost-effective.

What type of evidence is needed?

An assessment of the likely consequences of the uncertainty described above is required. This assessment can directly inform the judgement of what evidence is needed and whether or not the type of research required to generate it will be possible with approval. The expected consequences of uncertainty associated with each parameter are reported in *Table 134*. This decomposes the overall expected consequences into the contribution that each parameter (or group of parameters) makes. Note that the overall expected consequences of uncertainty will not,

TABLE 132 Range of parameter values leading to the decision to adopt each of the treatments of interest

		Parameter	Mean value	1: Infliximab	2: Etanercept	3: Adalimumab	4: Palliative care
Natural history[a]	1	LT_NH_HAQ	0.018	–	0.014 to 0.044	–	0.004 to 0.014
	2	PsARC_NH	0.249	–	0.001 to 0.920	–	–
	3	HAQ_NH	−0.218	–	−0.961 to 0	–	–
	4	PASI_50_NH	0.130	–	0 to 0.934	–	–
	5	PASI_75_NH	0.044	–	0 to 0.688	–	–
	6	PASI_90_NH	0.016	–	0 to 0.530	–	–
	7	U_HAQ	−0.298	–	−0.279 to 0.315	–	–
	8	U_PASI	−0.004	–	−0.003 to 0.005	–	–
Treatment effect	9	Withdrawal	−1.823	–	−1.189 to 2.711	–	–
	10	Prob_PSARC[b]	E: 0.713	–	0.464 to 0.883	–	–
			A: 0.587				
			I: 0.795				
	11	HAQ_noPSARC	E: −0.190	–	−0.122 to 0.478	–	–
			A: −0.130				
			I: −0.194				
	12	HAQ_resp	E: −0.630	−0.838	−0.493 to 0.930	−0.371 to 0.661	–
			A: −0.103				
			I: −0.650				
	13	PASI_50	E: 0.403	–	0.120 to 0.738	–	–
			A: 0.738				
			I: 0.913				
	14	PASI_75	E: 0.177	–	0.044 to 0.434	–	–
			A: 0.477				
			I: 0.769				
	15	PASI_90	E: 0.074	–	0.015 to 0.206	–	–
			A: 0.257				
			I: 0.557				
Cost	16	C_psoriasis	198	–	129.11 to 250.30	–	–
	17	C_HAQ	187	–	166.38 to 222.44	–	–

A, adalimumab; E, etanercept; I, infliximab.
a   Means rebound equal to initial gain.
b   For parameters for which there are different values for each of the biologics (etanercept, adalimumab and infliximab) the ranges required to change the decision to adopt relate to the relevant treatment only, e.g. ranges −0.371 to −0.661 for HAQ response for treatment 3 relate to adalimumab.

in general, equal the sum of the expected consequences for each of the parameters (or groups of parameters) separately. This is because the overall consequences take account of the joint effect of uncertainty in all parameters simultaneously. Even if parameters are independent they will be related to differences in NHEs in different ways (see *Table 134*); sometimes the effect of uncertainty in one parameter may, to some extent, substitute for or complement the effect of uncertainty in the others. *Table 134* confirms that it is uncertainty regarding the natural history of progression and the treatment effect (HAQ gain for biologics) that contributes towards the value of further research. There is potentially the most to be gained by resolving this uncertainty through additional research (8694 and 1201 QALYs or £17M and £2.4M, respectively).

TABLE 133 Overall probabilities of observing parameter values leading to the decision to adopt each of the treatments of interest

		Parameter	1: Infliximab	2: Etanercept	3: Adalimumab	4: Palliative care
Natural history	1	LT_NH_HAQ	0.646	–	–	0.354
	2	PsARC_NH	–	1	–	–
	3	HAQ_NH	–	1	–	–
	4	PASI_50_NH	–	1	–	–
	5	PASI_75_NH	–	1	–	–
	6	PASI_90_NH	–	1	–	–
	7	U_HAQ	–	1	–	–
	8	U_PASI	–	1	–	–
Treatment effect	9	Withdrawal	–	1	–	–
	10	Prob_PSARC	–	1	–	–
	11	HAQ_noPSARC	–	1	–	–
	12	HAQ_resp	0	0.812	0.187	–
	13	PASI_50	–	1	–	–
	14	PASI_75	–	1	–	–
	15	PASI_90	–	1	–	–
Cost	16	C_psoriasis	–	1	–	–
	17	C_HAQ	–	1	–	–

TABLE 134 Consequences of uncertainty associated with parameter values

			Expected consequences at the population level (based on PSA), decomposed by treatment choice, QALYs				
		Parameter	1: Infliximab	2: Etanercept	3: Adalimumab	4: Palliative care	Overall
Natural history	1	LT_NH_HAQ	–	–	–	8694	8694
	2	PsARC_NH	–	–	–	–	–
	3	HAQ_NH	–	–	–	–	–
	4	PASI_50_NH	–	–	–	–	–
	5	PASI_75_NH	–	–	–	–	–
	6	PASI_90_NH	–	–	–	–	–
	7	U_HAQ	–	–	–	–	–
	8	U_PASI	–	–	–	–	–
Treatment effect	9	Withdrawal	–	–	–	–	–
	10	Prob_PSARC	–	–	–	–	–
	11	HAQ_noPSARC	–	–	–	–	–
	12	HAQ_resp	–	–	1201	–	1201
	13	PASI_50	–	–	–	–	–
	14	PASI_75	–	–	–	–	–
	15	PASI_90	–	–	–	–	–
Cost	16	C_psoriasis	–	–	–	–	–
	17	C_HAQ	–	–	–	–	–

The potential benefits of resolving the uncertainty associated with groups of parameters are presented in *Table 135*. The most significant consequences of uncertainty associated with parameters relate to the natural history of PsA (8697 QALYs or £17.4M). Value is also associated with uncertainty regarding the treatment effect (1201 QALYs or £2.4M).

The results in *Table 135* suggest that a further observational (registry) study may well be worthwhile. In addition, there may be reason to believe that a study looking at the effectiveness of adalimumab, infliximab and etanercept in terms of the HAQ gain for responders might be worthwhile. It is likely that this research would have to be undertaken as part of a randomised trial design (RCT). Non-randomised designs, although still imposing significant costs, are likely to be cheaper than a RCT. It is therefore reasonable to assume that the research associated with the greatest value (natural history of progression) could be undertaken if infliximab, palliative care or adalimumab were approved for use in the NHS.

Types and categories of guidance resulting from points 3 and 4

Points 3 and 4 of the checklist are critical because if research is not judged to be worthwhile then no further assessments are required. In the case of PsA, more research appears to be worthwhile; the expected consequences of uncertainty are high. Whether the research required to generate the evidence needed can be conducted while the technology is approved for widespread use will determine whether AWR or OIR is a possibility. *Table 136* summarises the categories and types of guidance that could ultimately result from the sequence of assessments up to point 4.

Do the benefits of research exceed the costs?

The judgements made at points 5 and 6 of the checklist are critical because if the benefits of research are not judged to exceed the costs then no further assessments are required (unless there are significant irrecoverable costs; see *Appendix 4*). If they are and research can be conducted with approval, then AWR would be appropriate; however, other sources of uncertainty need to be assessed first as they will influence the potential benefits of research and, even when research is not conducted, they will also influence the appropriate category of guidance when there are significant irrecoverable costs.

TABLE 135 Overall expected consequences for the population as a result of uncertainty over groups of parameters at a threshold of £20,000 per QALY
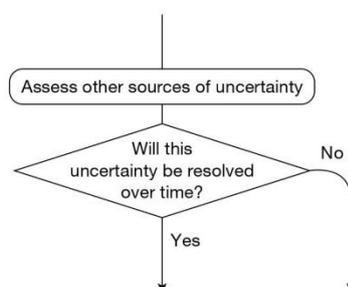
Group of parameters	Overall expected consequences, QALYs	Overall expected consequences, costs (£M)
Natural history	8694	17.4
Utility	0	0
Treatment effect	1201	2.4
Costs	0	0

TABLE 136 Types and categories of guidance that could ultimately result from assessments 3 and 4

Assessment	1	2	3	4	5	6	7	Guidance
7	No	No	Yes	Yes	Yes/no	Yes	–	OIR[2]
8	No	No	Yes	Yes	Yes/no	No	–	Reject[1]

Point 5: Will other sources of uncertainty resolve over time?

The fifth point on the checklist requires an assessment of whether or not changes are likely to occur in the future that will influence the cost-effectiveness of the alternative technologies and the potential benefits of research, that is, at the following point in the algorithm:

This assessment requires information about (1) changes in the prices of the technology and its comparators, (2) the emergence of new technologies that might make existing ones obsolete or change their cost-effectiveness and (3) other relevant research reporting. A number of potential sources of information and evidence were examined to inform this assessment for each case study (see *Appendix 4*). However, many potentially useful sources were either proprietary or public access was restricted, making it surprisingly difficult to inform these assessments with publically available information. When information and estimates were available they were often not directly relevant to a UK context.

Changes in the prices of the technology and its comparators

Changes in prices influence not only expected cost-effectiveness but also uncertainty and the potential benefits of research to future patients, for example if the price of a technology expected to be cost-effective is likely to fall significantly just before research reports the potential benefits will not be realised because approval of the technology will be less uncertain and there may be much less or little to gain from the results of the research. This assessment requires information about when major changes in prices are likely and some evidence about the likely extent of any changes. A major event in the life cycle of a pharmaceutical technology is the date at which the patent expires and cheaper generic versions of the brand become available. Although the date of patent expiry is, of course, known, it is surprisingly difficult to obtain the relevant date for particular products in the UK from publically available sources. Evidence of the extent to which the prices of generic versions are below the original brand price is also difficult to obtain and likely to differ by health-care system, type of technology, indication and time since patent expiry.

Although it was possible in the PsA case study to find patent expiry dates for etanercept (Enbrel), infliximab (Remicade) and adalimumab (Humira) in the USA (2012, 2014 and 2017, respectively) they were not available for the UK on the national patent database (Intellectual Property Office). There is a need to consider how access to the type of information required during NICE appraisal can be provided and how estimates of likely changes in prices relevant to the UK can be made readily available, if these assessments are to be routinely made.

Entry of new technologies

The entry of a new technology may make the existing technology that is expected to be cost-effective obsolete (no longer the most cost-effective alternative). Even when it does not, it will tend to change the relative cost-effectiveness of the alternatives, influencing how uncertain a decision to approve the original technology will be for future patients and the potential gains from research. A number of potential sources of information were examined to identify new

technologies relevant to the indications that were likely to become available. These included a variety of sources related to NICE topic selection, information about licence applications and clinical research in Phases I, II and III as well as evidence of the probability that earlier phase research leads to entry (probability of successful licence) and the likely time of entry (time to launch from initiating Phase I, II and III research). Again, this information and evidence is fragmented and in some cases restricted, for example NHS Horizon Scanning Centre[9].

It is known from the STA update (golimumab)[237] that a new comparator was available within 1 year of TA199.[125] There is a single trial for this comparing golimumab with placebo. Results showed that golimumab was not as effective as other strategies. To update the searches from this period onwards, new licence applications since the golimumab STA were searched using www.ukmi.nhs.uk/applications/NDO/dbSearch.asp. The terms 'etanercept', 'infliximab' and 'adalimumab' were used. A total of 742 records were found of which two were relevant (apremilast and certolizumab); however, these were not ongoing at the time of the NICE appraisal.

As no new evidence would have been available at the time of TA199[125] (golimumab trial anticipated but results not known), information about the technology is assumed to be limited. As such, scenarios are used in *Point 6: Are the benefits of research greater then the costs?* to explore the implications of this.

Other research reporting of the technology and its comparators

Research that is already under way, commissioned or likely to be undertaken, whether in the UK or elsewhere, is relevant for two reasons. First, if it is research based in the UK then guidance might impact on recruitment and the successful completion of this research. Second, when this research reports there is a chance that it will change the estimates of cost-effectiveness and resolve some of the current uncertainties. In other words, there is little to be gained by recommending OIR or AWR if the uncertainty is likely to be resolved in the near future when other research reports. A number of potential sources of information were examined to identify clinical research under way at the time of the relevant appraisal, including national and international trial registries (e.g. ClinicalTrials.gov, WHO ICTRP, Current Controlled Trials) and other databases that report NHS-funded research and not just clinical trials (e.g. NRR and UKCRN). Searches were also conducted for the recent STA.[237]

Despite an assiduous search no relevant records were identified. This may suggest that no other research was ongoing or expected for these comparators in these indications or it may indicate that currently available sources are fragmented, incomplete and/or difficult to access.

Point 6: Are the benefits of research greater than the costs?

The sixth point on the checklist requires a judgement on whether or not the potential benefits of conducting further research (initially considered at point 3) are likely to exceed the costs, that is, at the following point in the algorithm:
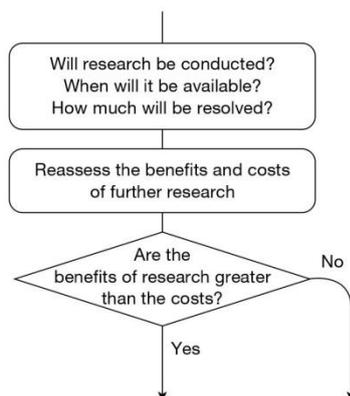

This requires an assessment of (1) whether or not the type of research that is required is likely to be conducted, (2) if conducted, when the results are likely to be available, (3) how much uncertainty is likely to be resolved and (4) the likely impact of any other sources of uncertainty on the longer-term benefits of research.

Will the research be conducted?

Even if research is recommended, it might not be undertaken by manufacturers or commissioned by research funders, and there is no guarantee that research will be able to recruit or complete. The expected consequences of uncertainty for PsA reported in *Point 3: Does more research seem worthwhile?* are illustrated in *Figure 81* for a range of probabilities that research will be successfully undertaken. The potential gains from research depend on a judgement of whether the research recommended as part of OIR or AWR will be successfully completed. The cost of research (in this case considered to be either £1.5M or £10M) can be compared directly with the potential benefits by either expressing the potential gains in population NHEs as the equivalent NHS resources (i.e. the resources that would be required to generate the same NHEs) or expressing the cost of research in terms of the QALYs that could be gained elsewhere in the NHS by using the same resources to provide access to health care. *Figure 81* shows the expected benefits of research minus its costs. Even assuming a cost of £10M for research the benefits of research outweigh this at all probabilities > 0 of research being conducted.

When will it be available?

Research, even if commissioned and successfully completed, will take time to complete and report; therefore, any assessment of the potential benefits should account for the fact that patient populations will not benefit from the results of research until they are available. If treatment decisions are irreversible (e.g. an acute indication or a irreversible chronic condition such as PsA) then it is only those patients who are incident after the research reports who will realise any of the potential benefits. For treatment decisions that are not irreversible, prevalent and incident patients can benefit from the results of the research. Patients prevalent while the research is undertaken will not benefit immediately but those who survive can benefit from the results once the research is completed.
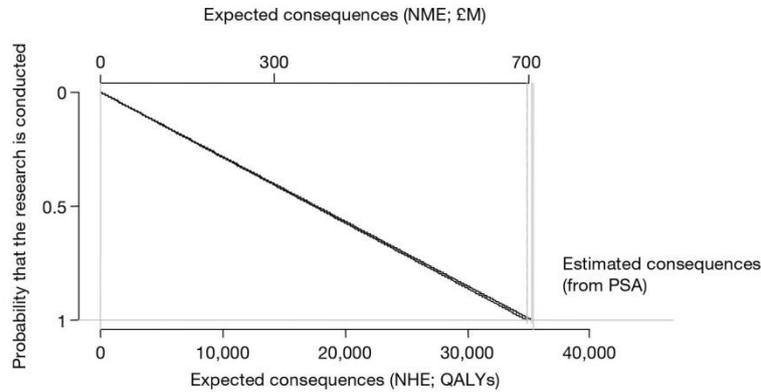
FIGURE 81 Expected potential benefits of research.

How long research might take to report will depend in part on the design (follow-up, sample size and end points), recruitment rates and size of the eligible patient population, as well as on how efficient the organisation and data collection might be.

The time needed for research to report may be unknown. *Table 137* shows that the time to report affects the maximum that one should be willing to invest in research. If research will take 5 years to report the maximum one should be willing to invest decreases by around £323M. *Figure 82* shows the boundary for the maximum that one should be willing to invest in research.

How much will be resolved?

How much of the uncertainty will be resolved depends on the type of research likely to be undertaken. In *Point 4: Is research possible with approval?* the potential benefits of different types of evidence were assessed. In the case of PsA, the most benefit can be gained by resolving the uncertainty regarding the natural history of PsA progression (measured using the HAQ). *Table 138* reports the potential benefits of research by delay in research reporting. As long as research reports before 4 years, the potential benefits are likely to exceed the costs. This table shows that one should be willing to invest the most in research relating to the natural history of disease progression (measured using the HAQ). This is associated with the highest value across all potential delays to research reporting.

What is the impact of other sources of uncertainty?

In *Point 5: Will other sources of uncertainty resolve over time?*, no information was identified relating to a change in price or other research reporting. There is evidence of new technologies emerging for PsA although it was unclear at the time of TA199[125] what the impact of this would be. Different scenarios are therefore explored below. In scenario A the new technology enters at year 5 and makes etanercept entirely obsolete, that is, not cost-effective and not uncertain. At this point there is no value in the evidence generated by research about PsA; therefore, in these circumstances the potential value of research is likely to exceed its costs only if it reports quickly. In scenario B the new technology enters at year 5 and has similar NHEs to etanercept and the uncertainty surrounding its expected cost-effectiveness is also similar. Now research about PsA has more potential value in the future as it will also help resolve some of the uncertainty in the choice between etanercept and the new technology for patients who become incident after that time.

TABLE 137 Maximum amount that one should be willing to invest in research by time to research reporting

Time from adoption to research reporting, years	Maximum cost of research, £M
0 (instant reporting)	707
1	647
2	584
3	520
4	453
5	384
6	312
7	238
8	161
9	82
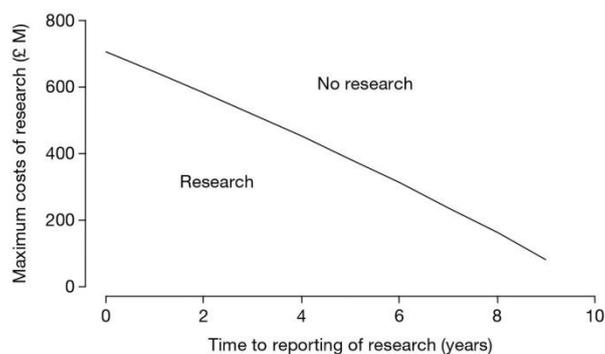
FIGURE 82 Boundary for the maximum costs that one should be willing to invest in research by time to research reporting.

TABLE 138 Potential benefits of research by time to research reporting

Time from adoption to research reporting	Maximum cost of research informing specific parameter sets, QALYs (£M)			
	Natural history	Treatment effect	Utilities	Costs
Immediate	8694 (17.40)	1201 (2.40)	0	0
1	7683 (15.38)	1061 (2.12)	0	0
2	6708 (13.42)	926 (1.85)	0	0
3	5765 (11.53)	796 (1.59)	0	0
4	4854 (9.71)	670 (1.34)	0	0
5	3974 (7.95)	549 (1.09)	0	0
6	3123 (6.25)	431 (0.86)	0	0
7	2302 (4.61)	372 (0.63)	0	0
8	1508 (3.02)	208 (0.42)	0	0
9	741 (1.48)	102 (0.20)	0	0

Tables 139 and *140* evaluate alternative times of introduction of the new technology in scenarios A and B. Under scenario B there is still some value in conducting further research even when research will not report until some point in the future. This is greater if the new technology is available sooner rather than later. It must be noted that the potential values of research presented in these tables, even after accounting for the type of evidence, follow-up and time until research reports, should still be regarded as upper bounds to the values that are likely to be realised by actual research for two reasons: (1) even well-designed research with large sample sizes will not fully resolve the uncertainty in the value that a parameter might take, especially in specific target populations and in a particular (future) context, and (2) insofar as implementation of NICE guidance is not 'perfect' and all clinical practice might not immediately respond to the results of research, the full benefits will be realised only over time or with additional implementation efforts. For these reasons a judgement of whether or not the benefits of research are likely to exceed the costs might be made conservatively, requiring evidence that, even in pessimistic scenarios, the research would still be worthwhile.

Types and categories of guidance resulting from points 5 and 6

The assessment that the benefits of research are likely to exceed the costs leads to a final decision at point 6 in the checklist, therefore negating the need to make the assessments at point 7 (*Table 141*). The decision is to recommend OIR. The model parameter associated with the greatest value is the progression of disease, as measured using the HAQ. There is also some value associated with establishing the effectiveness of biologics, measured using HAQ response. The first of these uncertainties could be addressed by commissioning an observational study, perhaps a registry. The second issue is likely to require a RCT.

TABLE 139 Potential value of research by time to report for a range of times of entry of the new technology under scenario A

Time from adoption to research reporting, years	Maximum cost of research for possible times of introduction of a new technology, £M										
	Not introduced	9 years	8 years	7 years	6 years	5 years	4 years	3 years	2 years	1 year	Immediate introduction
0 (instant reporting)	706.83	646.58	584.22	519.68	452.87	383.73	312.17	238.11	161.50	82.10	0
1	624.71	564.46	502.10	437.56	370.76	301.62	230.06	155.99	79.33	0	0
2	545.37	485.12	422.76	358.22	291.42	222.28	150.72	76.65	0	0	0
3	468.72	408.47	346.11	281.56	214.76	145.62	74.06	0	0	0	0
4	394.65	334.40	272.04	207.50	140.70	71.56	0	0	0	0	0
5	323.09	262.84	200.48	135.94	69.14	0	0	0	0	0	0
6	253.95	193.70	131.94	66.80	0	0	0	0	0	0	0
7	187.15	126.90	64.54	0	0	0	0	0	0	0	0
8	122.61	62.36	0	0	0	0	0	0	0	0	0
9	60.25	0	0	0	0	0	0	0	0	0	0

TABLE 140 Potential value of research by time to report for a range of times of entry of the new technology under scenario B

Time from adoption to research reporting, years	Maximum cost of research for possible times of introduction of a new technology, £M										
	Not introduced	9 years	8 years	7 years	6 years	5 years	4 years	3 years	2 years	1 year	Immediate introduction
0 (instant reporting)	706.8	1140.0	1588.7	2052.9	2533.4	3030.7	3545.4	4078.1	4629.4	5200.1	5790.7
1	624.7	1058.0	1506.6	1970.8	2451.3	2948.5	3463.0	3996.0	4547.3	5118.0	5118.0
2	545.3	978.7	1427.2	1891.4	2371.9	2869.2	3383.9	3916.0	4468.0	4468.0	4468.0
3	468.7	902.0	1350.6	1814.8	2295.3	2792.5	3307.2	3840.0	3840.0	3840.0	3840.0
4	394.6	828.0	1276.5	1740.7	2221.2	2718.5	3233.2	3233.2	3233.2	3233.2	3233.2
5	323	756.4	1204.9	1669.2	2149.6	2646.9	2646.9	2646.9	2646.9	2646.9	2646.9
6	253.9	687.3	1135.8	1600.0	2080.5	2080.5	2080.5	2080.5	2080.5	2080.5	2080.5
7	187.1	620.5	1069.0	1533.2	1533.2	1533.2	1533.2	1533.2	1533.2	1533.2	1533.2
8	122.6	555.9	1004.4	1004.4	1004.4	1004.4	1004.4	1004.4	1004.4	1004.4	1004.4
9	60.2	493.6	493.6	493.6	493.6	493.6	493.6	493.6	493.6	493.6	493.6

TABLE 141 Type and category of guidance that could ultimately result from assessments 5 and 6

Assessment	1	2	3	4	5	6	7	Guidance
7	No	No	Yes	Yes	Yes	Yes	–	OIR[2]

Appendix 11

Technical notes to the case studies

Model linearity and correlation between parameters

Evaluating cost-effectiveness and the need for further research (and its value) using a decision model may pose analytical challenges if, for example, the model is non-linear or there is correlation between input parameters. In a linear model that does not feature correlation between parameters it is possible to calculate cost-effectiveness using a deterministic analysis, and the value of further research using one loop of simulation, otherwise more burdensome simulation procedures may need to be implemented. Independence between parameters (or the absence of correlation) is important if parameters need to be evaluated separately for their impact on decision uncertainty and the value of further research.[a] The existence of correlation depends greatly on how the input parameters were estimated; however, correlation may not have a significant impact on the results (i.e. in the assessment of linearity and in estimating the value of further research for specific parameters). An investigation of model linearity is required if it is proposed to use a deterministic analysis to inform the assessments described in *Chapter 5*. An investigation of correlation is required to know whether or not a deterministic analysis may be appropriate in some circumstances, and also to know what information can be presented to inform the assessments at points 3, 4 and 6 of the checklist described in *Chapter 5*.

