ANALYTIC APPROACHES FOR INFORMING RESEARCH FUNDING DECISIONS: AN EXPLORATION OF THEIR ROLE AND VALUE USING CASE STUDIES OF CANCER CLINICAL TRIALS

by

Lazaros Andronis

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Health Economics Unit
School of Health and Population Sciences
College of Medical and Dental Sciences
The University of Birmingham
Birmingham, October 2012

Abstract

Patient-level evidence obtained from clinical trials is essential in assessing the cost-effectiveness of health care technologies. Given the increasing demand for primary evidence and limited public resources for health care research, research funding organisations are routinely called to make decisions on which clinical trials to fund. Such decisions need to be informed by evidence on the likely costs and benefits of competing research programmes. Two main analytic approaches have been proposed to provide such evidence, 'payback of research' and 'value of information'.

This work applied the 'payback' and 'value of information' methodologies to case studies representing proposals for clinical trials in cancer. This application gave estimates of the value of undertaking the trials and offered an insight into the strengths, limitations and usefulness of the methods.

'Payback of research' and 'value of information' can help with different funding decisions in the context of different funding streams, they are practical to undertake and can be readily incorporated into the existing research funding processes. It is suggested that the methods should be introduced and used as part of existing deliberative processes, to provide additional assurance that limited public resources are allocated to clinical trials which are likely to result in benefits to the population.

For Jenny, Mum, Dad, Roula and Vaggelis

Acknowledgements

This thesis has benefited from the advice and help of many people and it is a pleasure to acknowledge and thank them. First, I would like to thank my supervisors Dr Pelham Barton and Professor Lucinda Billingham for their insightful guidance and continuous support throughout this project. I am also thankful for the directions and advice I received from the project's advisory group, Professors Keith Abrams, Stirling Bryan, Nicholas James, Richard McManus and Dr Stuart Peacock.

I am particularly grateful to Professor Stirling Bryan for his suggestions on various aspects of this work, as well as for the opportunity to present my work at the Centre for Clinical Evaluation and Epidemiology in Vancouver and receive feedback from a diverse team of researchers. I would also like to acknowledge and thank Professor Nicholas James and Ms Sarah Pirrie for providing access to anonymised data from the TRAPEZE phase II clinical trial.

This work has been funded by the National Institute for Health Research through a Doctoral Researcher Fellowship award. The Institute's financial support has been instrumental; apart from giving me the opportunity to carry out doctoral research, the award has made it possible for me to attend a number of conferences and training courses, which has benefited both this work and me as a researcher.

I am grateful to a number of people who have shared their knowledge and offered their help with various aspects of this study; these are Ms Karen Metcalf, Mr Andrew Stanley and Dr Daniel Ford from the NHS Pan Birmingham Cancer Network; Ms Liz Sibley and Dr Peter

Davidson from the National Institute for Health Research, and Ms Natalie Reeve from Cancer Research UK.

This work has also benefited from comments, suggestions and feedback from colleagues at the University of Birmingham, researchers at the University of British Columbia as well as from participants in conferences and seminars where parts of this work have been presented. I am particularly grateful to Cristina, Raymond, Hareth, Hema, Lavanya, Paul and Jo who proofread chapters and commented on the content of this thesis.

Last but not least, I would like to thank my parents for their continuous encouragement, and, most importantly, my partner Jenny; she has been there for me and supported me in so many ways that, somehow, it feels like this project has always been a joint effort.

Table of Contents

List of Tables

List of Tables in Appendices

List of Figures

List of Figures in Appendices

Frequently used abbreviations

CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CEAF	Cost-effectiveness acceptability frontier
CMA	Cost-minimisation analysis
CRUK	Cancer Research UK
CUA	Cost-utility analysis
EME	Efficacy and Mechanism Evaluation programme
ENBS	Expected net benefit of sampling
EVPI	Expected value of perfect information
EVPPI	Expected value of perfect information for parameters
EVSI	Expected value of sample information
EVPIM	Expected value of implementation
rEVPI	Realisable expected value of perfect information
EVP	Expected value of perfection
EVIIT	Expected value of information-implementation tradeoff
HRPC	Hormone-refractory prostate cancer
ICER	Incremental cost effectiveness ratio
INMB	Incremental net monetary benefit
MRC	Medical Research Council
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NIHR	National Institute for Health Research
NIHR HTA	National Institute for Health Research Health Technology Assessment programme
NMB	Net monetary benefit
NSCLC	Non-small cell lung cancer
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
VoI	Value of information

PART I. Background

The first part of this thesis aims to provide the background to the topic. The part consists of Chapters 1 to 3. Chapter 1 discusses the role of primary evaluative research studies—such as clinical trials—in producing evidence for health care treatment adoption decisions, summarises the main sources of funding for such research and looks into the need for priority-setting among competing research programmes.

Chapter 2 discusses main aspects of priority-setting for evaluative research, distinguishes between deliberative and analytic approaches to priority-setting, describes analytic models put forward to assist with this task and categorises these models into two overarching frameworks: 'payback of research' and 'value of information'.

These frameworks are the focus of the last chapter of Part I, Chapter 3, which describes and discusses 'payback of research' and 'value of information' with regards to their aims, rationale, main components and methods by drawing on the relevant literature.