Exploring model linearity
Why is model linearity important?

A decision model is a function used to calculate NHEs based on a set of *P* input parameters $\theta = \{\theta_1, \theta_2, \ldots, \theta_p\}$. The model can be represented as a function $g(.)$ such that $\text{NHE} = g(\theta)$. The input parameters are uncertain and thus in estimating the expected NHEs the following expression needs to be evaluated: $E[\text{NHE}] = E[g(\theta)]$. It is current practice that this expectation is evaluated using a Monte Carlo simulation procedure (PSA), which entails (1) simulating from the distribution of the parameters, (2) evaluating the NHEs for each set of simulated parameters and (3) calculating the expectation by averaging the NHEs obtained from the simulations. Although the single Monte Carlo simulation procedure required to estimate cost-effectiveness is usually not too burdensome, the need to evaluate nested expectations in estimating the value of further research (for specific parameters) implies using nested simulations and therefore a much higher computation time.[b] If the analysis becomes unfeasible, analysts may either choose not to proceed or choose to apply simplifying assumptions (e.g. reduce the number of Monte Carlo simulations used).

The calculation of expected NHEs can however be simplified if $g(.)$ is linear or multilinear in its parameters. In this case, $E[g(\theta)]$ is equal to $g(E[\theta])$ and hence expected NHEs can be computed in a single calculation with the parameters set to their expected values (deterministic analysis), in this case eliminating the need for simulation procedures. However, the presence of non-linearity implies that the desired model output cannot be expressed as a function of the parameters set to their expected values, and proceeding with such a deterministic analysis will return biased results.

When is a model linear or multilinear?

Linearity happens when $g(.)$ represents a sum (or subtraction) of parameters (e.g. $g(\theta_1,\theta_2) = \theta_1 + \theta_2$, in which case $E[\theta_1 + \theta_2] = E[\theta_1] + E[\theta_2]$). For this relation to sustain, it is not necessary that the summands are mutually independent. When the NHE function contains products of (independent) parameters, $g(.)$ is said to be multilinear. If, for example, the NHE can be calculated using the product of θ_1 and θ_2, that is, $g(\theta_1,\theta_2) = \theta_1 \times \theta_2$, then $E[\theta_1 \times \theta_2]$ is equal to $E[\theta_1] \times E[\theta_2]$. This is valid only if θ_1 and θ_2 are independent. However, if the covariance (correlation) between these parameters is close to zero, this relation might be assumed to sustain.

How can model linearity be evaluated?

Evaluating model linearity is fairly straightforward if $g(.)$, the decision model, can be defined using algebraic descriptions. In this case, even if the model is apparently non-linear there may be a way to express the conditional NHEs as a linear (or multilinear) function of transformations of parameters by rearranging the model equations. When the model function is complex it may be more difficult to express it algebraically, which impedes a direct evaluation of linearity.

In complex models, evaluating non-linearity can be carried out by comparing deterministic and probabilistic results. A first step may be to evaluate its impact over the overall cost-effectiveness results (NHEs). For this, the expected NHEs using probabilistic, that is, $E[f(\theta)]$, and deterministic, that is, $f(E[\theta])$, estimates can be obtained and compared. Sufficient Monte Carlo simulations must be run to obtain a stable probabilistic estimate of $E[f(\theta)]$. If this differs significantly from the deterministic estimate, it can be said that the model is non-linear. It may happen that the bias differs but is the same for all strategies. In this case, the existence of non-linearity may not be significant in the sense that the conclusions drawn from the cost-effectiveness analysis will not change. Observing that the deterministic and probabilistic estimates do not differ in term of expected NHEs does not ensure that the model is linear, as the estimates of the value of resolving all uncertainty in a specific parameter may still be affected. It is thus important to also evaluate linearity using the expected NHEs when conditioned on specific values of a parameter of interest (spanning through its plausible range).[c] Estimates of the consequences of uncertainty calculated using or not using the assumption of linearity can also be compared.

If the assumption of linearity is rejected then model outputs should be estimated on the basis of probabilistic analysis. The evaluations of linearity proposed here require that some probabilistic analysis be conducted; however, it could potentially reduce computation time if, conditional on the model appearing to be linear, additional analysis could be based on deterministic results. Note again that observing a difference between deterministic and probabilistic results indicates non-linearity but not observing a difference does not guarantee linearity.

Exploring correlation

The existence of correlation between parameters depends on how these input parameters were estimated. Examples of when correlated quantities are used to inform a model may be when more than one coefficient derived using a regression analysis is used to predict prognosis, or potentially when multiple results from an evidence synthesis procedure are used.

Why is correlation important?

Correlation is important because under certain conditions it may determine non-linearity (see *Exploring model linearity*). Also, it may determine how the value of resolving uncertainty on specific parameters is evaluated. In undertaking these calculations it is necessary to evaluate the expected NHEs when the parameters of interest take particular values while the remaining parameters are still uncertain (conditional NHE estimates, as described in *Exploring model*

linearity; see notes a and b). Thus, in the presence of correlation, the distribution of the remaining parameters depends on the value assumed for the parameter(s) of interest, and this needs to be accounted for. On some occasions we may know these conditional distributions, in which case we may be able to explicitly use these in the estimation.[d] For example, coefficients from a linear regression analysis are commonly assumed multivariate normally distributed, where regression estimates inform both the means and the variance/covariance matrix. However, in other cases, information on the conditional distribution of sets of parameters may not be easy to obtain, for example when we have no access to more detailed results of analysis, or when correlation is implicitly introduced in the analyses, for example through the use of a random effect across parameters when relative treatment effects for multiple treatments were obtained jointly in a mixed-treatment comparison.

Note that if the parameters θ^C are probabilistically independent of the parameters θ^I, correlation does not play a part in computations, even if the subset of parameters in θ^C are not independent between themselves or the subset of parameters in θ^I are not independent between themselves. Parameters that are possibly correlated can thus be grouped (e.g. natural history parameters can be evaluated separately from treatment effect parameters) to avoid dealing explicitly with correlation.[e] In this case, however, conclusions over individual parameters cannot be drawn. In some situations it may be important to evaluate the parameters individually, especially if there is the need to further evaluate specific research designs (e.g. comparators for RCTs). In this context, correlation between parameters cannot be ignored as the results of investigations will be biased. If we suspect that there may be correlation between the parameter(s) of interest and its complementary set, and if this cannot be explicitly accounted for, it may be worth evaluating the presence of correlation and its impact on the results,[f] instead of proceeding to assume independence.

How can correlation and its impact be evaluated?
An evaluation of the existence and impact of correlation can be conducted by exploring:

1. *Between-parameter correlation*: The issue is that when non-monotonic transformations of parameters are used to define the model evaluating correlation between the untransformed parameters may not be informative.
2. *Impact of correlation in cost-effectiveness*: Consider two possibly correlated parameters, θ_1 and θ_2. We will examine the case in which the NHE of treatment A is a function of θ_1 and the NHE of treatment B is a function of θ_2. In the absence of correlation, alternative values of θ_1 should not impact on cost-effectiveness estimates of treatment B – only through correlation the latter may change.[g]
3. *Impact of correlation in decision uncertainty*: The previous assessments aim to evaluate the impact of correlation on the values assumed by the expected NHEs. However, even small changes in the expected NHEs due to correlation can meaningfully impact on the adopt decision and/or on the maximum NHEs attained. It may be tempting to assess the impact of correlation by calculating the value of resolving uncertainty for potentially correlated parameters individually (i.e. assuming independence), summing these and comparing this sum with the estimate obtained by assessing the value of resolving uncertainty for the group of potentially correlated parameters; however, even in the presence of independence these values may not be equal.

As with the explorations proposed for linearity, the absence of evidence of correlation does not mean that there is no significant correlation; thus, care is needed in using the assumption of independence in further analysis.

Accounting for the investment profile of displaced interventions

When an intervention is approved that is more costly than its comparator the additional cost of the new intervention must be funded at the expense of other interventions in the NHS. Thus, the approval of a new intervention displaces the comparator, a less effective or more costly intervention used to treat the same disease, and other interventions in the NHS used to treat other diseases. The amount of this additional displacement depends on the difference in budget impact between the newly approved intervention and its comparator. The opportunity cost of new interventions has been well described in the literature and is the key economic argument for needing a threshold in a budget-constrained system. If the new intervention is more cost-effective than the intervention displaced then the health system efficiency has improved.

As funders make decisions about providing a new intervention they are interested not only in the total costs and health effects but also in how these costs and health effects are accrued over time. The accrual of costs and health effects can be shown using the investment profile previously described. Additional research or other sources of uncertainty resolving over time can cause the reimbursement decision to be reversed before the end of the lifetime of the intervention (the time horizon that was used in assessing cost-effectiveness). For some interventions a change of guidance would not prevent the investment profile from being completed; however, in those circumstances in which the accumulation of NHEs would cease at the point of a change in guidance, the investment profile can inform the decision-maker about the potential risk in terms of the NHEs accumulated to that point.

The incremental investment profiles used in this report take into account the timing of the costs and health effects of both the new intervention and its comparator(s); however, the incremental investment profile does not take into account the investment profiles of the additionally displaced non-comparator interventions because at the time of approval it is unknown which current interventions will be displaced. Because the investment profiles of the displaced interventions are not known it is important to assess the implicit assumptions being used and their implications.

In the investment profiles presented in the main report and the case studies (see *Appendices 7–10*) we consider the overall NHEs of the displaced interventions using the cost-effectiveness threshold (k). What we have not taken into account is how the overall NHEs of the displaced interventions are accrued over time.

For many interventions upfront costs are offset over time by future health benefits. *Table 142* shows the population costs (C), health effects (H) and NHEs for two mutually exclusive interventions. The lifetime of use for the interventions is assumed to be 10 years, after which they are assumed to be obsolete.

Figure 83 shows the cumulative population NHEs for the two interventions from *Table 142*. The cumulative NHE of a new intervention must be higher than that of its comparators for it to be considered cost-effective. This means that the incremental cumulative NHE of the new intervention relative to the comparator will be positive at the end of the time horizon of interest.

Figure 83 shows that the NHEs of the new intervention do not become positive until year 4. The cumulative NHEs from the new intervention do not exceed those offered by the comparator until year 6. As can be seen from *Table 142*, discontinuing the existing comparator would release £20,000 to fund the new intervention, which leaves a shortfall of £8000. Additional interventions must be displaced to fund this £8000.[h] *Table 143* shows three alternative scenarios for how the

TABLE 142 Is the new intervention cost-effective and what are the risks?

Year	New intervention				Comparator				Incremental NHE	
	C (£)	H	NHE	Cumulative NHE	C (£)	H	NHE	Cumulative NHE	Per period	Cumulative
1	10,000	0.30	−0.20	−0.20	6000	0.30	0.00	0.00	−0.20	−0.20
2	10,000	0.30	−0.20	−0.40	6000	0.30	0.00	0.00	−0.20	−0.40
3	1000	0.32	0.27	−0.13	1000	0.25	0.20	0.20	0.07	−0.33
4	1000	0.34	0.29	0.16	1000	0.25	0.20	0.40	0.09	−0.24
5	1000	0.36	0.31	0.47	1000	0.25	0.20	0.60	0.11	−0.13
6	1000	0.38	0.33	0.80	1000	0.25	0.20	0.80	0.13	0.00
7	1000	0.40	0.35	1.15	1000	0.25	0.20	1.00	0.15	0.15
8	1000	0.40	0.35	1.50	1000	0.25	0.20	1.20	0.15	0.30
9	1000	0.40	0.35	1.85	1000	0.25	0.20	1.40	0.15	0.45
10	1000	0.40	0.35	2.20	1000	0.25	0.20	1.60	0.15	0.60
Total	28,000	3.60	2.20		20,000	2.60	1.60		0.60	
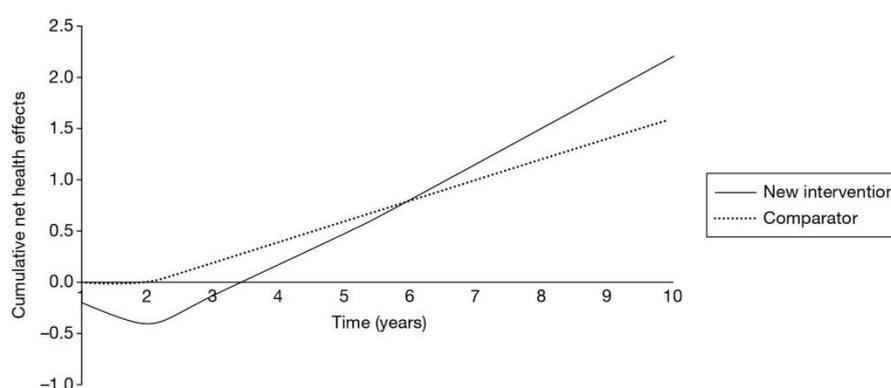
FIGURE 83 Cumulative NHEs of new and comparator interventions.

additional displaced interventions accrue incremental NHEs over time (relative to the relevant next best alternative comparator intervention).

In scenario 1 the additional displaced interventions could be described as 'more risky' as they do not break-even until year 10. In scenario 2 the accumulation of NHEs is constant over time and so the interventions could be described as 'no risk'. In scenario 3 the additional displaced interventions initially have positive incremental NHEs and so could be described as 'less risky'. *Figure 84* shows the cumulative incremental NHE curves of average displaced interventions for each of these scenarios. Note that the cumulative incremental NHEs of the additional displaced interventions are always equal to zero at year 10, which is consistent with the assumption underlying the use of a cost-effectiveness threshold.

Figure 85 shows the incremental NHE curve for the new intervention relative to its comparator (solid black line). This incremental curve is equivalent to those used in the main report and case studies (see *Appendices 7–10*), which take into account the timing of the incremental NHEs relative to the comparator and the opportunity costs of the additional displaced interventions, but not the timing of the NHEs of the displaced interventions. This would suggest that the

TABLE 143 Cumulative incremental effects of additional displaced interventions

Year	C (£)	H	NHE	Cumulative NHE
Scenario 1: more risky				
1	800	−0.06	−0.10	−0.10
2	800	−0.06	−0.10	−0.20
3	800	−0.06	−0.10	−0.30
4	800	−0.01	−0.05	−0.35
5	800	0.04	0.00	−0.35
6	800	0.04	0.00	−0.35
7	800	0.09	0.05	−0.30
8	800	0.14	0.10	−0.20
9	800	0.14	0.10	−0.10
10	800	0.14	0.10	0.00
Total	8000	0.40	0.00	
Scenario 2: no risk				
1	800	0.04	0.00	0.00
2	800	0.04	0.00	0.00
3	800	0.04	0.00	0.00
4	800	0.04	0.00	0.00
5	800	0.04	0.00	0.00
6	800	0.04	0.00	0.00
7	800	0.04	0.00	0.00
8	800	0.04	0.00	0.00
9	800	0.04	0.00	0.00
10	800	0.04	0.00	0.00
Total	8000	0.40	0.00	
Scenario 3: less risky				
1	800	0.14	0.10	0.10
2	800	0.14	0.10	0.20
3	800	0.14	0.10	0.30
4	800	0.09	0.05	0.35
5	800	0.04	0.00	0.35
6	800	0.04	0.00	0.35
7	800	−0.01	−0.05	0.30
8	800	−0.06	−0.10	0.20
9	800	−0.06	−0.10	0.10
10	800	−0.06	−0.10	0.00
Total	8000	0.40	0.00	

break-even point at which the cumulative NHEs became positive was at year 6. By considering different risk profiles of the displaced interventions we can see the difference in effect on the accumulation of NHEs, taking into account the total displaced interventions. If the investment profile of the displaced interventions is constant then the inclusion of the displaced interventions has no effect on the investment profile of the new intervention. If the investment profile of the displaced interventions is initially positive then the break-even point of the new intervention investment profile has been underestimated by not considering the displaced interventions. If the investment profile of the additional displaced interventions is initially negative ('more risky'), the break-even point will have been overestimated.
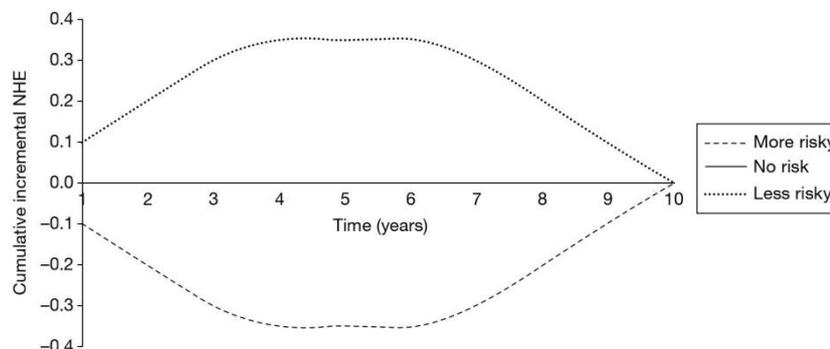
FIGURE 84 Cumulative incremental NHEs of additional displaced interventions.
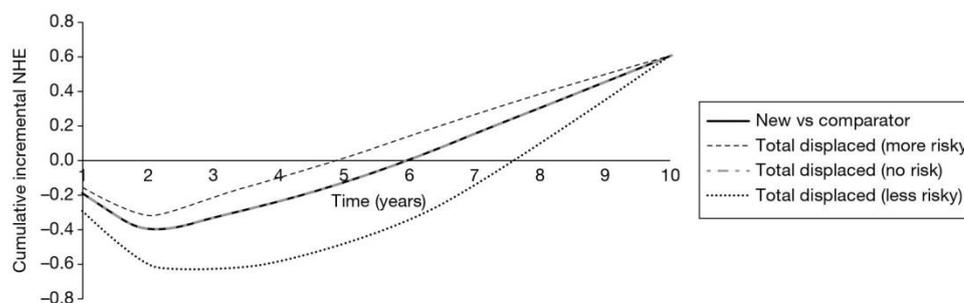
FIGURE 85 Cumulative incremental NHEs of the new intervention.

From the examples demonstrated the investment profile and break-even point of the new intervention is affected by its budget impact compared with the cost of the displaced interventions. The larger this difference the more interventions will be displaced and the more influence displaced interventions will have on the new intervention's risk profile. The investment profile of the new intervention is also affected by the shape of the investment profile of the displaced intervention. If the displaced intervention has a non-constant investment profile then the current evaluation will change by whether it is more or less risky than the displaced intervention. If the displaced intervention has a riskier profile than the new intervention then the investment profile of the new intervention will be underestimated by not taking into account the displaced intervention. If the displaced intervention has a less risky profile then the investment profile of the new intervention will be overestimated by not taking into account the displaced intervention. And if there is no risk to the displaced intervention then the investment profile of the new intervention is not affected.

It is tempting to think that displaced interventions are likely to have reached the positive NHEs of their investment profile and that the risk of new interventions is being underestimated; however, all previous NHEs of displaced interventions must be considered sunk and the investment profile should be considered starting at the point of the decision moving forward, thus for the incident population of those who would receive the displaced intervention.

Why the consequences of uncertainty differ from mean incremental effects

A decision based on expected NHEs may be incorrect if the true value of the underlying evidence suggests that an alternative intervention to the one reimbursed would have offered greater health gains. The probability that a decision to recommend a particular intervention is incorrect is equal to the probability of observing a state of the world in which a different intervention would maximise NHEs. The cost of making such an error, measured as health forgone, is determined by the difference in NHEs between the intervention that was selected and the best that could have been achieved. Thus, the expected cost of uncertainty, regarding the choice of the optimum treatment, is a function of the probability of making such an error and the amount of health forgone if the wrong treatment were selected. By avoiding reimbursement of an intervention that turns out to be cost-ineffective, the decision-maker would expect better health gains overall. As such, the expected cost of uncertainty describes the magnitude of the health gains that could be achieved by eliminating the decision uncertainty. An estimate of the cost of uncertainty is required to inform the assessments made from point 3 of the checklist described in *Chapter 5*.

Estimating the cost of uncertainty

Assessing the cost of uncertainty requires estimates of the probability that the decision will be incorrect and the size of the associated health losses. Consider the simple example in *Table 144* comparing two alternative interventions, $j = 1,2$. The available evidence (parameters) used to calculate the NHEs for a typical patient who receives either intervention is described by θ. This parameter can take one of four possible values.

Intervention 2 has the highest expected NHEs, $\max E_\theta NHE(j,\theta) = 0.6$, and would therefore be regarded as cost-effective based on current evidence. The incremental net health benefit from reimbursing intervention 2 instead of intervention 1 would be $E_\theta NHE(2,\theta) - E_\theta NHE(1,\theta) = 0.6 - 0.5 = 0.1$.

The probability of error and the consequences of that error in terms of health forgone can be estimated directly from the results of PSA. If, subsequent to intervention 2 being reimbursed, it was learned that the true value of θ was actually 3 or 4 (not 1 or 2), the decision based on current evidence would have been incorrect as intervention 1 would have offered greater NHEs and should have been reimbursed instead of intervention 2. The probability of making an error is therefore 0.5 (50%), assuming that all values of θ are equally likely. The consequences of this error in terms of health forgone are described by $E_\theta NHE(1,3) - E_\theta NHE(2,3) = 0.6 - 0.4 = 0.2$ for $\theta = 3$ and $E_\theta NHE(1,4) - E_\theta NHE(2,4) = 1.0 - 0.8 = 0.2$ for $\theta = 4$. This gives an expected health loss of 0.2. This expected health loss from making the wrong decision must be multiplied by the probability of error (50%) to express it in terms of the expected cost of uncertainty: $0.5 \times 0.2 = 0.1$.

TABLE 144 Expected and maximum NHEs per patient

θ	$NHE(1,\theta)$	$NHE(2,\theta)$	Max. NHE
1	0.4	0.5	0.5
2	0.1	0.7	0.7
3	0.6	0.4	0.6
4	1.0	0.8	1.0
E_θ	0.5	0.6	0.7

Max., maximum.

In the absence of a probabilistic analysis, an assessment of expected NHEs can be made based on a deterministic analysis if the underlying model is linear or is multilinear with independent parameters. However, a deterministic analysis provides no information on the probability and consequences of error, and so an assessment of the cost of uncertainty would have to be made on the basis of additional, perhaps informal, considerations. Note that in this example the expected health loss if intervention 2 was selected in error is 0.2, which is greater than the incremental NHE of intervention 2 compared with intervention 1 across all values of θ (0.1). In general, the difference between the expected NHEs of competing interventions cannot provide a good approximation of the consequences of uncertainty.

Why averaging scenarios may be misleading

One source of uncertainty in estimating the true value of the expected costs and health outcomes associated with alternative interventions has been described variously as modelling, structural or scenario uncertainty. Scenario uncertainty arises when alternative plausible modelling assumptions can be made. Potential sources of scenario uncertainty include alternative assumptions about which sources of evidence are relevant to the decision problem, the choice of mathematical model to estimate model parameters and the type and structure of the decision model that is used to generate estimates of costs and health outcomes. Unless the scenario uncertainty can be characterised by a parameter within the decision model, its impact on decision uncertainty will not be reflected in routine PSA based on Monte Carlo simulation and further analytical processes are required. This section describes in more detail how to estimate appropriately the value of further research in the presence of both parameter and scenario uncertainty. This is relevant when providing information for the assessments required at points 3 and 4 of the checklist described in *Chapter 5*.