CHAPTER 1. Introduction and project aims

This chapter seeks to introduce concepts relevant to the topic of interest, specify the research question and aims of the present work and outline the structure of this thesis.

1.1. Health care service provision and applied clinical research

The primary aim of health care systems around the world is to tackle disease and improve the health and wellbeing of the populations they serve[1]. Towards this aim, health care systems set up elaborate structures through which they make available a wide range of services, treatments and procedures—collectively known as health technologies—aiming to prevent, treat and rehabilitate[2;3].

Concerns over the fact that many commonly used technologies may be, in practice, ineffective or even harmful, have prompted the recognition that health care services provided to the population should be of proven effectiveness[4;5]. At the same time, efforts to contain rapidly growing health care expenditure[6;7] and the acknowledgment that needs are infinite but resources to fulfil them are limited, have led to the introduction of a further criterion: provided technologies must also be 'cost-effective', that is, they must represent efficient use of limited resources[8-10].

Information on health technologies' effectiveness and cost-effectiveness is nowadays used routinely to inform decisions on which treatments to adopt and provide to the population[11-14]. Such information, often called secondary evidence, is provided by health technology assessment studies, which evaluate the health and economic impact of

introducing and using a technology in clinical practice. Key input in such studies is primary, patient-level evidence generated through applied clinical research[4;15;16].

In contrast to basic research, which aims to generate general knowledge *per se*, applied clinical research is carried out to address specific questions. Such research encompasses a variety of activities, including studies on disease management, health care service delivery and evaluation of treatments and interventions[17]. Applied research that aims to generate primary, patient-level evidence on the effectiveness of health care technologies is often referred to as primary evaluative research[15;17]. With respect to their design, primary evaluative studies are either observational or experimental[4;18]. In observational studies, such as case-control and cohort studies, researchers aim to observe the course of a disease without intervening to alter its natural progression by providing a treatment, with a view to identifying possible associations between disease and exposure.

On the other hand, experimental studies are experiments where the investigators intervene to prevent or alter the natural progression of a disease by providing patients with a treatment and observing its impact. The most common form of experimental studies—and the most important type of primary evaluative research—are clinical trials[19;20].

1.2. Clinical trials

The primary aim of clinical trials is to generate and provide evidence on treatments' efficacy (i.e. whether the assessed treatment works under ideal conditions) and effectiveness (i.e. whether the treatment is expected to work in routine clinical

practice)[21]. In doing so, trials assume different forms and employ different design characteristics. Studies of a particular design—randomised controlled trials (RCTs)—have been described as *"the crown jewel"*[22[p.673]] and the *"archetypal primary research methods"*[4[p.7]] in clinical research, and they are considered as the best source of primary evidence for assessing and comparing health technologies[19;23;24]. The special status of RCTs in providing primary evidence on treatments' effectiveness warrants looking into such studies in more detail, with respect to their design characteristics and phases.

1.2.1. Design characteristics of randomised controlled trials

RCTs present a number of distinctive design characteristics. First, such studies typically involve two or more 'trial arms', that is, distinct groups of participants who receive different treatments. Typically, one (or more) group(s) are given the treatment(s) of interest, with other groups receiving the best alternative treatment, placebo, or no treatment at all[25]. A second key characteristic relates to the way in which participants are allocated to trial arms. In RCTs, this is typically done by using random allocation methods[26]. Random allocation of participants ensures that trial arms are as similar as possible with respect to their participants' characteristics, and reduces the chance of invalid comparisons due to dissimilar patient groups[20;27]. Last, RCTs differ in whether or not participants and investigators are aware of the trial arm they are assigned to, known as 'blinding'[26]. 'Blinding' for both patients and investigators is expected to alleviate possible psychological factors that may affect patients' compliance and response to treatment, and removes the possibility that patients whose trial arm is known to the investigators may be treated differently[20;28].

1.2.2. Clinical trials according to phase of drug development

A wide range of preventive and curative health care services and treatments can be investigated in clinical trials. Trials assessing pharmaceuticals are distinctive in that they typically involve phases representing specific stages in drug development. Four phases are commonly distinguished; these are summarised in Table 1.1.

Table 1.1: Phases of experimentation in clinical trials of pharmaceuticals

Trial phase	Aim
Phase I	Investigation of clinical pharmacology and toxicity
Phase II	Initial investigation of treatment efficacy
Phase III	Full-scale evaluation of treatment effectiveness
Phase IV	Post treatment evaluation and marketing surveillance

In phase I trials, the experimental drug is given to humans for the first time. The objective of such trials is to assess the drug's safety, determine the maximum tolerated dose which can be given to patients without causing serious side effects and establish the optimal dose schedule, should the assessment continues into the next phase[20;21;29]. Phase I results are scrutinised to verify if the drug meets predetermined safety thresholds and can be further evaluated in a phase II trial.

Phase II trials are aimed as the first investigation of a treatment's efficacy. When more than one drug is assessed, phase II trials also serve as a screening process, to filter out drugs that are considered ineffective or potentially unsafe. Traditionally, such trials make use of conventional non-randomised designs; nonetheless, randomised phase II trials—where participants are randomly assigned to different treatment arms—are becoming

increasingly common[20;30]. Data obtained from phase II trials inform decisions on whether the experimental drug appears effective enough to be taken forward into the next stage—a phase III trial.