How to estimate cost-effectiveness in the presence of scenario uncertainty

The estimation of expected costs, *c*, and health outcomes, *h*, will be informed by the available evidence, θ, and the set of assumptions, *s*, that characterise a particular scenario. The results of a cost-effectiveness analysis for each alternative intervention, *j*, can be expressed in terms of NHEs,

$$NHE(j, \theta, s) = h(j, \theta, s) - \frac{c(j, \theta, s)}{k},$$

where *k* is the cost-effectiveness threshold.

The assessment of cost-effectiveness is made by taking the expectation of NHEs across the parameter uncertainty and the range and likelihood of the alternative scenarios, $E_\theta E_s NHE(j, \theta, s)$.

Parameter uncertainty (uncertainty as to the true value of θ) can be characterised by assigning probability distributions to the model parameters. This allows the results of the model to be evaluated by PSA; a Monte Carlo simulation procedure can be used to repeatedly sample from those distributions a set of model inputs, and for each set calculate the corresponding model outputs. The expectation of the model outputs is found by averaging across the results of the Monte Carlo simulation. If scenario uncertainty has not been characterised by a probability distribution and sampled simultaneously alongside the parameter values, that Monte Carlo simulation procedure will estimate the expectation of NHEs across parameter uncertainty within a single scenario, for example $E_\theta NHE(j, \theta, s = 1)$. The Monte Carlo simulation procedure must then be repeated for each possible scenario. The expected NHEs across parameter uncertainty from each scenario can then be combined utilising the likelihood of each possible scenario to describe the expected NHEs across both parameter and scenario uncertainty.

How to estimate the value of research in the presence of scenario uncertainty

The best that a decision-maker could do would be to select the intervention that maximised health gains for a particular realisation of θ and s. The expected NHE associated with this error-free choice is $E_\theta E_s \max_j NHE(j,\theta,s)$. The maximum value of further research that would eliminate all uncertainty, including scenario uncertainty, is

$$EVPI_{\theta,s} = E_\theta E_s \max_j NHE(j,\theta,s) - \max_j E_\theta E_s NHE(j,\theta,s).$$

Given that we believe a certain scenario to be true, for example scenario 1, the maximum value of further research to eliminate parameter (θ) uncertainty is

$$EVPI_{\theta|s=1} = E_\theta \max_j NHE(j,\theta,s=1) - \max_j E_\theta NHE(j,\theta,s=1).$$

Combining the expected value of further research within each possible scenario using the likelihood of each scenario would not describe the value of research that would eliminate both scenario and parameter uncertainty. When alternative scenarios would suggest that different interventions would be expected to be cost-effective the NHEs of choosing the best intervention when integrating both parameter and modelling uncertainty cannot be found by averaging the NHEs of choosing the best intervention within each scenario. This is because

$$\max_j E_s E_\theta NHE(j,\theta,s) \neq E_s \left[\max_j E_\theta NHE(j,\theta,s) \right].$$

Averaging across the scenarios in this way ignores scenario uncertainty as it assumes that the decision-maker can select alternative treatments based on knowing how the scenario uncertainty is resolved. It is only when scenario uncertainty is not associated with decision uncertainty, that is, when the same intervention would be identified as cost-effective in all scenarios, that this would produce an unbiased estimate of the value of further research.

An appropriate method by which to evaluate the value of further research that would eliminate both parameter and scenario uncertainty would be to stack the Monte Carlo simulations, ensuring that the proportion of Monte Carlo simulations selected from the PSA within each scenario corresponds to the likelihood of that scenario being correct. To produce representative results this may require a larger number of simulations than would be selected based on consideration of parameter uncertainty alone, particularly if the likelihood of any particular scenario is low.

Measures quantifying the maximum value of eliminating uncertainty in either parameters or scenarios can also be of importance. The maximum value of further research to eliminate parameter (θ) uncertainty when scenario uncertainty is present is $EVPI_\theta = E_\theta \max_j E_s NHE(j,\theta,s) - \max_j E_\theta E_s NHE(j,\theta,s)$. The maximum value of further research to eliminate scenario uncertainty in the presence of parameter uncertainty is $EVPI_s = E_s \max_j E_\theta NHE(j,\theta,s) - \max_j E_\theta E_s NHE(j,\theta,s)$.

Example

Table 145 presents the NHEs for each intervention based on competing plausible forms of the function $NHE(j,\theta,s)$; $s=1,2$. It is assumed that each value of θ is equally likely and each scenario is equally plausible. Therefore, the expectation across parameter or scenario uncertainty can be found by averaging across the relevant set of results.

Based on scenario 1 it would appear that intervention 1 is cost-effective as it offers the greatest expected NHEs: $\max_j E_\theta NHE(j,\theta,s=1) = 5281$; however, in the alternative scenario 2 the

NHEs of intervention 2 are estimated to be larger than those associated with intervention 1 and so intervention 1 no longer appears cost-effective: $\max E_\theta NHE(j,\theta,s=2) = 5688$. In the presence of both scenario and parameter uncertainty it would appear that intervention 1 is cost-effective: $\max E_\theta E_s NHE(j,\theta,s) = 5281$. Averaging across the expected NHEs from within each scenario result would be equivalent to assuming that the decision-maker could select a different intervention based on how the scenario uncertainty resolved, giving expected NHEs of $E_s \max_j E_\theta NHE(j,\theta,s) = 5485$.

Table 146 presents the value of further research for a range of possible research questions.

TABLE 145 Population NHEs for alternative scenarios

θ	Overall population		
	NHE(1,θ)	*NHE*(2,θ)	Max. *NHE*
Scenario 1 (base case)			
1	4125	3250	4125
2	750	5250	5250
3	6000	2250	6000
4	10,250	6000	10,250
E_θ	5281	4188	6406
Scenario 2			
1	4125	4750	4750
2	750	6750	6750
3	6000	3750	6000
4	10,250	7500	10,250
E_θ	5281	5688	6938
Average across scenarios			
1	4125	4000	4125
2	750	6000	6000
3	6000	3000	6000
4	10,250	6750	10,250
E_θ	5281	4938	6594

Max., maximum.

TABLE 146 Population NHEs with equally likely alternative modelling assumptions

NHE	Value of further research	Calculation
$EVPI_{\theta\|s=1}$	1125	6406 − max.(5281, 4188)
$EVPI_{\theta\|s=2}$	1250	6938 − max.(5281, 5688)
$EVPI_\theta$	1313	6594 − max.(5281, 4938)
$EVPI_s$	203	Mean[max.(5281, 4188), max.(5281, 5688)] − max.(5281, 4938)
$EVPI_{\theta,s}$	1391	Mean(6406, 6938) − max.(5281, 4938)

Max., maximum.

The value of resolving parameter uncertainty in scenario 1 is 1125. In scenario 2 this value is 1250. Note that in this example the value of research to resolve parameter uncertainty is estimated to be greater when it takes account of scenario uncertainty (1313), even though the research would not resolve the scenario uncertainty. The value of resolving only scenario uncertainty is 203. The value of resolving both parameter and scenario uncertainty in this example is 1391.

Health Technology Assessment programme

HTA Commissioning Board *(continued)*

Professor Caroline Watkins,
Professor of Stroke and Older
People's Care, Chair of UK
Forum for Stroke Training, Stroke
Practice Research Unit, University
of Central Lancashire

Dr Duncan Young,
Senior Clinical Lecturer and
Consultant, Nuffield Department
of Anaesthetics, University of
Oxford

Observers

Dr Tom Foulks,
Medical Research Council

Dr Kay Pattison,
Senior NIHR Programme
Manager, Department of Health

HTA Clinical Evaluation and Trials Board

Chair,
Professor Sallie Lamb,
Director,
Warwick Clinical Trials Unit,
Warwick Medical School,
University of Warwick and Professor of
Rehabilitation,
Nuffield Department of Orthopaedic,
Rheumatology and Musculoskeletal Sciences,
University of Oxford

Deputy Chair,
Professor Jenny Hewison,
Professor of the Psychology of Health Care,
Leeds Institute of Health Sciences,
University of Leeds

Programme Director,
Professor Tom Walley, CBE,
Director, NIHR HTA programme,
Professor of Clinical Pharmacology,
University of Liverpool

Members

Professor Keith Abrams,
Professor of Medical Statistics,
Department of Health Sciences,
University of Leicester

Professor Martin Bland,
Professor of Health Statistics,
Department of Health Sciences,
University of York

Professor Jane Blazeby,
Professor of Surgery and
Consultant Upper GI Surgeon,
Department of Social Medicine,
University of Bristol

Professor Julia M Brown,
Director, Clinical Trials Research
Unit, University of Leeds

Professor Alistair Burns,
Professor of Old Age Psychiatry,
Psychiatry Research Group, School
of Community-Based Medicine,
The University of Manchester &
National Clinical Director for
Dementia, Department of Health

Dr Jennifer Burr,
Director, Centre for Healthcare
Randomised trials (CHART),
University of Aberdeen

Professor Linda Davies,
Professor of Health Economics,
Health Sciences Research Group,
University of Manchester

Professor Simon Gilbody,
Prof of Psych Medicine and Health
Services Research, Department of
Health Sciences, University of York

Professor Steven Goodacre,
Professor and Consultant in
Emergency Medicine, School of
Health and Related Research,
University of Sheffield

Professor Dyfrig Hughes,
Professor of Pharmacoeconomics,
Centre for Economics and Policy
in Health, Institute of Medical
and Social Care Research, Bangor
University

Professor Paul Jones,
Professor of Respiratory Medicine,
Department of Cardiac and
Vascular Science, St George's
Hospital Medical School,
University of London

Professor Khalid Khan,
Professor of Women's Health and
Clinical Epidemiology, Barts and
the London School of Medicine,
Queen Mary, University of London

Professor Richard J McManus,
Professor of Primary Care
Cardiovascular Research, Primary
Care Clinical Sciences Building,
University of Birmingham

Professor Helen Rodgers,
Professor of Stroke Care, Institute
for Ageing and Health, Newcastle
University

Professor Ken Stein,
Professor of Public Health,
Peninsula Technology Assessment
Group, Peninsula College
of Medicine and Dentistry,
Universities of Exeter and
Plymouth

Professor Jonathan Sterne,
Professor of Medical Statistics
and Epidemiology, Department
of Social Medicine, University of
Bristol

Mr Andy Vail,
Senior Lecturer, Health Sciences
Research Group, University of
Manchester

Professor Clare Wilkinson,
Professor of General Practice and
Director of Research North Wales
Clinical School, Department of
Primary Care and Public Health,
Cardiff University

Dr Ian B Wilkinson,
Senior Lecturer and Honorary
Consultant, Clinical Pharmacology
Unit, Department of Medicine,
University of Cambridge

Observers

Ms Kate Law,
Director of Clinical Trials,
Cancer Research UK

Dr Morven Roberts,
Clinical Trials Manager, Health
Services and Public Health
Services Board, Medical Research
Council

Current and past membership details of all HTA programme 'committees' are available from the HTA website (www.hta.ac.uk)

Diagnostic Technologies and Screening Panel

Members

Chair,
Professor Lindsay Wilson Turnbull,
Scientific Director of the Centre for Magnetic Resonance Investigations and YCR Professor of Radiology, Hull Royal Infirmary

Professor Judith E Adams, Consultant Radiologist, Manchester Royal Infirmary, Central Manchester & Manchester Children's University Hospitals NHS Trust, and Professor of Diagnostic Radiology, University of Manchester

Mr Angus S Arunkalaivanan, Honorary Senior Lecturer, University of Birmingham and Consultant Urogynaecologist and Obstetrician, City Hospital, Birmingham

Dr Diana Baralle, Consultant and Senior Lecturer in Clinical Genetics, University of Southampton

Dr Stephanie Dancer, Consultant Microbiologist, Hairmyres Hospital, East Kilbride

Dr Diane Eccles, Professor of Cancer Genetics, Wessex Clinical Genetics Service, Princess Anne Hospital

Dr Trevor Friedman, Consultant Liason Psychiatrist, Brandon Unit, Leicester General Hospital

Dr Ron Gray, Consultant, National Perinatal Epidemiology Unit, Institute of Health Sciences, University of Oxford

Professor Paul D Griffiths, Professor of Radiology, Academic Unit of Radiology, University of Sheffield

Mr Martin Hooper, Public contributor

Professor Anthony Robert Kendrick, Associate Dean for Clinical Research and Professor of Primary Medical Care, University of Southampton

Dr Nicola Lennard, Senior Medical Officer, MHRA

Dr Anne Mackie, Director of Programmes, UK National Screening Committee, London

Mr David Mathew, Public contributor

Dr Michael Millar, Consultant Senior Lecturer in Microbiology, Department of Pathology & Microbiology, Barts and The London NHS Trust, Royal London Hospital

Mrs Una Rennard, Public contributor

Dr Stuart Smellie, Consultant in Clinical Pathology, Bishop Auckland General Hospital

Ms Jane Smith, Consultant Ultrasound Practitioner, Leeds Teaching Hospital NHS Trust, Leeds

Dr Allison Streetly, Programme Director, NHS Sickle Cell and Thalassaemia Screening Programme, King's College School of Medicine

Dr Matthew Thompson, Senior Clinical Scientist and GP, Department of Primary Health Care, University of Oxford

Dr Alan J Williams, Consultant Physician, General and Respiratory Medicine, The Royal Bournemouth Hospital

Observers

Dr Tim Elliott, Team Leader, Cancer Screening, Department of Health

Dr Joanna Jenkinson, Board Secretary, Neurosciences and Mental Health Board (NMHB), Medical Research Council

Professor Julietta Patnick, Director, NHS Cancer Screening Programme, Sheffield

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health

Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

Dr Ursula Wells, Principal Research Officer, Policy Research Programme, Department of Health

Disease Prevention Panel

Members

Chair,
Professor Margaret Thorogood,
Professor of Epidemiology, University of Warwick Medical School, Coventry

Dr Robert Cook, Clinical Programmes Director, Bazian Ltd, London

Dr Colin Greaves, Senior Research Fellow, Peninsula Medical School (Primary Care)

Mr Michael Head, Public contributor

Professor Cathy Jackson, Professor of Primary Care Medicine, Bute Medical School, University of St Andrews

Dr Russell Jago, Senior Lecturer in Exercise, Nutrition and Health, Centre for Sport, Exercise and Health, University of Bristol

Dr Julie Mytton, Consultant in Child Public Health, NHS Bristol

Professor Irwin Nazareth, Professor of Primary Care and Director, Department of Primary Care and Population Sciences, University College London

Dr Richard Richards, Assistant Director of Public Health, Derbyshire County Primary Care Trust

Professor Ian Roberts, Professor of Epidemiology and Public Health, London School of Hygiene & Tropical Medicine

Dr Kenneth Robertson, Consultant Paediatrician, Royal Hospital for Sick Children, Glasgow

Dr Catherine Swann, Associate Director, Centre for Public Health Excellence, NICE

Mrs Jean Thurston, Public contributor

Professor David Weller, Head, School of Clinical Science and Community Health, University of Edinburgh

Observers

Ms Christine McGuire, Research & Development, Department of Health

Dr Kay Pattison, Senior NIHR Programme Manager, Department of Health

Professor Tom Walley, CBE, Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

External Devices and Physical Therapies Panel

Members

Chair,
Dr John Pounsford,
Consultant Physician North Bristol NHS Trust

Deputy Chair,
Professor E Andrea Nelson,
Reader in Wound Healing and Director of Research, University of Leeds

Professor Bipin Bhakta,
Charterhouse Professor in Rehabilitation Medicine, University of Leeds

Mrs Penny Calder,
Public contributor

Dr Dawn Carnes,
Senior Research Fellow, Barts and the London School of Medicine and Dentistry

Dr Emma Clark,
Clinician Scientist Fellow & Cons. Rheumatologist, University of Bristol

Mrs Anthea De Barton-Watson,
Public contributor

Professor Nadine Foster,
Professor of Musculoskeletal Health in Primary Care Arthritis Research, Keele University

Dr Shaheen Hamdy,
Clinical Senior Lecturer and Consultant Physician, University of Manchester

Professor Christine Norton,
Professor of Clinical Nursing Innovation, Bucks New University and Imperial College Healthcare NHS Trust

Dr Lorraine Pinnigton,
Associate Professor in Rehabilitation, University of Nottingham

Dr Kate Radford,
Senior Lecturer (Research), University of Central Lancashire

Mr Jim Reece,
Public contributor

Professor Maria Stokes,
Professor of Neuromusculoskeletal Rehabilitation, University of Southampton

Dr Pippa Tyrrell,
Senior Lecturer/Consultant, Salford Royal Foundation Hospitals' Trust and University of Manchester

Dr Nefyn Williams,
Clinical Senior Lecturer, Cardiff University

Observers

Dr Kay Pattison,
Senior NIHR Programme Manager, Department of Health

Dr Morven Roberts,
Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council

Professor Tom Walley, CBE,
Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy Research Programme, Department of Health

Interventional Procedures Panel

Members

Chair,
Professor Jonathan Michaels,
Professor of Vascular Surgery, University of Sheffield

Deputy Chair,
Mr Michael Thomas,
Consultant Colorectal Surgeon, Bristol Royal Infirmary

Mrs Isabel Boyer,
Public contributor

Mr Sankaran Chandra Sekharan,
Consultant Surgeon, Breast Surgery, Colchester Hospital University NHS Foundation Trust

Professor Nicholas Clarke,
Consultant Orthopaedic Surgeon, Southampton University Hospitals NHS Trust

Ms Leonie Cooke,
Public contributor

Mr Seumas Eckford,
Consultant in Obstetrics & Gynaecology, North Devon District Hospital

Professor Sam Eljamel,
Consultant Neurosurgeon, Ninewells Hospital and Medical School, Dundee

Dr Adele Fielding,
Senior Lecturer and Honorary Consultant in Haematology, University College London Medical School

Dr Matthew Hatton,
Consultant in Clinical Oncology, Sheffield Teaching Hospital Foundation Trust

Dr John Holden,
General Practitioner, Garswood Surgery, Wigan

Dr Fiona Lecky,
Senior Lecturer/Honorary Consultant in Emergency Medicine, University of Manchester/Salford Royal Hospitals NHS Foundation Trust

Dr Nadim Malik,
Consultant Cardiologist/Honorary Lecturer, University of Manchester

Mr Hisham Mehanna,
Consultant & Honorary Associate Professor, University Hospitals Coventry & Warwickshire NHS Trust

Dr Jane Montgomery,
Consultant in Anaesthetics and Critical Care, South Devon Healthcare NHS Foundation Trust

Professor Jon Moss,
Consultant Interventional Radiologist, North Glasgow Hospitals University NHS Trust

Dr Simon Padley,
Consultant Radiologist, Chelsea & Westminster Hospital

Dr Ashish Paul,
Medical Director, Bedfordshire PCT

Dr Sarah Purdy,
Consultant Senior Lecturer, University of Bristol

Dr Matthew Wilson,
Consultant Anaesthetist, Sheffield Teaching Hospitals NHS Foundation Trust

Professor Yit Chiun Yang,
Consultant Ophthalmologist, Royal Wolverhampton Hospitals NHS Trust

Observers

Dr Kay Pattison,
Senior NIHR Programme Manager, Department of Health

Dr Morven Roberts,
Clinical Trials Manager, Health Services and Public Health Services Board, Medical Research Council

Professor Tom Walley, CBE,
Director, NIHR HTA programme, Professor of Clinical Pharmacology, University of Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy Research Programme, Department of Health

Current and past membership details of all HTA programme 'committees' are available from the HTA website (www.hta.ac.uk)

Pharmaceuticals Panel

Members

Chair,
Professor Imti Choonara,
Professor in Child Health,
University of Nottingham

Deputy Chair,
Dr Yoon K Loke,
Senior Lecturer in Clinical
Pharmacology, University of East
Anglia

Dr Martin Ashton-Key,
Medical Advisor, National
Commissioning Group, NHS
London

Dr Peter Elton,
Director of Public Health, Bury
Primary Care Trust

Dr Ben Goldacre,
Research Fellow, Epidemiology
London School of Hygiene and
Tropical Medicine

Dr James Gray,
Consultant Microbiologist,
Department of Microbiology,
Birmingham Children's Hospital
NHS Foundation Trust

Dr Jurjees Hasan,
Consultant in Medical Oncology,
The Christie, Manchester

Dr Carl Heneghan,
Deputy Director Centre for
Evidence-Based Medicine and
Clinical Lecturer, Department of
Primary Health Care, University
of Oxford

Dr Dyfrig Hughes,
Reader in Pharmacoeconomics
and Deputy Director, Centre for
Economics and Policy in Health,
IMSCaR, Bangor University

Dr Maria Kouimtzi,
Pharmacy and Informatics
Director, Global Clinical Solutions,
Wiley-Blackwell

Professor Femi Oyebode,
Consultant Psychiatrist and Head
of Department, University of
Birmingham

Dr Andrew Prentice,
Senior Lecturer and Consultant
Obstetrician and Gynaecologist,
The Rosie Hospital, University of
Cambridge

Ms Amanda Roberts,
Public contributor

Dr Gillian Shepherd,
Director, Health and Clinical
Excellence, Merck Serono Ltd

Mrs Katrina Simister,
Assistant Director New Medicines,
National Prescribing Centre,
Liverpool

Professor Donald Singer,
Professor of Clinical
Pharmacology and Therapeutics,
Clinical Sciences Research
Institute, CSB, University of
Warwick Medical School

Mr David Symes,
Public contributor

Dr Arnold Zermansky,
General Practitioner, Senior
Research Fellow, Pharmacy
Practice and Medicines
Management Group, Leeds
University

Observers

Dr Kay Pattison,
Senior NIHR Programme
Manager, Department of Health

Mr Simon Reeve,
Head of Clinical and Cost-
Effectiveness, Medicines,
Pharmacy and Industry Group,
Department of Health

Dr Heike Weber,
Programme Manager, Medical
Research Council

Professor Tom Walley, CBE,
Director, NIHR HTA
programme, Professor of Clinical
Pharmacology, University of
Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy
Research Programme, Department
of Health

Psychological and Community Therapies Panel

Members

Chair,
Professor Scott Weich,
Professor of Psychiatry, University
of Warwick, Coventry

Deputy Chair,
Dr Howard Ring,
Consultant & University Lecturer
in Psychiatry, University of
Cambridge

Professor Jane Barlow,
Professor of Public Health in
the Early Years, Health Sciences
Research Institute, Warwick
Medical School

Dr Sabyasachi Bhaumik,
Consultant Psychiatrist,
Leicestershire Partnership NHS
Trust

Mrs Val Carlill,
Public contributor

Dr Steve Cunningham,
Consultant Respiratory
Paediatrician, Lothian Health
Board

Dr Anne Hesketh,
Senior Clinical Lecturer in Speech
and Language Therapy, University
of Manchester

Dr Peter Langdon,
Senior Clinical Lecturer, School
of Medicine, Health Policy and
Practice, University of East Anglia

Dr Yann Lefeuvre,
GP Partner, Burrage Road Surgery,
London

Dr Jeremy J Murphy,
Consultant Physician and
Cardiologist, County Durham and
Darlington Foundation Trust

Dr Richard Neal,
Clinical Senior Lecturer in General
Practice, Cardiff University

Mr John Needham,
Public contributor

Ms Mary Nettle,
Mental Health User Consultant

Professor John Potter,
Professor of Ageing and Stroke
Medicine, University of East
Anglia

Dr Greta Rait,
Senior Clinical Lecturer and
General Practitioner, University
College London

Dr Paul Ramchandani,
Senior Research Fellow/Cons.
Child Psychiatrist, University of
Oxford

Dr Karen Roberts,
Nurse/Consultant, Dunston Hill
Hospital, Tyne and Wear

Dr Karim Saad,
Consultant in Old Age Psychiatry,
Coventry and Warwickshire
Partnership Trust

Dr Lesley Stockton,
Lecturer, School of Health
Sciences, University of Liverpool

Dr Simon Wright,
GP Partner, Walkden Medical
Centre, Manchester

Observers

Dr Kay Pattison,
Senior NIHR Programme
Manager, Department of Health

Dr Morven Roberts,
Clinical Trials Manager, Health
Services and Public Health
Services Board, Medical Research
Council

Professor Tom Walley, CBE,
Director, NIHR HTA
programme, Professor of Clinical
Pharmacology, University of
Liverpool

Dr Ursula Wells,
Principal Research Officer, Policy
Research Programme, Department
of Health

Paper 2.

Soares MO, Welton NJ, Harrison DA, Peura P, Shankar-Hari M, Harvey SE, Madan J, Ades AE, Rowan KM, Palmer SJ. Intravenous immunoglobulin for severe sepsis and septic shock: clinical effectiveness, cost-effectiveness and value of a further randomised controlled trial. Critical Care. 2014 Dec 1;18(6):649.

CRITICAL CARE

Intravenous immunoglobulin for severe sepsis and septic shock: clinical effectiveness, cost-effectiveness and value of a further randomised controlled trial

Marta O Soares[1]*, Nicky J Welton[2], David A Harrison[3], Piia Peura[1], Manu Shankar-Hari[4], Sheila E Harvey[3], Jason Madan[5], Anthony E Ades[2], Kathryn M Rowan[3] and Stephen J Palmer[1]

Abstract

Introduction: Prior to investing in a large, multicentre randomised controlled trial (RCT), the National Institute for Health Research in the UK called for an evaluation of the feasibility and value for money of undertaking a trial on intravenous immunoglobulin (IVIG) as an adjuvant therapy for severe sepsis/septic shock.

Methods: In response to this call, this study assessed the clinical and cost-effectiveness of IVIG (using a decision model), and evaluated the value of conducting an RCT (using expected value of information (EVI) analysis). The evidence informing such assessments was obtained through a series of systematic reviews and meta-analyses. Further primary data analyses were also undertaken using the Intensive Care National Audit & Research Centre Case Mix Programme Database, and a Scottish Intensive Care Society research study.

Results: We found a large degree of statistical heterogeneity in the clinical evidence on treatment effect, and the source of such heterogeneity was unclear. The incremental cost-effectiveness ratio of IVIG is within the borderline region of estimates considered to represent value for money, but results appear highly sensitive to the choice of model used for clinical effectiveness. This was also the case with EVI estimates, with maximum payoffs from conducting a further clinical trial between £137 and £1,011 million.

Conclusions: Our analyses suggest that there is a need for a further RCT. Results on the value of conducting such research, however, were sensitive to the clinical effectiveness model used, reflecting the high level of heterogeneity in the evidence base.

Introduction

Sepsis is a clinical syndrome defined by the presence of both infection and a systemic inflammatory response; sepsis is defined as severe when associated with, or complicated by, organ dysfunction [1]. Severe sepsis may induce septic shock, defined as hypotension persisting despite adequate fluid resuscitation [2]. There is evidence indicating an increasing incidence of severe sepsis treated in critical care in England, Wales and Northern Ireland, rising from 50 to 70 cases per 100,000 population per year

between 1995 and 2005 - these cases being associated with approximately 31,000 episodes of severe sepsis and 15,000 in-hospital deaths per year [3]. Being a serious, life-threatening condition, severe sepsis is expected to be associated with substantial healthcare costs and a significant impact on quality of life.

Intravenous immunoglobulin (IVIG) is a scarce blood product derived from human donor blood; it is currently subject to a Demand Management Programme by the United Kingdom (UK) Department of Health [4]. This product has been proposed as an adjuvant therapy for severe sepsis/septic shock since the 1980s. However, the mechanisms of action of IVIG are complex and are not yet fully understood. Despite this, a number of (predominantly

* Correspondence: marta.soares@york.ac.uk
[1]Centre for Health Economics, University of York, Heslington, York, YO10 5DD, UK
Full list of author information is available at the end of the article

small) randomised controlled trials (RCTs) have been conducted, and numerous systematic reviews and meta-analyses have been undertaken to synthesise their findings [5-8]. As a result of the heterogeneity across studies and the inconsistencies in their results, the majority conclude that there is currently insufficient evidence to recommend IVIG as an adjuvant therapy and that more evidence, in the form of a large, well-conducted RCT, is required.

Prior to investing in a large, multicentre RCT, the Health Technology Assessment (HTA) Programme of the National Institute for Health Research in the UK called for an evaluation of the feasibility and value for money of undertaking such a trial (that is, whether or not the costs of undertaking the trial are outweighed by the potential benefit of the resulting information). The aim of this manuscript is to report our findings in response to this call by assessing the clinical and cost-effectiveness of IVIG in severe sepsis/septic shock in adults, and evaluating the value of conducting a large, multicentre RCT using an expected value of information (EVI) analysis. EVI offers a methodological framework that explicitly considers the uncertainty surrounding the decision by a healthcare system to adopt a health technology and values the additional information, which may be generated by further research, in a way that is consistent with the objectives and resource constraints of heathcare provision [9].

A full technical report of this research is published elsewhere [10]. Here, we provide a summary of the evaluation of clinical effectiveness, cost-effectiveness, and value of information of IVIG in critically ill adults with severe sepsis/septic shock, undertaken to inform the UK policy context. We discuss the implications arising from this policy-driven evidence review.

Methods

We conducted a series of formal systematic reviews and undertook additional primary data analysis to develop and populate a decision analytic model. Details of each review and data sources are presented in the full technical report [10]. The decision model evaluated the cost-effectiveness of IVIG as an adjunctive treatment to standard care for the management of adults with severe sepsis (we have used the term severe sepsis in this manuscript to include septic shock). The base-case population in the model reflected the baseline characteristics of the severe sepsis population in the Intensive Care National Audit & Research Centre (ICNARC) Case Mix Programme Database (CMPD), a high-quality clinical database of admissions to adult general (mixed medical/surgical) critical care units in the UK, considered to be more representative of current UK National Health Service (NHS) practice than the populations recruited

into RCTs. We used data for the years 2007 to 2009, corresponding to a sample of 26,249 patients.

The decision model evaluated costs from the perspective of the NHS and Personal Social Services (PSS), and expressed these in British pounds sterling at a 2009 price base. Outcomes were expressed in quality-adjusted life years (QALYs). Both costs and outcomes were discounted using a 3.5% annual discount rate, in line with current guidelines [11]. The model was probabilistic, that is, uncertainty over the input parameters was propagated through the model in such a way that the results of the analysis could be presented with their associated uncertainty [12]. The expected costs and QALYs for IVIG and standard care (SC) were estimated and compared using incremental cost-effectiveness ratios (ICERs) that represent the incremental cost per additional QALY. The ICER was compared against thresholds used to establish value for money in the NHS (currently in the region of £20,000 to £30,000 per QALY) [11].

The probabilistic analysis also provided a formal approach to quantifying the consequences associated with the uncertainty surrounding the model results, which were then used to inform the EVI analyses. The maximum amount the NHS should be willing to invest to reduce uncertainty in the decision can be informed by the expected value of perfect information (EVPI) [13,14]. The EVPI evaluates the expected cost of current uncertainty by accounting both for the probability that a decision based on existing evidence is wrong and for the magnitude of the consequences of making the wrong decisions. EVPI can be expressed at a population level based on the size of the population (yearly incidence = 33,160) and the number of years research is assumed useful (10 years). EVPI can also be estimated for individual parameters (or for groups of parameters) contained in the model, termed partial EVPI or EVPPI. Five groups of uncertain parameters were considered: i) baseline mortality during the initial acute hospitalisation with SC; ii) clinical effectiveness of IVIG; iii) long-term mortality estimates for survivors of severe sepsis; iv) long-term costs for survivors of severe sepsis; v) quality of life for survivors of severe sepsis. The groups of parameters also reflect potentially different research designs. For example, while an RCT would ideally be required to further inform the clinical effectiveness of IVIG, evidence on the other parameters could be generated using record linkage of existing databases.

The EVPI estimates set an upper limit on the returns to further research. However, to fully inform the research decision, the most efficient research design also needs to be established, for example, the type of study, the optimal sample size, the appropriate duration of follow-up and appropriate end points. The same framework of analyses can be extended to establish the expected value of sample

information (EVSI) for a particular research design. To obtain the societal payoff to the proposed research, the population EVSI needs to be compared with the costs of sampling: the difference between the EVSI and the costs of sampling gives the expected net benefit of sampling (ENBS). The ENBS provides a necessary and sufficient condition for deciding to conduct more research that is, if the ENBS is greater than zero for any sample size then the benefits of gathering the sample information exceed the costs, and further research is potentially justified. The ENBS also provides a framework for the efficient design of the clinical trial, where the optimal sample size, n^*, for the proposed trial is where the ENBS reaches its maximum. This optimal sample size thus indicates how many patients should be enrolled for the trial to provide the highest payoff.

All stages of the work were informed through discussions with an expert clinical advisory group who provided feedback on specific aspects of the analyses including the model structure, inputs and assumptions.

Decision model structure

A simplified schematic of the structure of the decision model is presented in Figure 1. The model evaluated the lifetime prognosis of severe sepsis in order to capture the longer-term costs and consequences associated with the natural history in the absence of IVIG. The model structure considered two related elements reflecting short- and long-term consequences:

i. short-term: the short-term consequences of the initial severe sepsis episode reflected the initial acute hospitalisation period (acute hospital and critical care) and related to the probability of surviving the initial acute hospitalisation. Baseline data from the CMPD were used to estimate the risk of acute hospital mortality associated with SC and the results of a systematic review of the clinical

Figure 1 Model schematic.

effectiveness of IVIG were applied to estimate the risk of acute hospital mortality associated with IVIG.

ii. longer-term: conditional on having survived the initial acute hospitalisation, a Markov structure was used to characterise the long-term prognosis over the remainder of a patient's lifetime. Such a model represents disease progression by defining important mutually exclusive events/health states and characterising how patients may move between states over regular time intervals. Here, we used a simple structure to represent the possibility of patients being alive or dead (a survival model). Annual cycles were employed to reflect the annual probability of death for each year after the initial severe sepsis episode. Duration of stay alive or transition to death can be associated with a quality of life score and with costs incurred, allowing for long-term QALYs and costs to be evaluated. If IVIG reduces the risk of mortality during the initial acute hospitalisation period, then the use of this longer-term model will allow differences in long-term costs and QALYs to be translated.

The model was developed in the statistical programming package R [15].

Clinical effectiveness of IVIG

Evidence on the effectiveness of IVIG was sought and used to inform the decision model and further analyses. A systematic review of the literature was conducted to inform the effect of IVIG in severe sepsis on all-cause mortality during the initial acute hospitalisation. Full details of the methods of the systematic review are reported in the full technical report [10]. In summary, RCTs within a critical care setting that compared any standard polyclonal IVIG or immunoglobulin M-enriched polyclonal IVIG (IVIGAM) with either no intervention, placebo (usually albumin) or another standard polyclonal IVIG or IVIGAM preparation were considered eligible for inclusion. Studies were included if the majority of patients were aged ≥18 years and clinical judgement deemed the population studied to have severe sepsis.

The primary outcome measure extracted was all-cause mortality, summarised on the odds ratio (OR) scale. Fixed- and random-effects meta-analyses were fitted using inverse variance weights. Heterogeneity was assessed using I^2, and Cochrans Q test. Stata, version 11.0 (StataCorp LP, College Station, TX, USA), was used.

Heterogeneity was further explored comparing model fit from meta-regressions, estimated through Bayesian [16] Markov chain Monte Carlo simulation [17] using WinBUGS version 1.4.3 (MRC Biostatistics Unit, Cambridge, UK) [18,19]. More detail on the use of Bayesian methods in meta-analysis and evidence synthesis can be found in Sutton and Abrams [20], and in Spiegelhalter *et al.* [21].
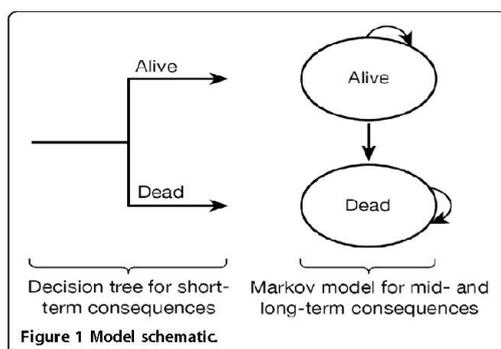
Results are presented using means and 95% credible intervals, 95%CrI (the Bayesian equivalent to confidence intervals) [20]. The posterior mean residual deviance (Dres) was used to measure model fit and the deviance information criterion (DIC), a composite measure of model fit and model complexity, was used to choose between competing models [22]. For the random-effects models, the posterior mean of the between-study standard deviation (SD) parameter (τ) was used to investigate the impact of the inclusion of the covariates on explaining (reducing) heterogeneity. Meta-regressions aimed to: i) identify key covariates responsible for heterogeneity; ii) consider more complex treatment models that compared different types and preparations of IVIG; and iii) adjust for potential confounding by considering combinations of covariates. The covariates evaluated related to: characteristics of treatments such as type and dose of IVIG; features of study design related to, or a proxy for, study quality (for example Jadad score - a composite measure ranging from 0 to 5 where 5 indicates best quality and sample size); setting; acute severity of disease; and follow-up period. Key covariates that explained some of the heterogeneity using model fit statistics (Dres, DIC and τ) were identified. In addition, combinations of key potential covariates were explored to identify which of the covariates best explained the heterogeneity, after having adjusted for other covariates. Other treatment models were also explored, in which the type of IVIG preparation and type of control were not grouped together. All treatment and covariate models were compared using the model fit statistics (Dres, DIC and τ). Results were reported for the best-fitting, competing models.

Other inputs of the decision model

Baseline event rates for standard care (initial acute hospitalisation)

Data from the CMPD (n = 26,249) were used to inform the baseline risk of acute hospital mortality applied to SC during the initial hospitalisation. Mortality risk was estimated by conditioning on characteristics of patients and the severity of illness at presentation: age and gender; Acute Physiology and Chronic Health Evaluation II (APACHE II) score; ICNARC physiology score and number of dysfunctional organ systems. Logistic regressions (robust standard errors (SEs) adjusting for clustering on critical care unit) were used.

Longer-term survival

We undertook additional primary data analysis to inform the longer-term survival estimates for severe sepsis survivors based on a cohort of 345 subjects from the Scottish Intensive Care Society (SICS) prospective, observational, multicentre, epidemiological study of severe sepsis [23]. Only patients (n = 271) for whom organ system dysfunction

was clearly reported were selected. Average follow-up for survival was 787 days (range 0 to 2,062) days.

Parametric survival analyses were undertaken to estimate longer-term mortality (goodness of fit was assessed using Akaike information criterion (AIC) statistics). Three separate models were fitted including additional covariates for: age; APACHE II score at admission; and organ system dysfunction. The covariates were included to adjust for potential imbalances between the baseline characteristics from the CMPD (used to estimate short-term mortality) and the SICS study cohort. Also, these covariates allowed for consideration of subgroup-specific estimates for longer-term survival.

Resource use, unit costs and health-related quality of life

The length of stay (LOS) during the index severe sepsis episode (from the CMPD) was used to cost the acute hospital stay. Unit costs were derived from national databases. A summary of the information used is presented in Table 1 and further details of the results of two separate literature reviews, on health-related quality of life (HRQoL) and resource use, are available in the full technical report [10].

Results

Table 1 provides a summary of the model input parameters used (both from the systematic reviews and from the primary data analysis) to inform the cost-effectiveness of IVIG.

Clinical effectiveness of IVIG

Seventeen studies were identified that met the inclusion criteria with a large degree of heterogeneity in treatment effect between studies [27-40]. Figure 2 presents a funnel plot of the SE of the effect size (log-OR) plotted against study effect size (OR on the log scale). The asymmetry suggests there may be publication bias with this evidence and further exploratory analyses were conducted. Unadjusted fixed- and random-effects meta-analyses were implemented, initially comparing two treatment groupings (IVIG or IVIGAM vs. albumin/no treatment) (Figure 3). Given the heterogeneity (I^2 = 46.9%), a random-effects model fitted the data well while the fixed-effect model showed substantial lack of fit. The pooled OR from the random-effects model was 0.47 (95% CI 0.32 to 0.69), showing a stronger effect than the fixed-effect model (pooled OR = 0.68; 95% CI 0.54 to 0.84).

From the meta-regression analyses, key covariates appearing to explain the existing heterogeneity between studies were: dosing regimen (duration of treatment (days), daily dose (g kg^{-1} day^{-1}), volume (ml kg^{-1} day^{-1})); and study quality (that is use of albumin as control - as a proxy for proper blinding to treatment, Jadad score, publication date, and a measure of sample size; 1/ N).

Table 1 Inputs of the decision model for the evaluation of IVIG for severe sepsis/septic shock: parameter values and uncertainty over parameter values

Parameter	Source	Base case	Notes on scenario analyses undertaken	Notes on subgroup analyses
Cohort characteristics				
Mean age of a severe sepsis patient at admission to hospital	CMPD	63 years old	Same as base case	Assumed to vary in subgroups defined using age. Sourced from ICNARC database
Proportion of males in a severe sepsis population at admission to hospital	CMPD	0.53	Same as base case	Assumed to vary in subgroups defined using gender. Sourced from CMPD
Short-term outcomes (ST)				
Probability of dying in hospital when SC is used in the treatment of severe sepsis (baseline risk)	CMPD	40.6%, 95% CI (40%, 41.2%)	Same as base case	Assumed to vary per subgroup (all). Sourced from CMPD
Odds ratio, when IVIG is used to complement SC in the treatment of severe sepsis (based on Model M1)	Evidence synthesis ('Clinical effectiveness of IVIG')	0.75 , 95% CI (0.58, 0.96)	Alternative models tested; see Table 2	Same as base case
Longer-term outcomes (LT)				
Age specific probability of dying in yearly intervals, conditional on patients having survived up to the start of the year.	Cuthbertson database and general population life tables 2010	Figure 4. Varies with time.	(1) time horizon (2) time points at which patients reverted to survival of general population	Assumed to vary for subgroups defined using age and APACHE II score. Sourced from CMPD
Cost-related parameters				
Costs of overall IVIG therapy	Non-stochastic, BNF	£5,539.05	Same as base case	Same as base case
Costs of SC, when only SC is used in the treatment of severe sepsis	Non-stochastic,	£0	Same as base case	Same as base case
LOS in ICU for patients remaining alive until discharge from hospital	CMPD	8.48 (SE = 0.086)	Same as base case	Assumed to vary for all subgroups. Sourced from CMPD
LOS in ICU for patients dying in hospital	CMPD	7.40 (SE = 0.108)	Same as base case	Assumed to vary for all subgroups. Sourced from CMPD
Costs associated to a day in ICU for a patient with severe sepsis	Non-stochastic, reference costs [24]	£1,393	Same as base case	Same as base case
Overall hospital LOS for patients remaining alive until discharge from hospital	CMPD	21.29 (SE = 0.292)	Same as base case	Assumed to vary for all subgroups. Sourced from CMPD
Overall hospital LOS for patients dying in hospital	CMPD	39.07 (SE = 0.325)	Same as base case	Assumed to vary for all subgroups. Sourced from CMPD
Costs associated to a day in wards other than ICU for a patient with a severe sepsis episode	Non-stochastic, reference costs [24]	£196	Same as base case	Same as base case
Costs incurred between year t-1 and year t after hospital discharge	Manns [25]	$t = 1$: £13,654 and $t > 1$: £4,466.5/year	(1) ± 50% of Manns' estimates (2) average annual per capital NHS cost for the general population	Same as base case
Utilities				
In-hospital HRQoL weight associated to severe sepsis patients	Drabinsky [26]	0.53	Same as base case	Same as base case
HRQoL weight associated to severe sepsis patients at year t	Cuthbertson database, Drabinsky [26]	$t = 1$: 0.62 and $t > 1$: 0.6833	Same as base case	Same as base case

See NIHR HTA full technical report [10] for a comprehensive reporting of parameter values used to inform the decision model, including those used in subgroup analyses. IVIG, intravenous immunoglobulin; CMPD, Case Mix Programme Database; ICNARC, Intensive Care National Audit & Research Centre; CI, confidence interval; BNF, British National Formulary; SC, standard care; APACHE II, Acute Physiology and Chronic Health Evaluation II; LOS, length of stay; HRQoL, health-related quality of life.

Figure 2 Evidence on the clinical effectiveness of IVIG for severe sepsis (and septic shock): publication bias - funnel plot (with pseudo 95% confidence limits) for mortality of IVIG and IVGAM versus control. IVIG, intravenous immunoglobulin; IVIGAM, immunoglobulin M-enriched polyclonal IVIG.
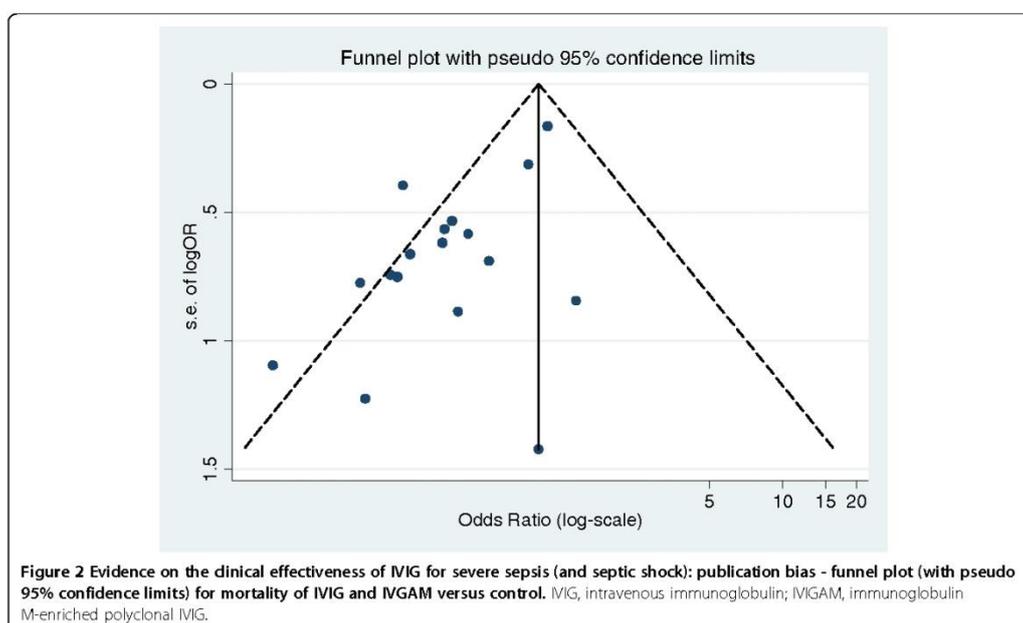
These two key covariates explained the majority of the heterogeneity in treatment effect across the studies. Further detail on these analyses is presented in the full technical report [10].

Across all the models considered [10], the best-fitting model assessed three treatments: IVIG/IVIGAM vs. albumin vs. no treatment and included duration of IVIG therapy as a treatment effect-modifying covariate. Results are reported for the most commonly used duration of therapy reported: three days (model M1 in Table 2). The OR for IVIG/IVIGAM vs. albumin was estimated at 0.75 (with a 95% CrI of 0.58 to 0.96) indicating a reduction in the odds of all-cause mortality for patients with severe sepsis compared with albumin.

Discussions with the expert clinical advisory group highlighted that there was no clear clinical rationale why duration of treatment would affect treatment effectiveness. For this reason, random-effects models with solely study quality covariates were also considered. The heterogeneity that can be explained with the dosing regimen covariates was left unexplained in these models, reflecting a belief that these covariates were a proxy for other, unmeasured, differences between the included studies. The results of a range of alternative models are presented in Table 2 (Models M2 to M4b).