Phase III trials seek to carry out the first full-scale assessment of the investigated treatment(s). Such trials aim to provide definitive evidence on a drug's effectiveness and, increasingly, they provide information on patients' use of health care resources associated with the assessed treatment(s)[31]. Phase III trials usually involve hundreds, or even thousands, of patients which are often followed up for several years. Employing great numbers of patients helps to attain high levels of statistical power and increases the capacity to detect clinically relevant differences between treatments, but, on the other hand, it increases the study's cost[32]. Given the above, RCTs are considered to be "*the most rigorous and expensive type of clinical investigation*"[21[p.3]].

Often, a further step in the drug assessment sequence involves conducting a phase IV trial. Such trials are carried out to monitor the long-term impact of treatments on morbidity and mortality, as well as to identify possible side effects. Typically phase IV trials take place after the treatment has been licensed and involve monitoring cohorts of patients which have been given the treatment over extensive periods of time[20].

1.2.3. Pragmatic clinical trials

A distinction between phase III trials is often made with regards to the purpose of the trial and the type of evidence generated from it. In this respect, phase III trials have been classified as 'explanatory' or 'pragmatic'[33;34]. As Schwartz and Lellouch[33] explain,

explanatory phase III trials aim to advance scientific knowledge and explore whether a drug is efficacious under ideal conditions. On the other hand, pragmatic phase III trials aim to determine a treatment's effectiveness by investigating whether the drug would work in routine clinical practice, and establishing the 'real world' benefits and resource use associated with it[31;35;36]. Although pragmatic trials are a preferable source of primary patient-level data on the effectiveness and use of resources associated with a treatment, evidence obtained from exploratory trials is also useful and it is often used as input in health technology assessment studies and economic evaluations[37;38].

1.3. Funding for clinical trials

The importance of evaluative research in assisting the National Health Service (NHS) to achieve its objectives has been recently highlighted in a comprehensive review of the existing research funding structures in the UK[17]. Given this, it is not surprising that there exist a number of programmes to support and fund evaluative research. In the UK, funding for such research is made available from three main sources: the public sector, charitable institutions and the pharmaceutical industry. These are discussed briefly below.

1.3.1. Public funders of evaluative research

The two main pillars of governmental, publicly funded evaluative research are the NHS National Institute for Health Research (NHS NIHR) and the Medical Research Council (MRC).

The NHS NIHR co-ordinates and manages publicly funded research on behalf of the NHS. The organisation was established in 2006 with the primary aim to invigorate publicly funded research and produce evidence to support decision-making within the NHS[39]. The organisation provides funding for evaluative research mainly through the NIHR HTA programme, which is the largest of all the NIHR programmes with a budget of over £80 million a year[40]. The NIHR HTA supports research with a view to producing evidence on the effectiveness and cost-effectiveness of technologies provided by the NHS. The programme funds primary research, mainly pragmatic clinical trials, but also evidence synthesis studies and economic evaluation studies[40;41]. Within the programme, funding is distributed through its 'researcher-led', 'commissioned' and 'themed calls' streams[41-43].

The other main pillar of publicly funded primary evaluative research in the UK is the MRC. The MRC supports research across a wide range of disease areas and various designs. Following suggestions on optimising research funding in the Cooksey report[17], the MRC has turned its focus on basic and translational research; however, it also supports primary evaluative studies, such as RCTs, through the Efficacy and Mechanism Evaluation (EME) programme. Established by the National Institute for Health Research (NIHR) and the Medical Research Council (MRC) as a coordinated strategy for clinical trials, this programme supports clinical trials with the objective of advancing scientific knowledge in order to provide the best care to the population[44]. The programme provides funding for studies that seek to evaluate new diagnostic technologies, medical devices and public health interventions[44]. In 2009, the amount EME allocated to clinical trials reached £13

million. Similarly to NIHR HTA, EME funding is provided through 'researcher-led', 'commissioned' and 'themed calls' work streams[45;46].

1.3.2. Charitable organisations

Medical charities have traditionally been an important source of funding for medical research in the UK, providing financial support for both basic and applied evaluative research in a wide range of disease areas[43]. A prominent funder of evaluative research is Cancer Research UK (CRUK). The aim of the programme is to fund and support studies on diagnostic, curative or palliative treatments aimed to improve cancer patients' survival and wellbeing[47].

CRUK funding for clinical trials is provided through a number of programmes (Feasibility Study Project Grants, New Agents Committee (NAC) trials, Phase III Clinical Trial Grants, Population Research Committee (PRC) Programme Grants and the Prospective Sample Collections Project Grants scheme), though the charity's main programme for late phase clinical trials is the Phase III Clinical Trial Grants[48]. The programme makes available funding for phase III therapeutic trials and large scale phase II trials with expected follow-up of more than two years. The programme is administered and run by the Clinical Trials Awards and Advisory Committee (CTAAC) which holds the responsibility of reviewing and prioritising research proposals[49].

1.3.3. Pharmaceutical industry

Pharmaceutical companies invest heavily on research, with a significant number of clinical trials carried out in the UK being fully or partially sponsored by the pharmaceutical industry[50]. Industry-funded clinical trials are primarily carried out for regulatory purposes, to ascertain the safety and efficacy of developed pharmaceutical compounds and support cases for reimbursement within the NHS. As the main objective of pharmaceutical companies in funding research is profit maximisation, commercially funded research falls outside the purposes of this study, although some of the methods discussed in this thesis may be, to some extent, applicable to commercial companies[51;52].