When the heterogeneity explained by duration of IVIG therapy in M1 was treated as unexplained (that is using a random-effects model, models M2 to M4b), the majority

of results were fairly comparable and still showed a reduction in the odds of all-cause mortality in patients with severe sepsis treated with IVIG but the 95% CrI widened suggesting a larger degree of uncertainty. The only exception was model M4b, which reported an increase in the odds of all-cause mortality, albeit with very wide credible intervals. However, some caution should be applied to this result since it involves extrapolation beyond the available data. In the absence of a single best-fitting model that made clinical sense, sensitivity analyses were subsequently used in the cost-effectiveness modelling.

Other input parameters to the decision model
Model input parameters and uncertainty around their estimates are described in Table 1 (and, in greater detail, in the full technical report [10]). Briefly, the probability of dying during the initial hospitalisation was estimated from the CMPD to be 40.6% (95% CI 40.0%, 41.2%). With respect to the evidence on longer-term mortality, we initially investigated the plausibility of the different parametric predictions beyond the five years of observed data by comparing these to age-adjusted estimates from the general population (Figure 4a). It was considered implausible that the longer-term mortality estimates for severe sepsis patients would become lower than that for the general population. Consequently, in the model, we further assumed that the probability of mortality would
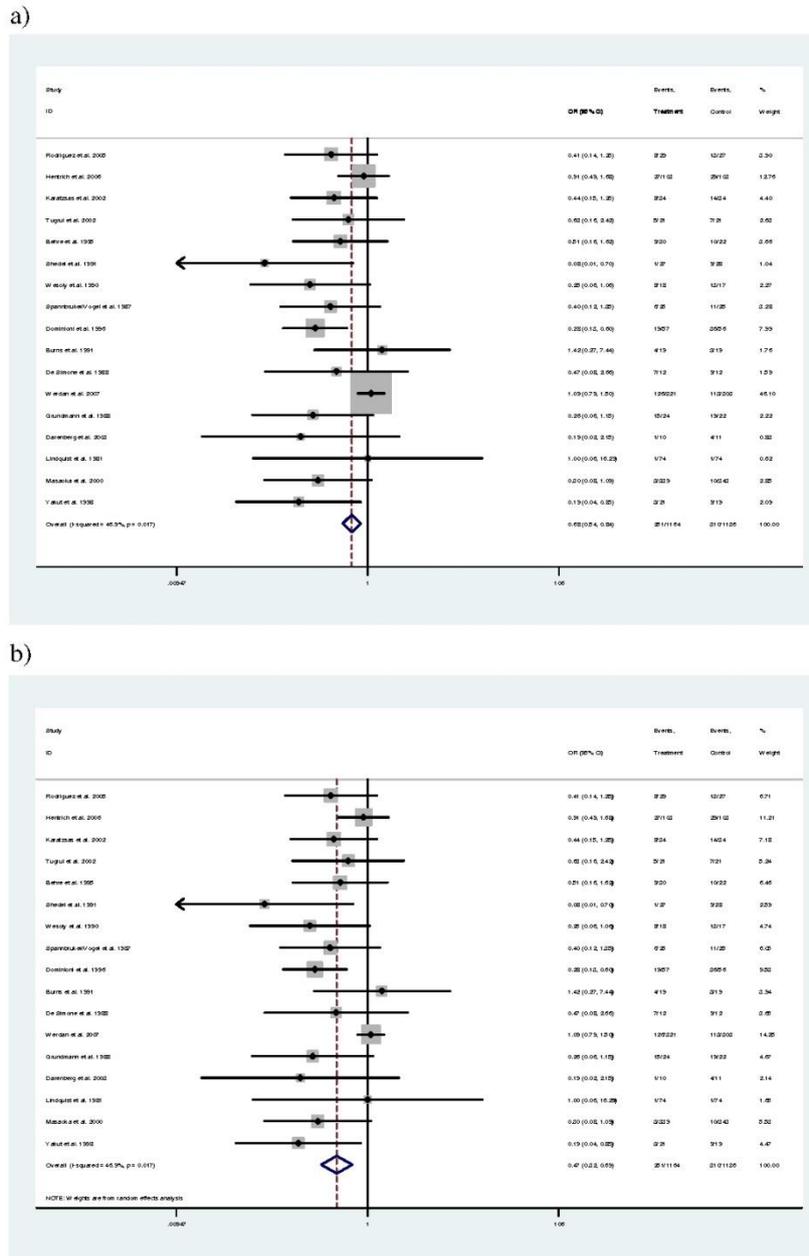
Figure 3 Evidence on the clinical effectiveness of IVIG for severe sepsis (and septic shock): Forest plots for (a) fixed-effects model using inverse variance weights, and (b) random-effects model using inverse variance weights. Both evaluate IVIG and IVIGAM treatments versus control. IVIG, intravenous immunoglobulin; IVIGAM, immunoglobulin M-enriched polyclonal IVIG.

Table 2 Evidence on the clinical effectiveness of IVIG for severe sepsis/septic shock: estimates from the best-fitting models for the synthesis of evidence

Model for the synthesis of clinical effectiveness evidence	Odds ratio (95% CrI)
M1: Fixed-effect model considering three treatments: IVIG/IVIGAM vs. albumin vs. no treatment, with covariate on duration of IVIG therapy. Relative effectiveness estimate reported for IVIG/IVIGAM vs. albumin for a duration of therapy of three days.	0.75 (0.58, 0.96)
M2: Random-effects model considering three treatments: IVIG/IVIGAM vs. albumin vs. no treatment, no covariates. Relative effectiveness estimate reported for IVIG/IVIGAM vs. albumin.	0.68 (0.16, 1.83)
M3: Random-effects model considering two treatments: IVIG/IVIGAM vs. albumin or no treatment, with covariate on Jadad score. Relative effectiveness estimate reported assuming a Jadad score of 5.	0.83 (0.18, 2.13)
M4a: Random-effects model considering two treatments: IVIG/IVIGAM vs. albumin/no treatment, with covariate representing $1/\sqrt{N}$. Relative effectiveness estimate reported assuming a sample size of 339 patients*	0.92 (0.23, 2.10)
M4b: Random-effects model considering two treatments: IVIG or IVIGAM vs. albumin or no treatment, with covariates representing $1/\sqrt{N}$. Relative effectiveness estimate assumes an infinitely large sample size*	1.27 (0.25, 3.17)

*With model M4 two cases were considered: for M4a, the sample size N was set equal to the maximum arm size in the studies in our review - avoiding extrapolation beyond the dataset; for M4b, sample size was set to infinity; this demonstrates the effect on model estimates of the absence of bias associated with study quality, here proxied by finite (and small) sample sizes. CrI, credible interval, the Bayesian equivalent to confidence intervals; IVIG, intravenous immunoglobulin; IVIGAM, immunoglobulin M-enriched polyclonal IVIG.

be the maximum of the predicted parametric distributions and the observed yearly probability of mortality for the general population (age- and sex-adjusted). The 'modified' parametric survival functions are reported in Figure 4b. The distribution with the best statistical goodness of fit was the Weibull function, and this was further used in the decision model [10].

With respect to resource use, we obtained estimates from the literature of total costs per annum and used these to describe costs after discharge from the index acute hospitalization [25]. The evidence used distinguishes the first year from subsequent years. With respect to HRQoL weights, we found evidence in the literature that the quality of life of survivors could be represented by a value of 0.69; also based on the available literature we assigned additional decrements to the within acute hospital period (0.09) and to the first month after acute hospitalisation (0.06).

Cost-effectiveness of IVIG

Table 3 reports the cost-effectiveness results using the best-fitting clinical effectiveness model for all-cause mortality and for each of the alternative models from the clinical effectiveness review considered within the sensitivity analyses. The results for the best-fitting model show that the ICER of IVIG is £20,850 per QALY, which is within the borderline region of estimates considered to represent value for money in the NHS. At a threshold of £20,000 per QALY, the probability that IVIG is cost-effective is 0.505. As the threshold cost per QALY increased, the probability that IVIG is cost-effective increased (that is increasing to 0.789 at a threshold of £30,000).

For the alternative clinical effectiveness models, the ICER estimates vary between £16,177 per QALY to IVIG being dominated by SC alone (that is IVIG being both less effective and more costly). These results clearly demonstrate

that any conclusions regarding the cost-effectiveness of IVIG appear highly sensitive to the choice of model used for clinical effectiveness. Furthermore, the level of decision uncertainty, expressed in terms of the probability that IVIG is cost-effective, remains high across all these scenarios.

Value of further information

Table 4 presents a summary of the population EVPI estimates based on a cost-effectiveness threshold of £20,000 per QALY. The results demonstrate a considerable range in the population EVPI estimates depending on the model applied to estimate the relative clinical effectiveness of IVIG. As expected, the random-effects model gave higher EVPI estimates given the additional between-study heterogeneity that is included. For a time horizon of 10 years, population EVPI varied between approximately £393 million and £1.4 billion. These results clearly suggested that further primary research would appear to be potentially worthwhile given the high cost of current decision uncertainty across all scenarios.

Figure 5 presents EVPPI estimates for the five groups of uncertain parameters for each of the clinical effectiveness models. The EVPPI associated with the relative treatment effect of IVIG consistently emerged as having significant influence on the overall decision uncertainty: the lowest estimate of EVPPI for the relative effect of IVIG was £173.7 million. The longer-term costs of severe sepsis also seem to be an important driver, with significant value in all except one of the scenarios. It should be appreciated that the costs of undertaking research on parameters such as quality of life would be significantly lower for these than for those required to undertake a large, multicentre RCT.

The population ENBS and the optimal sample size, n^*, for a proposed trial are reported in Table 5 (for a threshold of £20,000). Calculations assumed the costs of the trial to be based on a fixed cost component of

Figure 4 Longer-term survival of patients after an acute episode of severe sepsis (or septic shock): comparison of parametric survival functions (a) and their modified versions (b) to general population. The curves in Figure 3a investigate the plausibility of the different parametric (Weibull, Exponential and Lognormal) predictions beyond the five years of observed data by comparing these to age-adjusted estimates from the general population. These show that the longer-term mortality parametric estimates for severe sepsis patients become lower than that for the general population. This was considered implausible and, consequently, we modified these distributions by further assuming that the probability of mortality would be the maximum of the predicted parametric distributions and the observed yearly probability of mortality for the general population (age- and sex-adjusted). The 'modified' parametric survival functions are reported in Figure 4b.

£2 million and a variable cost component of £2,000 per patient recruited (+ £5,500 for patients receiving IVIG). Across scenarios, the maximum payoff from conducting this research (the ENBS) varies between £137 and £1,011 million. The optimal sample size always exceeded 800 subjects for each arm.

Discussion

Despite the existence of a large number of RCTs and numerous previous systematic reviews and meta-analyses of the relative effectiveness of IVIG, there remains controversy surrounding the value of IVIG as an adjunctive treatment in severe sepsis. Our study is the first to combine a formal systematic review of clinical effectiveness together with other epidemiological, resource use and quality of life data in order to robustly assess both the cost-effectiveness of IVIG based on existing evidence as well as the value of conducting further research.

Within this work, we re-analysed existing relative effectiveness RCT evidence and conducted a new meta-analysis, the first to simultaneously allow for type of IVIG (IVIG or IVIGAM), choice of control (no treatment or albumin), study quality/publication bias and other potential covariates. Our results indicated that treatment with IVIG may be associated with lower mortality but the evidence base shows a large degree of heterogeneity between individual studies. Given it was unclear what was the relevant source of heterogeneity in the evidence base, alternative clinical effectiveness models were evaluated.

Table 3 Cost-effectiveness of IVIG for severe sepsis/septic shock using the best-fitting and alternative synthesis models of effectiveness evidence (see Table 2 for detailed specification of the models)

(Best-fitting model, **M1**) Fixed-effect model estimate (IVIG/IVIGAM compared with albumin) considering 3 days duration of IVIG therapy				Probability of being cost-effective for cost-effectiveness threshold	
Treatment	Mean cost	Mean QALY	ICER	£20,000/QALY	£30,000/QALY
IVIG	£54,901	4.35	£20,850	0.505	0.789
Standard care	£45,593	3.90		0.495	0.211
(Alternative model **M2**) Random-effects model estimate (IVIG/IVIGAM compared with albumin)					
Treatment	Mean cost	Mean QALY	ICER	£20,000/QALY	£30,000/QALY
IVIG	£57,200	4.62	£16,177	0.597	0.707
Standard care	£45,593	3.90		0.403	0.295
(Alternative model **M3**) Random-effects model (IVIG/IVIGAM compared with albumin/no treatments) considering Jadad score = 5					
Treatment	Mean cost	Mean QALY	ICER	£20,000/QALY	£30,000/QALY
IVIG	£55,238	4.39	£19,968	0.502	0.611
Standard care	£45,593	3.90		0.498	0.389
(Alternative model **M4a**) Random-effects model (IVIG/IVIGAM compared with albumin/no treatment) considering a sample size of 339					
Treatment	Mean cost	Mean QALY	ICER	£20,000/QALY	£30,000/QALY
IVIG	£53,518	4.18	£28,520	0.404	0.514
Standard care	£45,593	3.90		0.596	0.486
(Alternative model **M4b**) Random-effects model (IVIG/IVIGAM compared with albumin/no treatment) considering a sample size of infinity					
Treatment	Mean cost	Mean QALY	ICER	£20,000/QALY	£30,000/QALY
IVIG	£50,024	3.76	Dominated	0.275	0.348
Standard care	£45,593	3.90		0.725	0.652

IVIG, intravenous immunoglobulin; IVIGAM, immunoglobulin M-enriched polyclonal IVIG; QALY, quality-adjusted life years; ICER, incremental cost-effectiveness ratio.

Our results on the cost-effectiveness of IVIG appear within the borderline region of estimates considered to represent value for money in the NHS, but these results are associated with significant decision uncertainty and appear highly sensitive to the alternative clinical effectiveness models applied. Using an expected value of information

Table 4 Value of further research on IVIG for severe sepsis/septic shock: population EVPI estimates (WTP = £20,000), according to alternative synthesis models of effectiveness evidence (see Table 2 for detailed specification of the models)
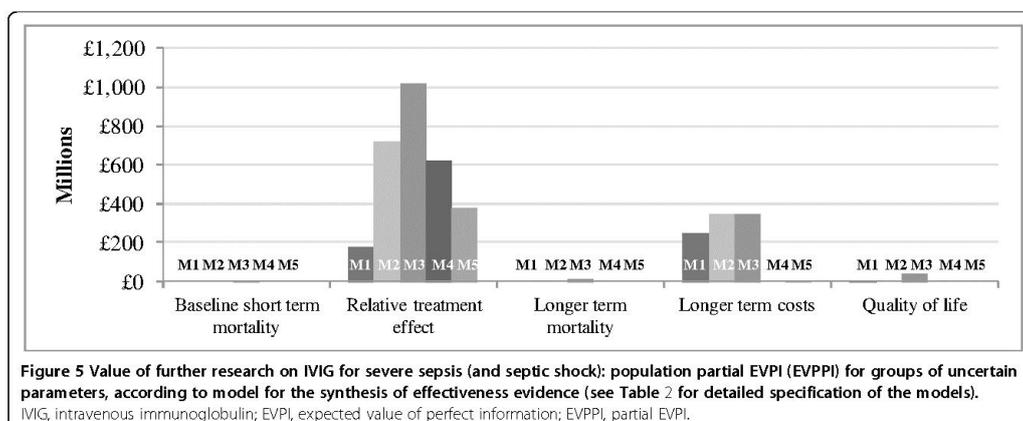
Scenarios for alternative synthesis models of effectiveness evidence	EVPI per patient	Population EVPI (*Time horizon* = 10 years)
M1	£1,377	£392,994,216
M2	£3,563	£1,017,023,732
M3	£4,791	£1,367,426,550
M4a	£3,146	£897,945,285
M4b	£2,113	£603,018,958

EVPI evaluates the expected cost of current uncertainty by accounting both for the probability that a decision based on existing evidence is wrong and for the magnitude of the consequences of making the wrong decisions. EVPI, expected value of perfect information.

framework, we established the value of collecting further information. Results show that a study collecting data on the relative effectiveness of IVIG (in comparison with standard care) appeared the most efficient research design to invest in. However, results on the value of conducting such research were also sensitive to the clinical effectiveness model used. Given that it was unclear what the clinical rationale for the effects explored was within each of the clinical effectiveness models and that the need for a further RCT exists, designing this study will be complex when uncertainties exist at this level.

Policy implications

Our study did not find evidence that current guidance on the use of IVIG should change (that is IVIG should not be recommended for use in severe sepsis, unless further evidence becomes available to support its use). Although the EVI analyses suggested substantial potential value from a large, multicentre RCT evaluating the clinical effectiveness of IVIG in this population, there remain significant uncertainties around the design of a study (with respect to, for example, the dose or duration of therapy with IVIG). Without greater understanding of the existing

Figure 5 Value of further research on IVIG for severe sepsis (and septic shock): population partial EVPI (EVPPI) for groups of uncertain parameters, according to model for the synthesis of effectiveness evidence (see Table 2 for detailed specification of the models). IVIG, intravenous immunoglobulin; EVPI, expected value of perfect information; EVPPI, partial EVPI.

variation (for example through a smaller-scale dosing RCT), there is a danger of investing in a large trial that is suboptimal (for example including a number of different dosing regimens), which may be a less efficient use of resources than deferring a definitive trial until we better understand the existing variation. Thus, our current recommendations are for research that focuses on filling the knowledge gaps that exist with a view to informing the design of a future multicentre RCT. These recommendations include: (i) research on the mechanism(s) of action of IVIG in severe sepsis (and on understanding the heterogeneity of the severe sepsis syndrome) - commencing with a rigorous review of existing research prior to embarking on any new studies and; (ii) dose-ranging/finding studies to identify dose, timing of dose and safety data to inform the intervention(s).

In addition to clarifying the results on clinical effectiveness, it may be that research informing other parameters is also worthwhile, especially with respect to data on the

Table 5 Value of further research on IVIG for severe sepsis/septic shock: ENBS and optimal sample size of a trial, according to model for the synthesis of effectiveness evidence (see Table 2 for detailed specification of the models)

Scenarios for alternative synthesis models of effectiveness evidence	Maximum ENBS	Optimal sample size per arm (*n**)
M1	£ 136,703,882	1900
M2	£ 687,441,146	1200
M3	£ 1,010,953,361	800
M4a	£ 605,931,859	900
M4b	£ 365,050,246	800

The ENBS provides the societal payoff to the proposed research. It is computed as the difference between the value of sample information (EVSI) for a particular research design and the costs of sampling. IVIG, intravenous immunoglobulin; ENBS: expected net benefit of sampling.

longer-term survival and costs of severe sepsis. This research could be conducted using relatively cheaper, non-RCT designs, for example using record linkage between existing databases or conducting a prospective cohort study providing that the period for which patients are observed is sufficiently long enough to capture the impact on costs for several years after the initial episode.

Conclusions

Our study examined the evidence for the clinical effectiveness of IVIG in severe sepsis (including septic shock) and found significant heterogeneity (variation) between studies. An in-depth look at the potential sources for the heterogeneity identified publication bias and dosing regimen as possible explanatory factors. The rationale for the latter being a relevant factor was unclear, and for this reason we examined the implications of the existing heterogeneity and concluded it affected both the potential clinical and cost-effectiveness of this treatment and, hence, current recommendations for its use.

We assessed the value of a new, large, multicentre RCT to evaluate the clinical and cost-effectiveness of IVIG in this population and, while our analyses suggest that such research may be of value, there remains significant uncertainties around the design for such a study (with respect to, for example, the dose or duration of IVIG). We concluded that, prior to investing in a new RCT, further research is needed both into the mechanism (s) of action of IVIG and from a dose-ranging/finding study.

Our work illustrates the use of an explicit framework to quantify the value of investing in new research. This work also highlights that explicit consideration of the sources of heterogeneity in the current evidence base is key to informing the design of new research.

Key messages

- The effect of IVIG on mortality is associated with a large degree of heterogeneity between individual studies. It was unclear what the relevant source of heterogeneity in the evidence base was.
- The cost-effectiveness of IVIG appears within the borderline region of estimates considered to represent value for money in the NHS, but these results are associated with significant uncertainty and appear highly sensitive to the alternative clinical effectiveness models applied.
- Collecting data on the relative effectiveness of IVIG appears an efficient research design in which to invest. However, the unclear clinical rationale for the heterogeneity in the evidence base suggests that, despite the need for a further RCT, designing this study will be complex when uncertainties exist at this level.

Abbreviations

95%CrI: 95% credible intervals; AIC: Akaike information criterion; APACHE II: Acute Physiology and Chronic Health Evaluation II; CMPD: Case Mix Programme Database; DIC: deviance information criteria; Dres: residual deviance; ENBS: expected net benefit of sampling; EVI: expected value of information; EVPI: expected value of perfect information; EVPPI: partial EVPI; EVSI: expected value of sample information; HRQoL: health-related quality of life; HTA: Health Technology Assessment; ICER: incremental cost effectiveness ratio; ICNARC: Intensive Care National Audit & Research Centre; IVIG: intravenous immunoglobulin; IVIGAM: immunoglobulin M-enriched polyclonal IVIG; LOS: length of stay; NHS: National Health Service; OR: odds ratio; PSS: Personal Social Services; QALY: quality-adjusted life years; RCT: randomised controlled trial; SC: standard care; SD: standard deviation; SE: standard error; SICS: Scottish Intensive Care Society; UK: United Kingdom.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MS conducted analyses of the data (cost-effectiveness and value of information) and drafted the manuscript. NW contributed to the acquisition of the data and its analyses (clinical effectiveness). DH analysed data (analyses of CMPD). PP contributed to the acquisition of data (cost-effectiveness and value of information). MSH contributed to the interpretation of the data and findings. SEH contributed to the interpretation of the data and findings. JM contributed to the acquisition of the data (clinical effectiveness). AEA contributed to the interpretation of data and findings (clinical effectiveness). KMR contributed to the interpretation of the data and coordinated the study. SP oversaw analyses (cost-effectiveness and value of information) and drafted the manuscript. All authors commented on drafts of the manuscript and provided final approval of the version to be published.

Author details

[1]Centre for Health Economics, University of York, Heslington, York, YO10 5DD, UK. [2]School of Social and Community Medicine, 39 Whatley Road, University of Bristol, Bristol, BS8 2PS, UK. [3]Intensive Care National Audit & Research Centre, 24 High Holborn, London, WC1V 6AZ, UK. [4]Guy's and St Thomas' Hospital NHS Foundation Trust, Westminster Bridge Road, London, SE1 7EH, UK. [5]University of Warwick, Coventry, CV4 7AL, UK.