1.4. Priority-setting for evaluative research

The increasing use of effectiveness and cost-effectiveness evidence for treatment adoption decisions has led to increased demand for primary evaluative studies. At the same time, public funds available for research are limited.

Within the constraints of a finite budget for research, the decision to fund a particular research programme, such as a clinical trial, incurs an opportunity cost: the loss of benefits that would accrue from funding and conducting an alternative programme. Similarly to allocating resources between health care services, decisions are needed on how to allocate the available research budget to competing proposals for evaluative research.

Such decisions need to be made in a way that ensures that public resources—coming from the government's pot or donations—are spent efficiently. This would require a systematic and explicit assessment of the costs and benefits expected to arise from allocating resources to alternative research programmes. A series of analytic methods aiming to provide such information with a view to assisting research funding decisions have been proposed in the literature since the early 1980s. Although these methods are well-defined, their value and role in informing research priority-setting remains ambiguous.

1.5. Study aims and objectives

With this in mind, this study sets out to assess the usefulness, potential role and value of prominent analytic frameworks for research priority-setting. Specific objectives are to identify methods proposed to assist with priority-setting for primary evaluative research, and, by applying these methods to case studies, assess their performance with relation to their practicality, theoretical and methodological soundness, ability to assist with relevant decisions, and potential to fit into existing research prioritisation processes.

1.6. Thesis structure

The thesis is structured as follows. The remainder of Part I sets out the background to the research questions. In particular, Chapter 2 reviews the literature to identify approaches put forward to assist and inform research priority-setting, and discusses currently used and alternative analytic approaches. Chapter 3 describes and discusses the most

prominent analytic approaches proposed for priority-setting—'value of information' (VoI) and 'payback of research'—by drawing on relevant published literature.

The next part, Part II, focuses on the practical application of the selected analytic approaches to two case studies, which represent proposals for clinical trials aimed to provide evidence on treatment adoption-related decisions in non-small cell lung cancer (NSCLC) and hormone-refractory prostate cancer (HRPC). This practical application was undertaken through a two-stage process: the first stage required identifying and summarising the available evidence around the relevant treatment adoption decisions before a decision on whether or not to fund the trials was made, and involved carrying out searches in the literature and constructing decision analytic models. Work undertaken at this stage is reported in Chapters 4 and 5 for the NSCLC and HRPC case studies, respectively. The second stage involved applying the 'payback of research' and 'value of information' methods, to determine the expected benefits from, and the potential value of, undertaking the proposed trials. The methods involved in these analyses, as well as the obtained results, are reported in Chapters 6 and 7.

The final part of the thesis, Part III, brings the study findings together and draws conclusions. Specifically, Chapter 8 interprets and discusses the results, and assesses the approaches with respect to their methodological and theoretical strengths and weaknesses, their sensitivity to different assumptions, their practicality, as well as their potential role in research priority-setting. The last chapter of this thesis, Chapter 9, discusses the findings, draws conclusions and formulates recommendations for further research.

CHAPTER 2. Review of proposed approaches to priority-setting for research

The aim of this chapter is to identify and discuss methods and approaches to priority-setting for evaluative research. The first part of the chapter gives the methods used in conducting a literature review to identify relevant material on the topic. Information obtained from this review forms the basis for the second part of this chapter, where identified approaches are described and discussed.

2.1. Literature review methods

The focus of this literature review is on methodology work and, as such, it presents a number of distinctive characteristics in comparison to widely used Cochrane-type systematic reviews. While the latter aim to provide synthesised information about interventions' effectiveness and cost-effectiveness, methodology reviews seek to identify and discuss methods, 'schools of thought' and arguments on a specific topic or idea. Although the methods and requirements for undertaking Cochrane-type systematic reviews of health care interventions are well documented[53;54], guidelines for carrying out review of methodological literature are sparse[55]. Despite this, the present review aimed to be as systematic as possible, following, wherever applicable, methods set out in well-established guidelines for undertaking literature reviews[56].

2.1.1. Identification of relevant literature

Different approaches to identifying relevant material were used in an attempt to maximise the yield of useful articles and limit the chance of potentially relevant articles being left out. Searches in electronic bibliographic databases—typically being the main focus of searches in systematic reviews—were complemented by additional techniques such as 'citation searching', related articles searches, 'reference list scanning' and general internet searches.

Bibliographic databases are virtual, electronic libraries containing a large number of journal citations and abstracts, covering a wide range of fields. As such, they are usually the starting point for literature searches. Databases searched for the purposes of this review were MEDLINE, EMBASE, CINAHL, EconLit and the databases of the Centre for Reviews and Dissemination (DARE, NHS EED and HTA). Searches in these databases were conducted in a stepwise manner. First, pilot searches aiming to give an insight into the volume of the relevant literature were carried out using simple free-text terms. These were followed by a series of main searches using more elaborate search strategies. Strategies employed in main searches, listed in Appendix 1.A., consisted of combinations of key words, synonyms, term variants, expressions and indexing (MeSH) terms, and had a focus on sensitivity rather than specificity, in that they aimed to maximise the yield of relevant studies.

In addition to searching electronic datasets using search strategies, 'related article' searches were carried out by using a built-in search feature available in MEDLINE

(PubMed portal). Such searches involved specifying key articles and retrieving publications related to these articles, identified through an algorithm that matches words in a key article with words in other articles' titles and abstracts[57]. Further searches for relevant articles were carried out using citation searching techniques, where relevant references are sought by searching for articles that cite already identified key papers. Citation searching was carried out through the ISI Web of Knowledge platform and Google Scholar®. Searches were also carried out through scanning the reference lists of key articles as well as through general, free-text searches using the Google Scholar® engine. Further information was obtained through personal communication with employees in public and charitable research funding organisations (NIHR Health Technology Assessment, Efficacy and Mechanisms Evaluation, Cancer Research UK), official documentation, as well as from funding organisations' websites.

Searches were carried out in an iterative manner, in that, identifying an article and classifying it as a key source of evidence would trigger a new series of searches for articles cited by, and citing, this publication. Searches in databases using search strategies retrieved a total of 1285 unique articles (Appendix 1.A., Table 1.a) while additional searches (i.e. citation searching, related articles searches etc.) gave an additional 16 references (Appendix 1.A, Table 1.b).

2.1.2. Selection of studies

Contrary to systematic reviews of effectiveness studies where inclusion is guided by specific requirements on the type of a study, in a methodology review there is a wealth of

different types of articles that may provide useful input, including letters, qualitative analyses and discussion papers[53;58]. In this review, all identified articles were considered against predetermined inclusion and exclusion criteria, irrespective of their type. Criteria were intentionally broad, to ensure that potentially useful material was not missed. Articles were included if:

a. they described, discussed or proposed methods for identifying or setting priorities for evaluative research;

and were excluded if:

a. they were concerned with priority-setting in areas other than health care research, or

b. they were concerned with priority-setting for basic or translational research.

For pragmatic reasons, articles or textbooks published in languages other than English or Greek were not considered. The application of inclusion and exclusion criteria to the whole set of identified studies was conducted by one researcher (L. Andronis). In cases where inclusion or exclusion was unclear, a decision was made after discussion with the project supervisors Dr P. Barton and Professor L. Billingham.

Selection of articles was carried out through a two-stage process. The first stage aimed to exclude clearly irrelevant papers and involved applying the inclusion criteria on the basis of each identified article's title and abstract. Of the 1285 unique articles found, 1098 clearly did not meet the inclusion criteria and were excluded. The remaining 187 were

articles that met the inclusion criteria, as well as papers for which an exclusion or inclusion decision could not be made on the basis of their abstract alone. These 187 papers were deemed potentially relevant and were forwarded to the second stage. There, the full text of all of the 187 papers was obtained and assessed against the inclusion/exclusion criteria, resulting in 107 relevant papers being included in the review. Articles found through methods other than bibliographic database searches (i.e. 'citation searching', related articles searches, 'reference list scanning' and general internet searches) were considered for inclusion upon identification; such searches gave an additional 16 relevant publications. A flow chart showing the number of studies at different states of the identification and selection process is given in Appendix 1.A, Figure 1.a.

2.1.3. Quality assessment and data extraction

In systematic reviews of effectiveness and cost-effectiveness studies, quality assessment aims to ensure that evidence from studies of poor quality is given a lower weight or it is filtered out[53;59]. In such reviews, good practice dictates assessing the quality of the considered studies by using published checklists[60-63]. In the present review, no articles were excluded on the basis of agreement with a quality checklist; nonetheless, in cases where methodological flaws were identified, these are explicitly highlighted and discussed. Useful information from included studies was extracted by one reviewer (LA).

2.2. Literature review results

A total of 123 relevant articles were retrieved, the majority of which discuss the current 'state of play' regarding research funding in different countries, comment on existing and proposed institutional arrangements, and examine mechanisms and criteria for setting priorities in research. Specific approaches to priority-setting are mentioned, reviewed, discussed or proposed in 29 of the identified articles. On the basis of this literature, the remainder of this chapter describes and discusses the main aspects of processes and approaches to priority-setting for research. The chapter closes with a description of specific priority-setting approaches identified in the literature.

2.2.1. Need for priority-setting in evaluative research

With regards to health care research, the term priority-setting is typically used to describe the outcome of an exercise aimed at prioritising and selecting research programmes for funding[64-67]. The necessity of identifying and setting priorities for evaluative research is evident and well documented in the literature[65;68-71]. Programmes such as the NIHR HTA Clinical Evaluation and Trials (HTA CET), the Efficacy and Mechanisms Evaluation (EME) and the CRUK Phase III Clinical Trial Grants receive a large number of proposals for research, of which only a small proportion can be funded. In 2008, the NIHR Health Technology Assessment programme alone received 90 proposals for primary evaluative research, of which only about one-third, 28 studies, were supported[40]. Similarly, of the 75 outline proposals and 22 full proposals submitted to EME in 2009, only 12 received funding[72].