References

1. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, Cohen J, Opal SM, Vincent JL, Ramsay G: SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med* 2003, 2001:1250–1256.
2. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Sevransky JE, Sprung CL, Douglas IS, Jaeschke R, Osborn TM, Nunnally ME, Townsend SR, Reinhart K, Kleinpell RM, Angus DC, Deutschman CS, Machado FR, Rubenfeld GD, Webb SA, Beale RJ, Vincent JL, Moreno R, Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup: Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013, 41:580–637.
3. Harrison DA, Welch CA, Eddleston JM: The epidemiology of severe sepsis in England, Wales and Northern Ireland, 1996 to 2004: secondary analysis of a high quality clinical database, the ICNARC Case Mix Programme Database. *Crit Care* 2006, 10:R42.
4. Department of Health: Demand management plan for immunoglobulin use. In *Demand management plan for immunoglobulin use.* 2nd edition; 2008. [http://www.ivig.nhs.uk/documents/Demand_Management_Plan_SECOND_EDITION.pdf].
5. Turgeon AF, Hutton B, Fergusson DA, McIntyre L, Tinmouth AA, Cameron DW, Hebert PC: Meta-analysis: intravenous immunoglobulin in critically ill adult patients with sepsis. *Ann Intern Med* 2007, 146:193–203.
6. Laupland KB, Kirkpatrick AW, Delaney A: Polyclonal intravenous immunoglobulin for the treatment of severe sepsis and septic shock in critically ill adults: a systematic review and meta-analysis. *Crit Care Med* 2007, 35:2686–2692.
7. Alejandria MM, Lansang MA, Dans LF, Mantaring Iii JB: Intravenous immunoglobulin for treating sepsis, severe sepsis and septic shock. *Cochrane Database Syst Rev* 2013, 9, CD001090.
8. Kreymann KG, de Heer G, Nierhaus A, Kluge S: Use of polyclonal immunoglobulins as adjunctive therapy for sepsis or septic shock. *Crit Care Med* 2007, 35:2677–2685.
9. Claxton K, Posnett J: An economic approach to clinical trial design and research priority-setting. *Health Econ* 1996, 5:513–524.
10. Soares M, Welton N, Harrison D, Peura P, Hari M, Harvey S, Madan J, Ades A, Palmer S, Rowan K: An evaluation of the feasibility, cost and value of information of a multicentre randomised controlled trial of intravenous immunoglobulin for sepsis (severe sepsis and septic shock): incorporating a systematic review, meta-analysis and value of information analysis. *Health Technol Assess* 2012, 16:1–186.
11. National Institute for Health and Clinical Excellence: *Guide to the methods of technology appraisal.* [http://www.nice.org.uk/article/pmg9: NICE; 2004].
12. Briggs A: Probabilistic analysis of cost-effectiveness models: statistical representation of parameter uncertainty. *Value Health* 2005, 8:1–2.
13. Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S: A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. *Health Technol Assess* 2004, 8:1–103.
14. Briggs AH, Claxton K, Sculpher MJ: *Decision modelling for health economic evaluation.* Oxford: Oxford University Press; 2006.
15. R Development Core Team: *The R project for statistical computing.* Vienna, Austria; The R Foundation for statistical computing. 2010. [http://www.r-project.org/]
16. Bernardo JM, Smith AFM: *Bayesian Theory.* Chichester, UK; New York: John Wiley & Sons Ltd; 1994.

17. Gilks WR: *Markov chain Monte Carlo.* New York, NY; John Wiley & Sons, Ltd: In Encyclopedia of Biostatistics; 2005.
18. Lunn D, Spiegelhalter D, Thomas A, Best N: The BUGS project: evolution, critique and future directions. *Stat Med* 2009, **28**:3049–3067.
19. Lunn DJ, Thomas A, Best N, Spiegelhalter D: WinBUGS - A Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000, **10**:325–337.
20. Sutton AJ, Abrams KR: Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res* 2001, **10**:277–303.
21. Spiegelhalter D, Abrams K, Myles J: *Bayesian approaches to clinical trials and health-care evaluation.* West Sussex, UK: John Wiley & Sons, Ltd; 2004.
22. Spiegelhalter DJ, Best NG, Carlin BP, Linde A: Bayesian measures of model complexity and fit. *J R Stat Soc Ser B (Stat Methodol)* 2002, **64**:583–639.
23. Cuthbertson BH, Elders A, Hall S, Taylor J, MacLennan G, Mackirdy F, Mackenzie SJ, the Scottish Critical Care Trials G, the Scottish Intensive Care Society Audit G: Mortality and quality of life in the five years after severe sepsis. *Crit Care* 2013, **17**:R70.
24. Department of Health: NHS Reference costs 2007/08. In *Book NHS Reference costs 2007/08.* Department of Health; 2008. [http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_098945]
25. Manns BJ, Lee H, Doig CJ, Johnson D, Donaldson C: An economic evaluation of activated protein C treatment for severe sepsis. *N Engl J Med* 2002, **347**:993–1000.
26. Drabinski A, Williams G, Formica C: Observational evaluation of health state utilities among a cohort of sepsis patients. *Value Health* 2001, **4**:128–129.
27. Rodriguez A, Rello J, Neira J, Maskin B, Ceraso D, Vasta L, Palizas F: Effects of high-dose of intravenous immunoglobulin and antibiotics on survival for severe sepsis undergoing surgery. *Shock* 2005, **23**:298–304.
28. Hentrich M, Fehnle K, Ostermann H, Kienast J, Cornely O, Salat C, Ubelacker R, Buchheidt D, Behre G, Hiddemann W, Schiel X: IgMA-enriched immunoglobulin in neutropenic patients with sepsis syndrome and septic shock: a randomized, controlled, multiple-center trial. *Crit Care Med* 2006, **34**:1319–1325.
29. Karatzas S, Boutzouka E, Venetsanou K, Myrianthefs P, Fildisis G, Baltopoulos G: The effects of IgM-enriched immunoglobulin preparations in patients with severe sepsis: another point of view. *Crit Care* 2002, **6**:543–544. author reply 545.
30. Tugrul S, Ozcan PE, Akinci O, Seyhun Y, Cagatay A, Cakar N, Esen F: The effects of IgM-enriched immunoglobulin preparations in patients with severe sepsis [ISRCTN28863830]. *Crit Care* 2002, **6**:357–362.
31. Behre G, Ostermann H, Schedel I, Helmerking M, Schiel X, Rothenburger M: Endotoxin concentrations and therapy with polydonal IgM-enriched immunoglobulins in neutropenic cancer patients with sepsis syndrome: pilot study and interim analysis of a randomized trial. *Anti Infect Drugs Chemother* 1995, **13**:129–134.
32. Schedel I, Dreikhausen U, Nentwig B, Hockenschnieder M, Rauthmann D, Balíkcioglu S, Coldewey R, Deicher H: Treatment of gram-negative septic shock with an immunoglobulin preparation: a prospective, randomized clinical trial. *Crit Care Med* 1991, **19**:1104–1113.
33. Wesoly C, Kipping N, Grundmann R: Immunglobulintherapie der postoperativen sepsis. *Z Exp Chir Transplant Kunstliche Organe* 1990, **23**:213–216.
34. Spannbrucker N, Munch H, Kunze R, Vogel F: Auswirkungen von immunglobulinsubstitution bei sepsis. *Intensivmedizin* 1987, **6**:314.
35. Dominioni L, Bianchi V, Imperatori A, Minoia G, Dionigi R: High-dose intravenous IgG for treatment of severe surgical infections. *Dig Surg* 1996, **13**:430–434.
36. Burns ER, Lee V, Rubinstein A: Treatment of septic thrombocytopenia with immune globulin. *J Clin Immunol* 1991, **11**:363–368.
37. De Simone C, Delogu G, Corbetta G: Intravenous immunoglobulins in association with antibiotics: a therapeutic trial in septic intensive care unit patients. *Crit Care Med* 1988, **16**:23–26.
38. Werdan K, Pilz G, Bujdoso O, Fraunberger P, Neeser G, Schmieder RE, Viell B, Marget W, Seewald M, Walger P, for the Score-Based Immunoglobulin Therapy of Sepsis (SBITS) Study Group, Stuttmann R, Speichermann N, Peckelsen C, Kurowski V, Osterhues H-H, Verner L, Neumann R, Müller-Werdan U: Score-based immunoglobulin G therapy of patients with sepsis: the SBITS study. *Crit Care Med* 2007, **35**:2693–2701.
39. Grundmann R, Hornung M: Immunoglobulin therapy in patients with endotoxemia and postoperative sepsis–a prospective randomized study. *Prog Clin Biol Res* 1988, **272**:339–349.
40. Yakut M, Cetiner S, Akin A, Tan A, Kaymakcioglu N, Simsek A: Sepsisdeki hastalarda immunglobulin G (IgG) kullaniminin mortalite oranina etkisi. *GATA Bulteni* 1998, **40**:76–81.

Paper 3.

Soares MO, Dumville JC, Welton N, Ades T. Treatment comparisons for decision making: facing the problems of sparse and few data. Journal of the Royal Statistical Society: Series A (Statistics in Society), 177(1), 2014, 259-279

Treatment comparisons for decision making: facing the problems of sparse and few data

Marta O. Soares and Jo C. Dumville

University of York, UK

and A. E. Ades and Nicky J. Welton

University of Bristol, UK

Summary. Advanced evidence synthesis techniques such as indirect or mixed treatment comparisons provide powerful analytic tools to inform decision making. In some cases, however, existing research is limited in quantity and/or existing research data are 'sparse'. We demonstrate how modelling assumptions in evidence synthesis can be explored in the face of limited and sparse data by using an example where estimates of relative treatment effects were required in a synthesis of the available evidence regarding treatments for grade 3 or 4 pressure ulcers.

Keywords: Elicited evidence; Evidence synthesis; Mixed treatment comparison; Network meta-analysis; Observational studies; Randomized controlled trials evidence; Sparse data

1. Introduction

Informed decision making on the use of health technologies often requires available evidence on alternative treatment options to be synthesized by using data from multiple sources (Sutton and Higgins, 2008). A key parameter is the relative effectiveness of the technologies of interest, evidence for which is ideally derived from randomized controlled trials (RCTs). Aggregate, or summary, data from such studies are commonly synthesized by using standard pairwise meta-analytic techniques (Higgins *et al.*, 2008; Sutton, 2000). However, when there are several different treatment options such approaches are limited, and the reader is left to compare separate meta-analysis outputs 'qualitatively' to estimate the relative effect of treatments that are not directly compared in existing RCTs. To solve this problem, extensions to meta-analysis have been developed that jointly synthesize evidence from RCTs that form a network of treatment comparisons. These are indirect or mixed treatment comparisons (MTCs) (Ades *et al.*, 2006; Lu and Ades, 2004) (also termed network meta-analysis or multiple-treatments meta-analysis).

Advanced evidence synthesis techniques such as MTCs are powerful analytic tools; yet, analyses remain dependent on the volume and the quality of the primary data included. In practice, available RCT data are often lacking, precluding the application of a single MTC analysis to all relevant treatments. Also, existing research data may be 'sparse', i.e. a small number of events

may be reported because, for example, few individuals are recruited or because of a short follow-up time. There are many challenges when analysing sparse data, one of which is that it may be difficult to justify the use of distributional approximations to statistics of interest in inferential procedures (Sweeting *et al.*, 2004). Furthermore, when no outcome events have been observed in a study, estimations may return undefined results. Such challenges may prevent the use of the available evidence in decision making, which could mean that the implementation of a potentially beneficial treatment is delayed (or the implementation of a harmful treatment is facilitated). Additionally, without an appropriate characterization of uncertainty from the existing data, research commissioners cannot appropriately assess the value of collecting further evidence (Ades *et al.*, 2006; Claxton, 2008).

Limited and/or sparse data should not, however, be seen as an insurmountable barrier to the conduct of evidence synthesis, but strong assumptions will normally be required to enable modelling and estimation. For example, in trying to form a network of evidence, analysts may group treatments, implicitly assuming that their effectiveness is equal. It is the very nature of the data that forces the use of assumptions, which makes it difficult to test these assumptions. However, MTCs offer some advantages compared with standard meta-analytic approaches as strength may be borrowed across the entire network through indirect evidence and the sharing of parameters across trials and/or treatments (Lu and Ades, 2004, 2009). Nevertheless, the process of synthesizing available evidence under sparseness of data remains challenging. On the one hand, the results must reasonably reflect the uncertainties that are shown in the data; on the other hand the analysis will be based on difficult-to-verify assumptions owing to the lack of robust data. This implies that we may not know how best to synthesize data, or that there may be other model options but no way of choosing between them.

This work uses a motivating example concerning negative pressure wound therapy (NPWT) and alternative treatments for severe pressure ulceration, where it was essential to quantify appropriately the uncertainty surrounding relative effectiveness for further use in a cost effectiveness and value of further research evaluation (Soares *et al.*, 2012). The evidence base was limited and sparse, and the work that is presented here focused on the key assumptions made to allow such data to be modelled, how consequences of assumptions could be tested (internal validity checks) and how evidence from external sources (elicited and observational data) could be used to verify these assumptions and to augment external validity. The work provides a summary of key issues and considerations, aimed at informing, shaping and promoting future evidence synthesis when data are limited and/or sparse.

2.   Motivating example: treatments for severe pressure ulceration

Pressure ulcers (or bedsores) are prevalent wounds (Kaltenthaler *et al.*, 2001) that are associated with ill health and immobility, representing a substantial burden for patients and high costs for health services (Bennett *et al.*, 2004). NPWT, which involves the application of a suction force (negative pressure) across a wound surface via a dressing, is now commonplace in the UK National Health Service and is recommended for use in the more severe pressure wounds, i.e. grade 3 or 4. Whereas grade 3 pressure ulcers involve full thickness skin loss involving damage to, or necrosis of, subcutaneous tissue that may extend down to, but not through, underlying fascia, grade 4 pressure ulcers are very deep wounds that involve extensive destruction, tissue necrosis or damage to muscle, bone or supporting structures with or without full thickness skin loss. There are several alternatives to NPWT in the treatment of pressure ulcers—most of which are wound dressings.

Relative effectiveness evidence from an MTC was required for further modelling work

inclusion in a decision analytical model (which is discussed elsewhere (Soares *et al.*, 2012)) that aimed to quantify the cost effectiveness of NPWT and comparator dressing treatments, decision uncertainty around this estimate and the value for money of conducting further primary research. For this work, relevant comparator treatments were defined in collaboration with expert clinicians as spun hydrocolloids (treatment HC), foam dressing (treatment F) and alginate dressing (treatment ALG). The effectiveness measure of interest was ulcer healing.

2.1. Description of randomized controlled trial evidence

Systematic searches were conducted to identify RCTs presenting data on the effectiveness of NPWT therapy and dressings to treat grade 3 and/or grade 4 pressure ulcers. 12 two-arm RCTs were identified (Table 1). Overall, the reported methodological quality of the trials was unclear or poor. Of the RCTs that were identified, only one evaluated the effect of NPWT (against an enzymatic treatment that is more commonly used in the USA; Ford *et al.* (2002)). Seven RCTs evaluated treatment HC, one an ALG dressing and one an F dressing (Table 1).

The follow-up periods of included RCTs were short, ranging from 4 to 16 weeks. The 12 RCTs reported the number of participants or ulcers healed at specific time points (evaluating outcomes for a total of 493 patients or wounds). It was uncommon for RCTs to report time to healing and, when they did, this estimate related only to healed individuals. Although it might be possible to incorporate these data into MTC models, we did not expect a significant contribution to the results and used the proportion of healed data only. In terms of grade of ulcer, four studies included only grade 3 ulcers, one included only grade 4 ulcers and seven included grade 3 and 4 ulcers, with two of these seven trials reporting data for grade 3 and 4 ulcers separately. We postulated that grade of ulcer may potentially impact on the rate of healing and thus investigated this further by using available data (see Section 3.1).

The MTC network formed from the RCT evidence (Fig. 1) shows that there was no direct evidence on the comparison between NPWT and each of the comparator treatments of interest.

2.2. Description of other sources of evidence

Systematic searches were conducted to identify relevant observational studies. In total three relevant observational studies were identified (Table 1). Observational evidence was not used to inform treatment effects directly, as this would generate the need to adjust for selection bias—rather, these data were used to explore for the presence of heterogeneity in baseline hazards (Section 4.2).

In addition to collecting research data, we elicited experts' beliefs over the treatment effects of interest. Judgements were elicited with uncertainty. Full methods and results of the elicitation exercise implemented have been published elsewhere (Soares *et al.*, 2011). Briefly, this exercise used formal procedures (O'Hagan, 2006) to elicit a number of quantities including the effectiveness of NPWT, F and ALG in relation to treatment HC from 23 experts. In eliciting relative effectiveness, a two-step approach was used. Firstly, each expert elicited the absolute effectiveness of HC (the proportion of patients healed in 6 months). Each expert was presented with the mode of the distribution they evaluated in the first stage (i.e. the value they believed the most) and was asked to assume that this value was true—under this assumption, in a second stage, the expert elicited the absolute effectiveness of the other treatments. Owing to time constraints these quantities were not elicited separately for grade 3 and grade 4 ulcer patients.

Judgements were elicited individually for each expert and then pooled mathematically (to generate a unique empirical distribution). A relative effectiveness measure, the (log-) hazard ratio, was calculated from the responses and a parametric distribution was fitted to the pooled distribution. The resulting subjective distributions are shown in Table 2.

Table 1. Evidence available: description of studies, treatments and healing outcomes evaluated in each arm†

s	Study	Type	Ulcer grade	Follow-up (weeks)	Treatment, k	Results for grade 3 and 4		Results for grade 3		Results for grade 4		Number grade 3, n/N
						$n_{g=3}$	$r_{g=3}$	$n_{g=1}$	$r_{g=1}$	$n_{g=2}$	$r_{g=2}$	
1	Graumlich et al. (2003)	RCT	3	4	HC, 1	—	—	7	0	—	—	—
					HG, 2	—	—	6	0	—	—	—
2	Belmin et al. (2002)	RCT	3, 4	8	HC, 1	53	8	—	—	—	—	—
					ALG-HC, 3	57	3	—	—	—	—	—
3	Kloth et al. (2002)	RCT	3, 4	12	W, 4	21	11	—	—	—	—	—
					MIX, 5	22	8	—	—	—	—	—
4	Sopata et al. (2002)	RCT	3	8	HG, 2	—	—	14	9	—	—	—
					F, 6	—	—	12	9	—	—	—
5	Honde et al. (1994)‡	RCT	3, 4	8	HC, 1	40	6	—	—	—	—	35/40
					FM, 7	29	9	—	—	—	—	24/29
6	Matzen et al. (1999)	RCT	3, 4	12	HC, 1	17	5	—	—	—	—	—
					G, 8	15	0	—	—	—	—	—
7	Burgos et al. (2000)	RCT	3	12	HC, 1	—	—	19	3	—	—	—
					ENZ, 9	—	—	18	3	—	—	—
8	Ford et al. (2002)	RCT	3, 4	6	ENZ, 9	15	2	—	—	—	—	—
					NPWT, 11	20	2	—	—	—	—	—
9	Whitney et al. (2001)	RCT	3, 4	8	W, 4	15	8	7	5	8	3	7/15
					MIX, 5	14	6	11	6	3	0	11/14
10	Muller et al. (2001)	RCT	4	16	HC, 1	—	—	—	—	12	7	—
					ENZ, 9	—	—	—	—	12	11	—
11	Price et al. (2000)	RCT	3	6	W, 4	25	3	—	—	—	—	20/25
					ALG, 10	25	2	—	—	—	—	23/25
12	Thomas et al. (2005)	RCT	3, 4	12	HC, 1	20	7	11	7	9	0	11/20
					W, 4	21	8	11	8	10	0	11/21
13	Brandeis et al. (1990)	Observational	3, 4	16	MIX, 5	718	197	369	116	349	81	369/718
14	Berlowitz and Wilking (1990)	Observational	3, 4	6	MIX, 5	37	3	22	3	15	0	22/37
15	Bergstrom et al. (2005)	Observational	3, 4	12	MIX, 5	234	12	—	—	—	—	—

†RCTs reporting the same number of ulcers and patients were assumed to have followed up only one ulcer per patient (patients may have multiple ulcers and it is often the largest or deepest eligible ulcer that is selected for follow-up). Treatment abbreviations: ALG, alginate: ENZ, enzymatic debridement agents; F, foam; FM, films and membranes; G, gauze; HC, hydrocolloid; MIX, mixed dressings: RD, redon drain; NPWT, topical negative pressure; W, warming.
‡Results on healing were derived from a time-to-event plot; data were extracted only for the subset of patients with grade 3 or 4 ulcers.
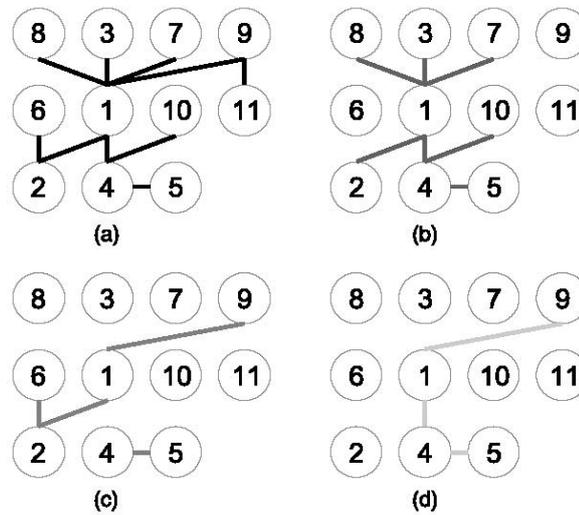
Fig. 1. Networks of evidence: in scenario (a) the network comprises all RCTs; scenarios (b) and (c) show the network of trials including only participants with grade 3 and grade 4 ulcers respectively; scenario (d) shows the network of trials including both grade 3 and grade 4 pressure ulcer patients; treatments are identified by numeric code, as in the column labelled *s* of Table 1

Table 2. Elicited distributions for relative effectiveness parameters (Soares *et al.*, 2011)

Comparison	*Elicited distribution*	
	Mean (95% credible interval)	*Distribution(mean, standard deviation)*
Log-hazard ratio of healing for F *versus* HC	−0.96 (−6.32, 4.40)	*N*(−0.96, 2.73)
Log-hazard ratio of healing for ALG *versus* HC	0.003 (−0.63, 0.64)	*N*(0.003, 0.32)
Log-hazard ratio of healing for NPWT *versus* HC	0.45 (−0.66, 1.56)	*N*(0.45, 0.57)

3. Synthesis models for the motivating example

Available RCT summary data were modelled by using MTCs (Lu and Ades, 2004), to estimate treatment effects for the four treatments of interest (F, HC, ALG and NPWT). Section 3.1 describes an initial analysis framework that used all available RCT data and considered grade of ulcer (framework A). The assumptions that were made in the implementation of the framework are summarized in Table 2. Section 3.2 presents a simplified framework where the grade of ulcer was not explicitly considered (framework B). In Section 4 we detail the implementation of the approaches that are now outlined and we describe results.

3.1. Defining the initial modelling framework (framework A)

3.1.1. Modelling framework A: incorporating heterogeneity

In synthesizing evidence from multiple trials it is important to consider differences between studies, not only in their design and conduct but also in patient populations or interventions. In the presence of such heterogeneity, true treatment effects may differ across studies even when

these studies address a common question. In general, small RCTs, like many of those in this example, have the potential to be more heterogeneous in their design, as well as less detailed in their reporting. Additionally, small sample populations run a higher risk of being unrepresentative—though poor reporting can make this difficult to assess. Although it is essential to evaluate the extent of heterogeneity (Higgins and Thompson, 2002), sparse data limit investigations. Under sparseness, uncertainty will be wide and statistically significant results are thus unlikely (Shuster *et al.*, 2007). In this example, we felt the difference in severity between grade 3 and 4 ulcers could be an important source of heterogeneity, influencing ulcer healing outcomes and the effectiveness of the alternative treatments. Thus, an initial analysis (framework A) explicitly used outcome data from grade-specific participant subgroups (Table 3, item 1).

3.1.2. Model and likelihood for the data

Formally, the data consist of J RCTs comparing K treatments. For trial j, n_{jkg} denotes the number of participants at risk at the beginning of the follow-up period, with a specific grade of ulcer g ($g = 1$, grade 3 ulcers; $g = 2$, grade 4 ulcers; $g = 3$, grade 3 and grade 4 pressure ulcers), allocated to treatment k. The number of participants from this set who became healed during the follow-up period of the study (time$_j$) is denoted by r_{jkg}. The data set can therefore be described by

$$\{(r_{jkg}, n_{jkg}), \text{ for } j = 1, \ldots, J, \ k \in t_j, \ g = \{1, 2, 3\}\}, \tag{1}$$

where t_j represents the set of treatments evaluated in trial j.

An RCT with limited follow-up can result in few outcome events within the study's timeframe. In our case-study, zero events were observed in some studies. This makes it difficult to justify the use of distributional approximations to statistics of interest (Sweeting *et al.*, 2004); exact methods were therefore required. The number of events observed at a specific time point for the gth set or subset of participants in the kth arm of RCT j, r_{jkg}, was characterized by a binomial distribution (Table 3, item 2): $r_{jkg} \sim \text{Bin}(p_{jkg}, n_{jkg})$.