Typically, priority-setting is required for specific research proposals; however, in some organisations, priorities also need to be set across research topics. Prioritising specific proposals is relevant to 'researcher-led' (reactive) programmes, where researchers are asked to submit proposals for applied research on topics of their choice[42;46;48]. On the other hand, priorities between topics are routinely set in 'commissioned' (proactive) programmes. In such programmes, decisions are needed about which of various topics to prioritise and commission research on[40;45]. Topics are typically identified through different ways, including consultations with institutes involved in making recommendations on the use of health technologies—such as the National Institute for Health and Clinical Excellence (NICE) in the UK—recommendations from organisations involved in the provision of health care services, direct consultation with patients and patient groups, as well as through 'horizon scanning' processes[73-75].

2.2.2. Approaches to priority-setting

Once potential topics and proposals have been gathered, decisions need to be made on which of them should be funded. Such decisions require determining, explicitly or implicitly, the value of carrying out a proposed piece of evaluative research. Identifying the value of a research project is a necessary but complex task, which, in principle, requires a comparison between the benefits a research project is expected to bring about against its cost[64;65].

Within this context, prioritisation is typically 'absolute', in the sense that topics or proposals are judged on their own, rather than 'relative' where all considered proposals

or topics are compared against each other and ranked[55;76-78][personal communication with Ms N. Reeve, research funding manager, Cancer Research UK, 31-03-2011]. Strictly, the distinction between absolute and relative prioritisation can be distorted by the fact that, when making absolute prioritisation decisions, it is possible that decision-makers may implicitly compare the proposal of interest against other proposals competing for funding.

Two main approaches to determining the value of research in order to assist decisions around research funding allocation are discussed in the literature:

a. 'Deliberative' approaches, which typically involve groups of experts and informed participants discussing and agreeing on the desirability of research in a deliberative and interpretive way[64;66;71;79], and

b. 'Analytic' approaches, which aim to calculate and provide numerical estimates of the expected benefits of research, typically by using statistical and mathematical techniques[64;79].

Deliberative and analytic approaches are discussed below, with regards to their main components and characteristics.

2.2.3. Deliberative approaches

Deliberative (or 'subjective'[80], 'interpretive'[79], 'implicit'[64]) approaches are used extensively by public research funding organisations and medical research charities, including the NIHR HTA, MRC and CRUK[40;44;49;81;82]. Depending on the funding stream,

such approaches aim to assist in setting priorities between possible topics for research or specific research proposals. In such a process, experts seek to weigh the potential benefits of proposed research studies against their cost, with a view to agreeing on which of them should be funded [personal communication with Dr P. Davidson, Director of NIHR Evaluation Trials and Studies Coordinating Centre, 13-01-2012].

Although deliberative approaches are broadly similar, different arrangements have been proposed and are in place in different research funding organisations. Specific components depend on a number of factors, including the goals of the programme for which priorities need to be set, the type of research that the organisation focuses on and the allocation of responsibilities between this and other organisations which fund or undertake similar research[65]. The main characteristics of deliberative approaches are described below.

2.2.3.A. Synthesis of decision-making panels and committees

Panels deciding on research funding typically consist of representatives of groups that have an expertise in the topic, or a direct interest in the results of the funding process. Typically, these are decision-makers, researchers, health care professionals and representatives of research funding bodies[40;64;68;82]. There is an increasing argument for greater public involvement in the process as the public are potential recipients of the evaluated health care interventions. Indeed, research in Canada and the UK has highlighted that the public's insights may assist in identifying important research questions[83-85]. The involvement of an interdisciplinary team including researchers,

decision-makers at local and central level, representatives of medical associations and patients has been also seen as a way to raise awareness of the value of the research programme and is expected increase the likelihood that research findings will be implemented[65;79].

2.2.3.B. Use of criteria

More often than not, research proposals are discussed and assessed with reference to predetermined criteria[41;76;78]. Using criteria helps to ensure that research topics and proposals are judged against considerations which are relevant to the aims of funding organisations. Moreover, the use of criteria is seen as a means of ensuring that the priority-setting process is not reduced to a solely implicit judgement process and enables those tasked with making decisions to look at the proposals from a common viewpoint[64;79].

According to the identified literature, criteria should facilitate a comprehensive assessment of the proposal's value, agree with the aims of the health care system and, at the same time, be relevant to the organisation which carries out the prioritisation exercise[65;86-88]. Criteria typically vary with respect to their intended use in the priority-setting process. For decisions on prioritising topics for research, relevant criteria are focused on the importance of the topic and include considerations around:

a. the burden of the disease or problem associated with the topic;

b. the topic's relevance to patients and the health care system, and

c. the extent of the existing uncertainty around the topic.

On the other hand, at the stage where specific research proposals are assessed, criteria tend to concentrate on the proposals' merits, by looking at:

a. the scientific rigours of the proposal, including considerations about the robustness of the research plan and the feasibility of recruiting the necessary number of participants, and

b. the ability of the team which submit the proposal to complete the proposed piece of research.

Criteria used by main funding organisations with respect to relevant stages are given in Table 2.1. It is worth reiterating that these criteria are aimed at priority-setting for applied evaluative research, rather than for basic or translational research. Moreover, identifying the value of a research proposal requires judging the benefits expected from conducting research, rather than the benefits from the use of a technology under assessment. Although it is sensible to assume that a relationship exists between the benefits that may result from evaluative research on a technology and the benefits from the technology itself, this relationship is typically not straightforward[65].