3.1.3. Effect measure

In this case-study the choice of effect measure required careful consideration (Deeks, 2002; Walter, 2000). The variable duration of follow-up across the RCTs (ranging from 4 to 16 weeks) made it difficult to justify the use of measures such as odds or risk ratios. Alternatively, a hazard function relating time to healing and the duration of follow-up (Table 3, item 3) was considered. Available evidence was, however, insufficient in allowing more flexible distributions for the time to healing (Welton *et al.*, 2010) and thus the exponential distribution was used. Consequently, a defining decision was the use of a constant hazard of healing (Table 3, item 4). This assumption was not challenged in the current work but is discussed in Section 5.

For RCTs reporting outcomes on (sub)sample(s) of participants for which $g = \{1, 2\}$, the probability parameter of the binomial distribution, p_{jkg}, was assumed to inform the grade-specific log-transformed hazard of healing, θ_{jkg}, using the algebraic expression of the cumulative density function of the exponential distribution:

$$p_{jkg} = 1 - \exp\{-\exp(\theta_{jkg}) \, \text{time}_j\}. \tag{2}$$

In this way, the observed proportions of participants healed were expressed as a function of duration of follow-up.

Table 3. Assumptions of modelling framework A and how these have been challenged in further analyses†

Assumptions of model A1	Tested?	Details
Heterogeneity		
Issue 1 Ulcer grade, a factor potentially contributing to heterogeneity, was considered	Yes	The effect of ulcer grade on treatment effects was considered; this was challenged by removing such an effect, and consequently simplifying the analysis framework (framework B)
Description of the data and likelihood		
Issue 2 Binomial likelihood was assumed to describe the occurrence of healing	No	Only exact methods could be used in the presence of sparse data; also, the presence of zero counts of events in arm-specific outcomes meant that the occurrence of events had to be modelled, instead of relative measures directly; no alternative analyses were implemented
Choice of effect measure for the synthesis		
Issue 3 Hazard of healing was modelled by relating proportion of patients healed with duration of follow-up	No	Owing to differences in the follow-up duration of the trials, measures that did not account for time (such as risk or odds) were not used here; no alternative analyses were implemented
Issue 4 Time to healing was assumed to be described by a uniparametric (exponential) distribution.	No	It was not possible to relax this assumption by using available data; appropriate external evidence, if available, could have been used to specify more complex distributions (see Section 5)
Defining baseline hazard and treatment effects		
Issue 5 Baseline hazards were assumed unconstrained between studies	No	Constant or exchangeable baselines would not be appropriate in reflecting the expected heterogeneity across studies; no alternative analyses were implemented
Issue 6 Grade was allowed to determine absolute healing hazards; this effect was assumed constant across studies	Yes	Alternative parameterization of the grade-specific effect on the hazard of healing was not possible owing to the few data points available; however, observational data were used subsequently to inform this parameter (issue 9)
Issue 7 Treatment effects were assumed exchangeable (random) across dressings	Yes	Alternative models assumed unconstrained treatment effects either by using a smaller data set for which inferences were possible, or by using subjective priors (issue 9)
Issue 8 A constant effect of grade on treatment effects was explored	Yes	Alternative models assumed that this effect was zero; a simpler framework of analysis (framework B) was implemented where the effect of grade on absolute hazard of healing was embedded in the study baselines
Validity		
Issue 9 Only indirect evidence was available on the comparisons of interest; in this situation, the validity of results cannot be assessed	Yes	Two sources of external evidence were used alongside existing evidence to augment or confirm external validity: (a) elicited evidence on relative treatment effects (within framework B only); (b) observational data were used to inform a grade-specific effect on healing (baseline hazard only) Consistency of results across models (i.e. elicited + RCT *versus* RCT alone and observational + RCT *versus* RCT alone) was evaluated qualitatively

†See Sections 4.2 and 4.3 for full model details.

3.1.4. Baseline hazards and treatment effects

The log-hazard rates of healing, θ_{jkg}, were regressed linearly on indicator variables for the treatments and on an interaction term with grade of ulcer:

$$\theta_{jkg} = \underbrace{\mu_j + \beta I_{g=2}}_{\text{baseline hazard}} + \underbrace{\delta_{jk} + \gamma I_{g=2} I_{k=b(j)}}_{\text{treatment effects}}, \qquad \text{if } g = \{1, 2\}, \tag{3}$$

where $I_{g=2}$ is an indicator variable assuming the value 1 when $g = 2$ and 0 when $g = 1$, and $I_{k \neq b(j)}$ is an indicator of treatment arm (rather than control arm). The baseline log-hazard for grade 3 ulcers, μ_j, was study specific and unconstrained by any further modelling, embedding potential heterogeneity across studies (Table 3, item 5) (Spiegelhalter *et al.*, 2004). The parameter β represents an added component of hazard for participants with $g = 2$ (expected to be negative, indicating a lower hazard of healing in relation to grade 3 ulcers). This parameter was assumed constant since sparse data prevented alternative parametric specifications from being explored (Table 3, item 6).

The relative treatments effects δ_{jk} are the log-transformed effects of treatment k in relation to trial baseline treatment and can be further expressed as a function of basic parameters:

$$\delta_{jk} = d_k - d_{b(j)}, \qquad \text{for } k \neq b(j). \tag{4}$$

By assuming $d_1 = 0$, treatment HC was taken as the reference treatment. The remaining parameters d_k, $k \neq 1$, represent the log-hazard ratios of each treatment of interest in relation to the reference treatment.

In models defining treatment effects as unconstrained (with uninformative or vague priors), the data from RCTs with no reported healing events did not contribute to the likelihood (Sweeting *et al.*, 2004). In our case-study, treatment F was compared with only one other treatment (HG) (Sopata *et al.*, 2002). In turn, the study that was key in linking treatment HG (and thus F) into the network (Graumlich *et al.*, 2003) reported no healing outcome data (Table 1). For this reason we anticipated being unable to achieve inferences on treatment F contrasted with the reference treatment (HC). A solution that allows estimations is to assume that the treatment effects for the dressing treatments (i.e. all treatments in the network except NPWT) come from a population of dressing treatment effects described by a normal distribution with a common mean D—i.e. a dressing 'class effect'—and between-treatment (within-class) precision prec. This is an assumption of exchangeability across treatments within the dressings class, which we also refer to as a random-effect distribution. The parameters of the random-effect distribution, D and prec, were assigned appropriate priors. We explicitly excluded NPWT (as this treatment does not belong to the same class) and evaluated its relative effectiveness as unconstrained, i.e. as a constant c. This was parameterized as

$$d_k \begin{cases} = 0, & \text{if } t_k = \text{HC}, \\ \sim N(D, \text{prec}), & \text{if } t_k \neq \text{HC, NPWT}, \\ = c & \text{if } t_k = \text{NPWT}. \end{cases} \tag{5}$$

Grade was assumed to modify the effect of treatment through D (constant across treatments), allowing us to explore whether the relative effectiveness of treatments differed by grade of ulcer. More flexible specifications could not be implemented because of the few data points that were available (i.e. a random effect across treatments, or a treatment-specific effect (Cooper *et al.*, 2009); Table 3, item 8).

3.1.5. Outcomes reported for a mixed grade sample of patients

Five RCTs included both grade 3 and grade 4 ulcers but did not report separate outcome data

for each subgroup (identified as $g = \{3\}$). The outcomes that were observed for such mixed samples of participants were thus expressed as a function of the outcomes reported for the corresponding homogeneous (sub)groups of participants, although the latter quantities were not observed within this subset of trials:

$$p_{jk3} = [1 - \exp\{-\exp(\theta_{jk1})\,\text{time}_j\}]q_{jk} + [1 - \exp\{-\exp(\theta_{jk2})\,\text{time}_j\}](1 - q_{jk}), \qquad \text{if } g = \{3\}, \tag{6}$$

where q_{jk} represents the probability of patients having grade 3 pressure ulcers (rather than grade 4) in arm j of trial k. This probability parameter was informed by the observed number of grade 3 ulcer patients, rq_{jk} (out of nq_{jk} patients in the trial arm) by assuming that rq_{jk} was binomially distributed (with probability parameter q_{jk}). Four out of eight RCTs did not, however, report rq_{jk} (Table 1). To overcome the fact that these data were missing, an imputation procedure was implemented. A hierarchical beta–binomial model was used with q_{jk} being assumed beta distributed:

$$
\begin{aligned}
rq_{jk} &\sim \text{Bin}(q_{jk}, nq_{jk}), \\
q_{jk} &\sim \text{beta}(a, b).
\end{aligned}
\tag{7}
$$

The parameters a and b are hyperparameters to which uniform priors were assigned, $(a, b \sim \text{Unif}(0, 100))$. Note that we here assume that data are missing completely at random, supported by the rationale that the missingness is due to unintentional omission of this information in reporting.

3.2. Simplified modelling framework (framework B)

Given the uncertainty regarding the treatment modifying effect of ulcer grade we evaluated the implications of assuming the parameter γ (equation (3)) to be 0. In this case, and since we aimed only to obtain inferences on relative treatment effects, the association between grade of ulcer and baseline hazard did not need to be explicitly evaluated. Hence, a simpler data set was evaluated where studies reporting outcomes per grade of ulcer have these outcomes aggregated. The data set can be described by

$$\{(r_{jk}, n_{jk}), \text{ for } j = 1, \ldots, J, \ k \in t_j\}$$

where g is no longer represented. An exact binomial likelihood was used to represent the number of events observed at a specific point in time, $r_{jk} \sim \text{Bin}(p_{jk}, n_{jk})$. As previously, the probability parameter p_{jk} was described as a function of the log-hazard of treatment k in trial j, θ_{jk}, and further defined as a function of the treatment effects δ_{jk}

$$p_{jk} = 1 - \exp\{-\exp(\theta_{jk})\,\text{time}_j\}.$$

The model was defined with log-hazard rates for the baseline treatments τ_j kept unconstrained (i.e. heterogeneity is implicitly embedded in these baselines):

$$\theta_{jk} = \underbrace{\tau_j}_{\text{baseline hazard}} + \underbrace{\delta_{jk}}_{\text{treatment effects}}.$$

The treatment effects were assumed to be exchangeable across treatments, analogously to equations (4) and (5). Vague prior distributions were assigned to all parameters.

3.3. Challenges in implementing the synthesis models and model checks

In any evidence synthesis it is desirable for multiple alternative models to be fitted to test alternative assumptions and to select the model best representing the input data. However, such testing can be challenging in cases of limited or sparse data, where it may not be possible to identify all parameters in the more flexible models. Additionally, with sparse data it may be difficult to use statistical criteria to choose between alternative models, because model comparison tests may not be appropriate (if, for example, these are based on asymptotic approximations although nested model comparisons are more robust) (McCullagh and Nelder, 1989). The evidence in our motivating example was also limited in the sense that only direct, or indirect, evidence was available for each link and never both (there were no data loops). In this case, it meant that the validity of the consistency equations, i.e. the existence of consistency of evidence, could not be tested (Dias *et al.*, 2010).

All these issues were relevant in our case-study: we needed to ensure that the results of our analysis reasonably reflected the uncertainties in the data in a situation when we were using difficult-to-verify assumptions because of insufficient data. Likewise, there were potentially alternative ways of structuring the data together but no way of choosing between these alternative models.

In the face of the need to use the available data, we adopted two alternative procedures. One was to govern model choice by internal validity checks. These were qualitative evaluations of the influence of assumptions on estimates and their uncertainty. The other was to consider external evidence to verify assumptions made in the analysis and to augment the external validity of the conclusions drawn; in the case-study, we used elicited expert opinions and observational data.

4. Models implemented and checks

We aimed to account for the potential effect of grade of ulcer on relative treatment effect within our initial modelling framework A. The full model (A1) as defined in Section 3.1 was implemented, followed by a series of subsequent models where alternative assumptions were made (which are outlined in Table 4 and Section 4.2). Finally a second, simpler, model framework (framework B) was applied to the data, where grade was not explicitly acknowledged (outlined in Sections 3.2 and 4.3).

4.1. Methods of analysis

All analyses were conducted as follows.

4.1.1. Implementation

All analyses were undertaken in a Bayesian framework, using Markov chain Monte Carlo with Gibbs sampling implemented in the software WinBUGS 1.4.3 (Lunn *et al.*, 2009; Spiegelhalter *et al.*, 2003). The WinBUGS code that was used to analyse the data can be obtained from

```
http://www.blackwellpublishing.com/rss
```

Models were implemented defining two chains, and distinct sets of initial values were considered. Estimates that were deemed stable to the distinct sets of initial values were presented. Convergence was assessed through the Brooks–Gelman–Rubin statistic (Brooks and Gelman, 1998). All simulations that were obtained before convergence were discarded (200000 simulations for the majority of models). The simulation procedure was continued until the Monte Carlo error was invariant to running more simulations. For the majority of models, this was achieved with

Table 4. Summary of the alternative models specified in the analysis

	Modelling framework A: explicit effect of grade, for the following models:				*Modelling framework B: no explicit effect of grade, for the following models:*		
	A1†	*A2‡*	*A3§*	*A4§§ (+ observational data)*	*B1**	*B2***	*B3†† (+ elicited evidence)*
Explicit inclusion of an effect of ulcer grade	Yes	Yes	Yes	Yes	No	No	No
Baseline hazards							
unconstrained	Grade 3	Grade 3	Grade 3	Grade 3	Not grade specific	Not grade specific	Not grade specific
random							
constant	Grade 4	Grade 4	Grade 4	Grade 4 (+ observational data)			
Treatment effects							
unconstrained		Not grade specific	Not grade specific		Not grade specific	Not grade specific	Not grade specific (+ elicited evidence)
random	Grade 3			Grade 3			
constant	Grade 4			Grade 4			

†Baseline, grade 3 ulcers—unconstrained, grade 4 ulcers—constant; treatment effect, grade 3 ulcers—random, grade 4 ulcers—constant.
‡Baseline, grade 3 ulcers—unconstrained, grade 4 ulcers—constant; treatment effect, random.
§Baseline, grade 3 ulcers—unconstrained, grade 4 ulcers—constant; treatment effect, unconstrained.
§§Baseline, grade 3 ulcers—unconstrained, grade 4 ulcers—constant; treatment effect, grade 3 ulcers—random, grade 4 ulcers—constant.
*Baseline, unconstrained; treatment effect, random.
**Baseline, unconstrained; treatment effect, unconstrained.
††Baseline, unconstrained; treatment effect, unconstrained.

200 000 additional simulations (before thinning). Because of auto-correlation, we defined a thin of 20, i.e. we kept 1 out of every 20 Markov chain Monten Carlo samples.

4.1.2. Model choice

Where possible, model adequacy was assessed in terms of overall goodness of fit by using the posterior mean residual deviance—the deviance for the fitted model minus the deviance for the saturated model. In a model with good fit this statistic is expected to be approximately equal to the number of unconstrained data points (Spiegelhalter *et al.*, 2002). In our example, the number of data points varied because distinct data sets were used; the number of data points was thus presented alongside the results. Where possible, model comparison was performed by using the deviance information criterion statistic DIC (Spiegelhalter *et al.*, 2002, 2004), which is a composite measure of fit and complexity. To aid interpretation, the measure of complexity that was used to compute DIC, pD (the 'effective' number of parameters), is also presented (pD is the difference between the posterior mean of the deviance and the deviance at the posterior estimates of the parameters of interest). Residual deviance was used to compare models with different numbers of data points (by judging how close this measure was to the number of data points) as using DIC-values for model comparison here was difficult. Broadly, where the difference between a model's residual deviance was no more than the difference in data points the fit of both models was assumed comparable.

In general, given the difficulties of performing statistical comparisons between models with sparse and few data, model choice was also governed by evaluating qualitatively the influence of assumptions on model estimates (and their uncertainty). Treatments were judged more effective, and ranked first, when the hazard ratio of healing (in relation to a common comparator) was highest.

4.1.3. Further results from selected models

Further results and inferences were explored for the treatments of interest by using only the models selected, following the model choice procedures. We estimated hazard ratios of healing for all possible comparisons of treatments of interest. Owing to the pronounced skewness of the posterior distributions, the posterior median of the hazard ratios was presented with credible intervals (CIs). The cumulative probability of healing for each treatment was calculated by using the average hazard observed in the biggest RCTs evaluating treatment HC (studies 2 and 5 in Table 1) and the results of the analyses on relative treatment effects. For each model, the treatments of interest were ranked by using the posterior probabilities that each of these treatments is the most effective (Caldwell *et al.*, 2005). The treatment with higher estimated hazard ratio of healing (in relation to a common comparator) is more effective than any other treatment and was ranked first.

4.2. Models using framework A

4.2.1. Description of models

In model A1 (as defined in Section 3.1), both baseline hazard and treatment effects were determined by grade. Alternative specifications of this model were tested that challenged assumptions regarding the effect of grade (Tables 3 and 4):

$$\theta_{jkg} = \begin{cases} \mu_j + \beta I_{g=2} + \delta_{jk} + \gamma I_{g=2} I_{k=b(j)}, \\ \underbrace{\mu_j + \beta I_{g=2}}_{\text{baseline hazard}} + \underbrace{\delta_{jk}}_{\text{treatment effects}} \end{cases}$$

for models A1 and A4 (upper equation), and A2 and A3 (lower equation). Models A2 and A3 differ from model A1 in that grade is no longer a treatment effect modifier ($\gamma = 0$). Model A2 differs from model A3 in the way that the treatment effects are defined: in model A2, the d_k are assumed exchangeable across treatments within the dressings class according to the random-effects distribution that is given in equation (5), whereas in model A3 these parameters were left unconstrained (a smaller data set for which inferences were possible was used where studies 1 (evaluating treatments HC and HG, respectively treatments 1 and 2) and 4 (evaluating treatments HG and F, respectively treatments 6 and 2) were eliminated).

Model A4 used the model A1 framework to explore the inclusion of identified observational study data (studies 13–15, Table 1) to complement RCT data. Studies 13 and 14 contributed with two data points each as outcomes for grade 3 and 4 were reported separately. Study 15 reports results for a mixed grade sample, thus added as one data point. The three studies evaluated a mix of treatments used in current practice (classified as treatment MIX, Table 1). These studies did not contribute to estimates of treatment effects, but the information that is contained in them helped to define the absolute healing hazard through β (directly for studies 13 and 14, and indirectly for study 15 through equation (6)).

In all the models, parameters were given vague priors: $\mu_j \sim N(0, 10^3)$, $D \sim N(0, 10^3)$ and standard deviation sd $\sim$ unif(0, 100), where prec $=$ sd^{-2} (Lu and Ades, 2006). Treatment effects, when unconstrained, were assigned uninformative priors ($d_k \sim N(0, 10^3)$, $\forall k$, $k \neq 1$).

All the models implemented impute the proportion of grade 3 wounds for the three trials that did not report this value (see Section 3.1). We have not tested alternative specifications for the imputation procedure.

4.2.2. Results: models using framework A

Results of inferences that were obtained with models A1–A4 are shown in Table 5 along with goodness-of-fit statistics. Across all models, treatment effects were estimated with large uncertainty reflected by their wide CI.

Model A1 was the most complex model that was fitted to the data and considered grade to determine baseline hazard through β, and to modify treatment effects through γ. The hazard of healing for grade 4 ulcers was lower than for grade 3, and this result was statistically significant (β was estimated at -2.77, with CI ranging from -4.37 to -1.42). The term γ was estimated to be positive (at a value of 0.23, with CI ranging from -1.29 to 1.64) suggesting that grade 4 ulcers may benefit more from treatment, i.e. the relative effect of treatment was higher than in grade 3 ulcers. Nonetheless, high uncertainty was associated with this estimate.

Model A2, which assumed that grade did not modify treatment effects, produced estimates that were similar to those of model A1 but with lower residual deviance (32.1 *versus* 33.2) indicating a better fit. The assumption of exchangeability across treatments was evaluated by qualitatively comparing results of inferences over the treatment effects d in models A2 (random treatment effects) and model A3 (unconstrained treatment effects). The results were concordant, although in model A2 these estimates were consistently shrunken towards the common effect with mean D. The biggest difference was seen with gauze (treatment 8, G) where model A3 returned highly uncertain estimates and lower values (-26.7, CI ranging from -72.4 to -3.0) than model A2 (-1.2, CI ranging from -4.1 to 0.45); these results were consistent with trial evidence where treatment G was reported to be less effective than treatment HC in a trial where no healings were observed in the participants treated with gauze (study 6).

Estimates from model A4, evaluating the inclusion of observational data to inform β, differed substantially from those from model A1. This was because study 13 (evaluating 718 participants)

Table 5. Results of models implemented in framework A: summary statistics and goodness-of-fit-measures†

	Results for model A1‡		Results for model A2§		Results for model A3§§		Results for model A4* (+ observational data)	
	Mean	CI	Mean	CI	Mean	CI	Mean	CI
Treatment								
d_2, HG	−0.32	(−2.59, 1.75)	−0.29	(−2.44, 1.47)	—	—	0.02	(−2.48, 2.13)
d_3, ALG+HC	−0.74	(−2.11, 0.40)	−0.70	(−2.12, 0.44)	−1.26	(−2.86, 0.15)	−0.54	(−1.96, 0.64)
d_4, W	0.19	(−0.70, 1.11)	0.22	(−0.66, 1.08)	0.23	(−0.84, 1.34)	0.46	(−0.47, 1.4)
d_5, MIX	−0.48	(−1.72, 0.61)	−0.43	(−1.65, 0.61)	−0.59	(−2.00, 0.77)	0.17	(−1.02, 1.33)
d_6, F	−0.09	(−2.37, 2.08)	−0.05	(−2.18, 1.81)	—	—	0.26	(−2.23, 2.42)
d_7, FM	0.54	(−0.44, 1.66)	0.56	(−0.37, 1.60)	0.95	(−0.16, 2.12)	0.79	(−0.16, 1.85)
d_8, G	−1.30	(−4.31, 0.43)	−1.20	(−4.10, 0.45)	−26.74	(−72.42, −2.97)	−1.16	(−4.32, 0.65)
d_9, ENZ	0.43	(−0.69, 1.83)	0.53	(−0.26, 1.40)	0.78	(−0.11, 1.70)	1.28	(0.07, 2.75)
d_{10}, ALG	−0.25	(−1.97, 1.22)	−0.20	(−1.82, 1.23)	−0.34	(−2.79, 1.97)	0.13	(−1.58, 1.74)
d_{11}, NPWT	0.09	(−2.62, 2.81)	0.21	(−2.34, 2.73)	0.45	(−2.04, 2.94)	1.14	(−1.46, 3.82)
Other parameters								
β	−2.77	(−4.37, −1.42)	−2.62	(−4.07, −1.45)	−2.69	(−4.16, −1.5)	−0.52	(−0.80, −0.24)
γ	0.23	(−1.29, 1.64)	—	—	—	—	−1.05	(−2.35, 0.09)
D	−0.22	(−1.50, 0.81)	−0.17	(−1.42, 0.77)	—	—	0.16	(−1.20, 1.32)
sd	0.99	(0.13, 2.65)	0.95	(0.15, 2.47)	—	—	1.11	(0.15, 2.87)
Goodness-of-fit and other summary measures								
Number of data points	28		28		24**		33	
Residual deviance††	33.2		32.1		29.5		49.1	
pD	20.2		19.1		20.1		23.4	
DIC	117.3		115		113.4		156.9	

†Results regarding unconstrained baseline hazards μ are not shown.
‡Baseline, grade 3 ulcers—unconstrained, grade 4 ulcers—constant; treatment effect, grade 3 ulcers—random, grade 4 ulcers—constant.
§Baseline, grade 3 ulcers—unconstrained, grade 4 ulcers—constant; treatment effect, random.
§§Baseline, grade 3 ulcers—unconstrained, grade 4 ulcers—constant; treatment effect, unconstrained.
*Baseline, grade 3 ulcers—unconstrained, grade 4 ulcers—constant; treatment effect, grade 3 ulcers—random; grade 4 ulcers—constant.
**The data set used comprises 24 data points as study 1 (evaluating treatments HC and HG, respectively treatments 1 and 2) and study 4 (evaluating HG and F, respectively treatments 6 and 2) were eliminated: results on HG and F were thus not attained, and care must be taken when comparing goodness-of-fit statistics with other models.
††Residual deviance: posterior mean residual deviance.

reported a more similar healing rate for grade 3 and 4 ulcers than the RCT data, suggesting inconsistency between the sources of evidence (high residual deviance relative to the number of data points; DIC also increased substantially). This impacted on the link between NPWT and treatment ENZ, resulting in higher estimates of the relative effectiveness of NPWT.