Table 2.1: Criteria for priority-setting used by selected research funding organisations

Funding programme	Criteria*	
	Topic-related criteria	**Proposal-related criteria**
NIHR HTA[41;78]	Burden of the health problem Cost of the health problem Current uncertainty Urgency for assessment Cost of the treatment	Scientific quality of the proposal Demonstration of the necessary skill mix, experience, project management and infrastructure for success Explanation and justification for estimated recruitment rates in primary research Ethical, legal and social implications of the research proposed have been considered Reasonable costs (value for money)
EME[89]	Importance of the topic	Scientific quality of the proposal Feasibility of the study Reasonable costs and value for money
CRUK Phase III Clinical Trials Grant[48]	Clinical importance of the research question Scientific importance of the research question Adequacy of background and preliminary data	Strength of study design, including statistical design Expected interest/appeal to patients and likelihood of adequate accrual Anticipated opening for trial in portfolio
* Reproduced from the NIHR HTA, EME and CRUK websites		

2.2.4. Analytic approaches

The view that priority-setting for evaluative research would benefit from methods that generate explicit and quantifiable estimates of the potential value of research has driven the development of a number of analytic models[64;90-92]. Such models acknowledge that the primary objective of evaluative research is to inform treatment adoption decisions and, on this basis, they aim to identify the potential benefits of conducting such research.

The focus of these approaches is on evaluative research for existing treatments which have already been proven safe and potentially effective, rather than for trials carried out to assess treatments' safety for regulatory purposes.

Both prospective and retrospective analytic methods exist in the literature[55;71]. Retrospective methods estimate the benefits from research that has already been conducted and given results, and may be useful in assessing the impact of research funding[93]. On the other hand, prospective methods aim to infer the value of a research proposal by estimating the benefits that the proposal is likely to bring about before the study is considered for funding, with a view to guiding funding decisions. Given the above, the focus of this thesis is on prospective approaches.

The literature search retrieved nine distinct prospective analytic models[67;90-92;94-98]. Three studies reporting literature searches for analytic approaches[55;67;80] were also found. The available literature, including identified models and existing work on analytic approaches is described and discussed below.

2.2.4.A. Weinstein's model for priorities in cancer prevention

In the early 1980s, Weinstein[94] developed and published an analytic model aimed to guide priority-setting for research. The model's specific objective was to determine which of two research programmes for identifying preventable causes of cancer should take priority: toxicological studies to identify possible carcinogenetics and limit human exposure to them, or epidemiological studies to improve the understanding of the impact of dietary agents on inhibiting cancer (a clinical trial on β-carotene). To answer this

question, the author[94] set out to estimate ratios representing the total costs and the benefits associated with each of the proposed research programmes, and rank these ratios in order of magnitude.

In this work, total cost was estimated as the sum of the costs of carrying out the study, although the cost of implementing the findings was not considered. As the aim of the studies in question was to prevent cancer mortality, benefits were measured in terms of avoided mortality under the assumptions that: a) carrying out each of the research programmes would establish the impact of carcinogenetics or dietary agents on cancer; b) research results would trigger changes in policy (e.g. industry regulations to reduce exposure to carcinogenetics, increased uptake of cancer-inhibiting dietary agents etc.) and c) changes in policy would have a beneficial impact on cancer prevention.

In this model, Weinstein[94] specified the magnitude of the possible effect that might be observed from each research programme, assumed a change in policy associated with observing those effect, and predicted the reduction in deaths due to the combination of observed effect and changes in policy. Separate analyses were carried out for each of the competing research programmes. The author concluded that conducting a trial to assess the benefits of β-carotene in preventing cancer is preferable to research on carconogenetics as it "*offers a greater expected reduction in cancer mortality per research dollar*"[94[p.17]].

Weinstein's[94] work is intuitive and follows a sequence of logical steps with regards to how research translates into benefits in the population. However, the model has limitations,

the most important of which relates to the fact that the occurrence of the events of interest, i.e. observation of study results, change in policy and health benefits accruing from this change, cannot be known in advance. Rather, these aspects must be treated as random variables, the occurrence of which depends on factors such as the prior probability of observing an effect (e.g. the probability that carcinogenetics trigger cancer), the "*sensitivity of the test system*"[94[p.18]](i.e. the power of the research programme to detect a difference), the magnitude of change in policy in the light of research results, and the rate at which this change will translate into health benefits. As Weinstein[94] admits, ambiguity around these aspects introduces uncertainty in the analysis and makes reliance to subjective guesses inevitable.

Although relatively simplistic, this approach has been valuable in that it laid the foundations for a series of subsequent analytic models[67]. The influence of Weinstein's early analytic model can been seen in other models, such as those by Eddy[90], Detsky[91;99], Drummond *et al*.[100], Townsend and Buxton[92], Davies *et al.*[98] and Townsend and colleagues[67].

2.2.4.B. Eddy's Technology Appraisal Priority-Setting System

Six years after Weinstein's work, Eddy[90] introduced his Technology Appraisal Priority-Setting System (TAPSS). This was developed in the context of work undertaken for the Institute of Medicine in the United States and its primary aim was to estimate the value of proposals for assessment in order to guide the Institute's research funding agenda.

Similarly to Weinstein's work[94], the TAPSS model aims to predict the benefits which would accrue had a research project proposed to investigate specific technologies been carried out and provided information to guide the use of the technologies in clinical practice. Given this, if the research project is expected to lead to beneficial changes in clinical practice and additional benefits to the population, the study should be considered as a good candidate for funding. The model is partitioned into two main stages. The first stage predicts the possible effect of the research project on a single patient, while the second stage extrapolates this effect to the population.