4.3. Models using framework B

4.3.1. Description of models

Model B1 (which was described in Section 3.2) was firstly considered. We further altered how inferences were being conducted with regard to the relative effectiveness parameters (see Table 3). Firstly, unconstrained effects with vague priors were assumed (model B2). Model B3 extended model B2 by using the available subjective prior distributions for treatment 6, 10 and 11 (see Section 2.2), and vague priors for remaining treatments:

Table 6.   Results of models implemented in framework B: summary statistics and goodness-of-fit-measures†

	Results for model B1‡		*Results for model B2*§		*Results for model B3*§§ *(+ elicited evidence)*	
	Mean	*CI*	*Mean*	*CI*	*Mean*	*CI*
Treatment						
d_2, HG	−0.25	(−2.35, 1.46)	—	—	−1.26	(−2.82, 0.30)
d_3, ALG + HC	−0.64	(−2.00, 0.40)	−1.22	(−2.76, 0.11)	−1.22	(−2.80, 0.12)
d_4, W	0.11	(−0.74, 0.92)	0.11	(−0.96, 1.16)	0.14	(−0.86, 1.17)
d_5, MIX	−0.25	(−1.32, 0.70)	−0.32	(−1.58, 0.95)	−0.29	(−1.53, 0.91)
d_6, F	−0.04	(−2.04, 1.73)	—	—	−0.96	(−2.16, 0.23)
d_7, FM	0.48	(−0.37, 1.50)	0.85	(−0.22, 1.96)	0.85	(−0.19, 1.99)
d_8, G	−1.06	(−3.78, 0.41)	−26.44	(−70.31, −2.85)	−26.7	(−71.88, −2.82)
d_9, ENZ	0.48	(−0.30, 1.39)	0.79	(−0.1, 1.71)	0.79	(−0.07, 1.69)
d_{10}, ALG	−0.21	(−1.82, 1.12)	−0.41	(−2.77, 1.76)	−0.27	(−2.16, 1.55)
d_{11}, NPWT	0.18	(−2.22, 2.61)	0.45	(−1.93, 2.82)	0.47	(−1.24, 2.20)
Other parameters						
D	−0.12	(−1.31, 0.75)	—	—	—	—
sd	0.85	(0.08, 2.38)	—	—	—	—
d_{new}	−0.12	(−2.64, 2.09)	—	—	—	—
Goodness-of-fit and other summary measures						
Number of data points	24		20*		24	
Residual deviance**	24.2		20.7		19.85	
pD	17.7		18.7		17.95	
DIC	105.7		103.2		101.6	

†Results regarding unconstrained baseline hazards τ are not shown.
‡Baseline, unconstrained; treatment effect, random.
§Baseline, unconstrained; treatment effect, unconstrained.
§§Baseline, unconstrained; treatment effect, unconstrained.
*The data set used comprises 24 data points as study 1 (evaluating treatments HC and HG, respectively treatments 1 and 2) and study 4 (evaluating HG and F, respectively treatments 6 and 2) were eliminated: results on HG and F were thus not attained, and care must be taken when comparing goodness-of-fit statistics with other models.
**Residual deviance: posterior mean residual deviance.

$$d_k \sim \begin{cases} N(0, 10^3), & k = \{2, 3, 4, 5, 7, 8, 9\}, \\ N(-0.96, 2.73), & k = \{6\}, \\ N(0, 0.32), & k = \{10\}, \\ N(0.45, 0.57), & k = \{11\}. \end{cases} \qquad (8)$$

4.3.2. Results: models using framework B

The results of applying models B1–B3 are shown in Table 6. As expected, CIs over the treatment effect estimates for which a random effect was assumed were slightly less wide in model B2 than in B1 ('shrinkage').

Model B3 included elicited evidence (see Section 2.2), making inferences possible for treatments 2 and 6 when unconstrained treatment effects were used. Estimates were only slightly more confident than with model B1, reflecting the uncertainty in experts' knowledge. The only marked difference was for treatment F which was judged less effective by experts.

4.4. Interpretation of results and model selection

All models in framework A (A1–A4) consistently evaluated lower baseline healing rates in grade

4 ulcers than in grade 3. When ulcer grade was also included as a treatment effect modifier (model A1, through γ), results suggested that treatments were more effective when used in grade 4 ulcers. However, because of the wide uncertainty the relevance of this finding is unclear. Although clinically it might be expected that the grade of ulcer impacts on healing rates *per se*, there is currently no suggestion in the literature that ulcer grade modifies the effectiveness of treatments.

Observational data (included in model A4) reported more comparable healing rates for grades of ulcer than data from the RCTs. This can be seen as an inconsistency between the two sources of evidence, and combining sources of evidence in this case-study was deemed inappropriate.

Given the uncertainty over grade being a relevant treatment effect modifier, framework B was also explored. Model B1 considered a 'class' effect on the effectiveness of alternative treatments. Results obtained were rather invariant to distinct model specifications, specifically when compared with model B2 where the assumption of a random but common effect was relaxed. Investigations over the specification of the baseline hazard (random and common effects) did not return better fitting models (the results are not presented here). In model B3, the use of subjective priors affected estimates of the effectiveness of treatment F when compared with the inference that was obtained with the exchangeability assumption (model B1).

Treatment effect estimates from frameworks A and B were broadly similar (bearing in mind that estimates of the *d*s in models A1 and A4 isolate the effectiveness over grade 3 ulcers by using γ to represent the added effectiveness of treatments on grade 4 ulcers; see Table 4). For NPWT, estimates that were produced from models using similar assumptions remained stable (models A3 and B2 both assuming unconstrained treatment effects; and models A2 and B1 assuming random treatment effects without an effect modifier). However, residual deviances were closer to the number of data points in models of framework B—this may suggest that the way in which we could consider the effect of grade in framework A was not sufficiently flexible, or that the existing heterogeneity was not fully explained by grade. On this basis, we selected models B2 and B3 for further work, but given existing uncertainty we still present model A1 results. The choice between models B2 and B3 is dependent on views regarding the use of subjective priors; consequently both these models were explored further.

4.5.   Further results from selected models

Estimated hazard ratios of healing (posterior medians) with models B3 and A1 are shown in Table 7 for all permutations of treatment comparisons of interest (even if unobserved in RCTs).

Table 7.   Estimated hazard ratios

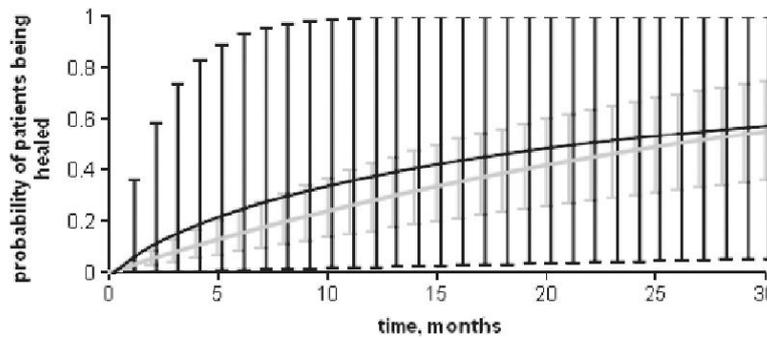| Treatment	Results for model B3 and the following treatments:				Results for model A1, grade 3 ulcers, and the following treatments:				Results for model A1, grade 4 ulcers, and the following treatments:			
	HC	F	ALG	NPWT	HC	F	ALG	NPWT	HC	F	ALG	NPWT
HC												
F	0.38				0.92				1.15			
ALG	0.76	0.82			0.78	0.85			0.98	1.07		
NPWT	1.60	1.23	1.49		1.10	1.20	1.41		1.38	1.51	1.77	

Fig. 2. Probability of patients being healed at different time points and 95% CIs, estimated using model B3:
▮, HC; ▮, NPWT

Table 8. Posterior probabilities that each of the treatments of interest is the most effective

Treatment	Posterior probability (%) for model B2	Posterior probability (%) for model B3	Posterior probability (%) for model A1
HC	16.5	16.4	20.0
F	25.2	1.8	24.5
ALG	15.1	22.3	15.0
NPWT	43.2	59.5	40.5

We also predicted the cumulative probability of healing (posterior medians) for the treatments of interest (Fig. 2 shows results for model B3). Up to 30 months, individuals who received NPWT had a higher expected probability of having their ulcers healed: these results for NPWT are, however, very uncertain as shown by the wide CIs. The uncertainty reflects the sparse data that the estimates are based on.

We then ranked treatments evaluated in relation to the probabilities (posterior) that each treatment was the most effective (Table 8). For all models, NPWT was the treatment with the highest likelihood of being the most effective (e.g. 43.2% in model B2). However, the chance that this is not the most effective treatment is also high (e.g. $56.8\% = 100 - 43.2\%$ in model B2), reflecting the existing uncertainty.

5. Discussion

5.1. Main findings

In this paper we demonstrate how modelling assumptions in evidence synthesis can be explored in the face of limited and sparse data. In this work we

(a) developed an analytical approach to allow the estimation of all treatment effects of interest (a 'class effect' was applied to dressing treatments),
(b) included external (elicited and observational) evidence and
(c) used extensive internal validity checks.

The strengths and limitations of each of these features are discussed in turn.

5.1.1. Use of a 'class effect'

In the motivating example inferences could not be obtained for one of the treatments of interest by using standard synthesis models (e.g. unconstrained treatment effects and baselines, no external data used). To overcome this we adopted a novel analytical approach: we assumed that the relative effectiveness observed for an element of a class of treatments, dressings, informed any other treatment also belonging to this class. This was accomplished by defining a random effect describing the relative effectiveness for the distinct treatments in this 'class' of dressings, but not for NPWT (its relative effectiveness parameter remained unconstrained). We felt that such a class effect could be justified since there is no published evidence that one dressing treatment is better than another in the treatment of pressure ulcers (this is also true in other complex wounds such as venous leg ulcers).

In using this approach, a conditional probability structure was defined and information transferred between treatments in the MTC while explicitly acknowledging possible differences—thus *all* dressing treatments were used to inform the estimate and not only the treatments of interest. If a standard model (e.g. standard specifications such as unconstrained baselines and treatment effects) had been employed treatments such as gauze (G) would not have contributed to inferences in the treatments of interest; yet it is important to include these treatments when exchangeable treatment effects are considered. This is an important feature of the work that is developed here and can be easily generalized to allow future MTCs to accommodate multiple classes of treatments.

The estimates obtained show that uncertainty was well represented according to the existing RCT evidence. However, the assumption of 'exchangeability' between treatments is, in our example, likely to influence the results—data on one of the treatments of interest, treatment F, was uninformative and for this reason its relative effectiveness was valued close to the mean class effect. To augment the (external) validity of the conclusions drawn we here further proposed the use of external sources of evidence. In the case-study we illustrate the use of elicited expert opinion and observational data alongside evidence from the literature.

5.1.2. Use of external evidence

In general, where there are sparse data appropriate external evidence offers the potential to challenge analytic assumptions. In this example, the available RCT data did not allow the use of distributions for time-to-event data other than the exponential distribution. If a second source of evidence had been available, a more complex distribution could have been tested. In what concerns elicited evidence, its use in this example was valuable: indeed, the judgements elicited from experts challenged analytical assumptions that were introduced in the analysis, specifically regarding the effectiveness of treatment F for which less data were available. In our example, the elicited judgements were consistent with the existing RCT evidence (and reflected the existing wide uncertainty). Key to any elicitation is to ensure that formal methods are employed and that judgements obtained are face valid and adequate (Soares *et al.*, 2011). However, even in these circumstances it is impossible to evaluate whether elicited probabilities or distributions truly translate the experts' views, i.e. the quality of the exercise.

We also explored the use of observational data to inform differences in healing by grade of ulcer. However, since we found inconsistencies with the available RCTs we decided not to use these data further. The inconsistencies may be due to the presence of heterogeneity; this could have been controlled for by conditioning inferences on covariates, or by possibly restricting the observational data to a subpopulation matching the experimental data sample in observed covariates. Future studies should consider requesting access to the individual patient data, when possible, to be able to overcome some of the limitations.

5.1.3. Use of internal validity checks

We employed a series of internal validity checks and compared alternative (and possible) model specifications. An important component of such work was exploration of the effect of the grade of ulcer. The grade of ulcer is a measure of severity of the condition that could affect healing. Also, it could also modify the relative effectiveness of treatments. In our analysis, an initial modelling framework was implemented in which baseline hazards were determined by grade and treatment effects were modified by grade (framework A). This approach fully utilized the available data from RCTs that often reported outcomes separately for subgroups of patients (according to grade). Results demonstrated that grade determines the absolute hazard of healing but showed no evidence of modifying the effectiveness of treatment—a simpler modelling approach (framework B) was implemented, where any effect of grade of ulcer on baseline hazard of healing was embedded in the study level baseline hazards. Given the sparse data, not obtaining a statistical significant result for an effect of grade on treatment effects is not surprising; thus we recommend further research to explore this issue.

Despite having conducted several analyses (which are not all shown here), further checks could have been implemented, specifically investigating how missingness in the proportion of grade 3 ulcers in the RCT samples was accounted for. Bayesian inference provides a natural framework for dealing with missing data in hierarchical models, by imputing the missing values at each step of the Markov chain Monte Carlo simulation as a draw from their predictive distribution. We specified a hierarchical beta–binomial model to describe the proportion of grade 3 ulcer patients. We could have extended the approach that was used in several ways, e.g. by exploring the influence of other sample characteristics or exploring alternative imputation procedures. In the example, we used the posterior distribution to impute; alternatively, we could have used the predictive distribution, or directly the observed proportion in studies for which this information was available.

When limited data are available it is less likely that loops of evidence exist (loops can be generated when both direct and indirect evidence exist on a particular comparison). This was so in our example. The absence of loops of evidence may reduce the complexity of analyses but, when loops exist, it is important to evaluate consistency of evidence (Dias *et al.*, 2010; Lu and Ades, 2006) to grant internal validity to inferences obtained.

5.1.4. Other strengths and limitations

Ultimately, rather than trying to choose models we could have used model averaging (Draper, 1995). Nevertheless, competing model specifications provided closely comparable explanations for the data, and consequently we deemed that there was little point in this.

Finally, we acknowledge that this work focuses on one case-study example and other strategies may be more appropriate in other cases. Nonetheless, the approach that we adopted here has generic themes: where analytic assumptions are needed to generate quantitative representations of uncertainty, internal validity checks and external evidence can be used to increase our confidence in the results.

5.1.5. Relevance of the findings and wider policy implications

The estimates that were obtained here have been used in a separate strand of work: the conduct of a decision analytical modelling and value of further information project aiming to inform the decision to adopt one of a set of treatments for grade 3 or 4 pressure ulcers and as well the need for further research (Soares *et al.*, 2012). The growing need to evaluate treatments explicitly often requires few and sparse data to be aggregated. Although there is a temptation to discard

such data, adoption and research decisions often rely on only sparse data. It is also when data are sparse that it is most important that further research needs are evaluated and prioritized. Value-of-information analyses applied to this case-study showed that more RCT information would be valuable to inform policy decisions on NPWT (Soares *et al.*, 2012). Although there is a constant need for improved conduct and reporting of primary research, the work that is reported here demonstrates approaches that aim to ensure that all available information is acknowledged and existing uncertainties correctly represented by using Bayesian MTC coupled with extensive sensitivity analysis and the use of external evidence.

Acknowledgement

References

Ades, A. E., Sculpher, M., Sutton, A., Abrams, K., Cooper, N., Welton, N. and Lu, G. (2006) Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics*, **24**, 1–19.

Belmin, J., Meaume, S., Rabus, M.-T. and Bohbot, S. (2002) Sequential treatment with calcium alginate dressings and hydrocolloid dressings accelerates pressure ulcer healing in older subjects: a multicenter randomized trial of sequential versus nonsequential treatment with hydrocolloid dressings alone. *J. Am. Geriatr. Soc.*, **50**, 269–274.

Bennett, G., Dealey, C. and Posnett, J. (2004) The cost of pressure ulcers in the UK. *Age Agng*, **33**, 230–235.

Bergstrom, N., Horn, S. D., Smout, R. J., Bender, S. A., Ferguson, M. L., Taler, G., Sauer, A. C., Sharkey, S. S. and Voss, A. C. (2005) The National Pressure Ulcer Long-Term Care Study: outcomes of pressure ulcer treatments in long-term care. *J. Am. Geriatr. Soc.*, **53**, 1721–1729.

Berlowitz, D. and Wilking, S. (1990) The short-term outcome of pressure sores. *J. Am. Geriatr. Soc.*, **38**, 748–752.

Brandeis, G. H., Morris, J. N., Nash, D. J. and Lipsitz, L. A. (1990) The epidemiology and natural history of pressure ulcers in elderly nursing home residents. *J. Am. Med. Ass.*, **269**, 2905–2909.

Brooks, S. P. and Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *J. Computnl Graph. Statist.*, **7**, 434–455.

Burgos, A., Gimenez, J., Moreno, E., Lamberto, E., Utrera, M., Urraca, E., Velez, F., Lopez, E., Martinez, M., Gomez, M. and Garcia, L. (2000) Cost, efficacy, efficiency and tolerability of collagenese ointment versus hydrocolloid occlusive dressing in the treatment of pressure ulcers. *Clin. Drug Invest.*, **19**, 357–365.

Caldwell, D. M., Ades, A. E. and Higgins, J. P. T. (2005) Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Br. Med. J.*, **331**, 897–900.

Claxton, K. (2008) Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics*, **26**, 781–798.

Cooper, N. J., Sutton, A. J., Morris, D., Ades, A. E. and Welton, N. J. (2009) Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Statist. Med.*, **28**, 1861–1881.

Deeks, J. J. (2002) Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statist. Med.*, **21**, 1575–1600.

Dias, S., Welton, N. J., Caldwell, D. M. and Ades, A. E. (2010) Checking consistency in mixed treatment comparison meta-analysis. *Statist. Med.*, **29**, 932–944.

Draper, D. (1995) Assessment and propagation of model uncertainty (with discussion). *J. R. Statist. Soc.* B, **57**, 45–97.

Ford, C. N., Reinhard, E. R., Yeh, D., Syrek, D., de las Morenas, A., Bergman, S. B., Williams, S. and Hamori, C. A. (2002) Interim analysis of a prospective, randomized trial of vacuum-assisted closure versus the healthpoint system in the management of pressure ulcers. *Ann. Plas. Surg.*, **49**, 55–61.

Graumlich, J. F., Blough, L. S., McLaughlin, R. G., Milbrandt, J. C., Calderon, C. L., Agha, S. A. and Scheibel, L. W. (2003) Healing pressure ulcers with collagen or hydrocolloid: a randomized, controlled trial. *J. Am. Geriatr. Soc.*, **51**, 147–154.

Higgins, J. P. and Thompson, S. G. (2002) Quantifying heterogeneity in a meta-analysis. *Statist. Med.*, **21**, 1539–1558.

Higgins, J. P. T., Green, S. and Cochrane Collaboration (2008) *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley–Blackwell.

Honde, C., Derks, C. and Tudor, D. (1994) Local treatment of pressure sores in the elderly: amino acid copolymer membrane versus hydrocolloid dressing. *J. Am. Geriatr. Soc.*, **42**, 1180–1183.

Kaltenthaler, E., Whitfield, M., Walters, S., Akehurst, R. and Paisley, S. (2001) UK, USA and Canada: how do their pressure ulcer prevalence and incidence data compare? *J. Wnd Care*, **10**, 530–539.

Kloth, L. C., Berman, J. E., Nett, M., Papanek, P. E. and Dumit-Minkel, S. (2002) A randomized controlled clinical trial to evaluate the effects of noncontact normothermic wound therapy on chronic full-thickness pressure ulcers. *Adv. Skin Wnd Care*, **15**, 270–276.

Lu, G. and Ades, A. E. (2004) Combination of direct and indirect evidence in mixed treatment comparisons. *Statist. Med.*, **23**, 3105–3124.

Lu, G. and Ades, A. E. (2006) Assessing evidence inconsistency in mixed treatment comparisons. *J. Am. Statist. Ass.*, **101**, 447–459.

Lu, G. and Ades, A. (2009) Modeling between-trial variance structure in mixed treatment comparisons. *Statist. Med.*, **28**, 792–805.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009) The BUGS project: evolution, critique and future directions. *Statist. Med.*, **28**, 3049–3067.

Matzen, S., Peschardt, A. and Alsbjørn, B. (1999) A new amorphous hydrocolloid for the treatment of pressure sores: a randomised controlled study. *Scand. J. Plast. Reconstr. Surg. Hand Surg.*, **33**, 13–15.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

Muller, E., van Leen, M. W. and Bergemann, R. (2001) Economic evaluation of collagenase-containing ointment and hydrocolloid dressing in the treatment of pressure ulcers. *Pharmacoeconomics*, **21**, 1209–1216.

O'Hagan, A. (2006) *Uncertain Judgements: Eliciting Experts' Probabilities*. Hoboken: Wiley.

Price, P., Bale, S., Crook, H. and Harding, K. (2000) The effect of a radiant heat dressing on pressure ulcers. *J. Wnd Care*, **9**, 201–205.

Shuster, J. J., Jones, L. S. and Salmon, D. A. (2007) Fixed vs random effects meta-analysis in rare event studies: the rosiglitazone link with myocardial infarction and cardiac death. *Statist. Med.*, **26**, 4375–4385.

Soares, M. O., Bojke, L., Dumville, J., Iglesias, C., Cullum, N. and Claxton, K. (2011) Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure wound ulceration. *Statist. Med.*, **30**, 2363–2380.

Soares, M. O., Dumville, J. C., Ashby, R. L., Iglesias, C. P., Bojke, L., Adderley, U., McGinnis, E., Stubbs, N., Torgerson, D. J., Claxton, K. and Cullum, N. (2012) Methods to assess cost-effectiveness and value of further research when data are sparse: negative-pressure wound therapy for severe pressure ulcers. *Med. Decsn Makng*, to be published.

Sopata, M., Luczak, J. and Ciupinska, M. (2002) Effect of bacteriological status on pressure ulcer healing in patients with advanced cancer. *J. Wnd Care*, **11**, 107–110.

Spiegelhalter, D., Abrams, K. and Myles, J. (2004) *Bayesian Approaches to Clinical Trials and Health-care Evaluation*. Chichester: Wiley.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.

Sutton, A. J. (2000) *Methods for Meta-analysis in Medical Research*. Chichester: Wiley.

Sutton, A. J. and Higgins, J. P. (2008) Recent developments in meta-analysis. *Statist. Med.*, **27**, 625–650.

Sweeting, M. J., Sutton, A. J. and Lambert, P. C. (2004) What to add to nothing?: use and avoidance of continuity corrections in meta-analysis of sparse data. *Statist. Med.*, **23**, 1351–1375.

Thomas, D. R., Diebold, M. R. and Eggemeyer, L. M. (2005) A controlled, randomized, comparative study of a radiant heat bandage on the healing of stage 3-4 pressure ulcers: a pilot study. *J. Am. Med. Direct. Ass.*, **6**, 46–49.

Walter, S. D. (2000) Choice of effect measure for epidemiological data. *J. Clin. Epidem.*, **53**, 931–939.

Welton, N. J., Willis, S. R. and Ades, A. E. (2010) Synthesis of survival and disease progression outcomes for health technology assessment of cancer therapies. *Res. Synth. Meth.*, **9**, 239–257.

Whitney, J. D., Salvadalena, G., Higa, L. and Mich, M. (2001) Treatment of pressure ulcers with noncontact normothermic wound therapy: healing and warming effects. *J. Wnd Ost. Contnce Nursng*, **28**, 244–252.