The starting point in Eddy's[90] work is the specification of results one would expect to see from research, which were termed 'Delta results'. Delta results, according to Eddy[90], must demonstrate some necessary features: they must a) be meaningful to practitioners, b) be mutually exclusive and collectively exhaustive, and c) must reflect the effect of the technology on relevant health outcomes. The approach makes the assumption that an assessment could result in one of two Delta results: a technology can be either effective or not effective.

Following this, the model requires asking experts to predict the probability of observing each of the specified Delta results, as well as to forecast how clinical practice would change in the light of each result. The next step involves estimating the effect that each particular Delta result would be expected to have on an individual patient. For a typical individual and given a specific Delta result, this can be approximated by the difference between the benefits (e.g. survival improvement) expected to accrue to a patient who

receives an assessed technology and the benefits realised by a patient who does not receive this technology.

Once the benefits of an assessment given a specific Delta result have been estimated, these can be used as a starting point for estimating the benefits likely to accrue to the whole population. This is found by multiplying the individual patient's benefits by the number of people expected to be in need of the technology within a specified time period over which the results of the assessment are expected to be usable. Last, weighted, 'expected' benefits are estimated as the sum of the products of the benefit of the assessment given a specific Delta result, multiplied (weighted) by the probability that the assessment will show the specific result. On the basis of the calculated benefits, assessments can be ranked and prioritised in order of desirability[90].

To illustrate the principles of his work, Eddy[90] used the TAPSS model to determine if an assessment of the cost-effectiveness of screening for maple syrup urine disease in newborn babies should take place. The study concluded that the predicted expected benefits from the study (0.17 severe retardations avoided per year) would justify devoting resources to it.

2.2.4.C. Detsky's model on the relationship between trial design, and costs and benefits of research

A similar approach to TAPSS[90] has been proposed by Detsky[91;99], who developed a conceptual framework that relates the value of evaluative research to specific trial research design characteristics. The model makes a link between the cost of a study and

the possible benefits that might result from undertaking the study (in terms of health outcomes such as reduction in the risk of death), the latter expressed in the form of a frequency distribution.

In this model, the distribution of all possible health benefits is used in place of Eddy's[90] 'Delta results'. In addition, Detsky[91;99] goes a step further by accounting for the fact that the ability of an assessment to detect a real difference of a given size as significant (i.e. the power of a trial) is related to the number of patients taking part in the trial. On the one hand, larger sample sizes allow detecting a smaller minimum clinically important difference δ, increase the chance of an effective intervention being detected as such, and are more likely to have an impact on clinical practice. On the other hand the larger the sample size, the greater the cost of the trial. On this basis, the model compares the cost required for obtaining more accurate information (i.e. the cost of increasing the trial's sample size) against the potential benefits expected to arise from changing clinical practice according to the trial results. The comparison is expressed as a cost-effectiveness ratio.

The application of the approach to seven randomised controlled trials showed that conducting each of these trials would be cost-effective, showing results which spanned from $2 to a maximum of $685 per life-year saved[91]. As expected, when the author changed the chosen minimum clinically important difference to a smaller value (i.e. when a larger trial was needed), the cost-effectiveness ratios increased substantially.

Detsky's[101] idea on the impact of trial design characteristics on the costs and benefits of research was subsequently taken forward by Coyle *et al.*[102], who looked into the link between different values of the minimal clinically important difference δ and 'time-to-payback', representing the number of years until the benefits of a trial cover its cost. The work was based on a probabilistic model which aimed to predict the costs and survival of patients with colorectal cancer in two situations: with the trial being conducted and being followed by a change in clinical practice, as well as without the trial and no change in practice. The results confirmed that the 'time-to-payback' of an effectiveness trial varied with the selected level of statistical significance and the number of trial participants recruited. The authors found that, for a trial comparing minimal versus intensive follow up for patients with colorectal cancer involving 1000 participants, a change in the significance level from 5 percent to 1 percent led to a six-fold increase in the expected 'time-to-payback' from 7 to 42 years.

In summary, the proposed models[91;99;102] expands on the methods found in previous work[90;94]. By considering how trial design characteristics impact on cost and benefits—the latter estimated as the benefits expected to arise from a beneficial change in clinical practice—this work is a useful addition to the existing literature, one that links the value of a trial with design choices.

2.2.4.D. Retrospective analysis of the costs and benefits from a trial on diabetic retinopathy

Drummond *et al*.[100] set out to develop a methodology for assessing the societal costs and benefits resulting from medical research. The purpose of the model was to find out

whether the benefits arising from the Diabetic Retinopathy Study, a major and costly clinical trial supported by the National Eye Institute in the US, outweighs the cost of the trial. Although this model has been used after the trial was carried out, Drummond *et al.*[100] point out that it can also be used prospectively to inform research funding decisions.

The authors structured their model in line with decision analytic methods and followed the core steps outlined in Eddy's TAPSS work[90]. In doing so, they specified two possible Delta results, 'positive' and 'negative', and structured their analysis in the form of a decision model (decision tree) assessing two states of the world: with the trial and without it. For each of these states, the decision tree showed the sequence of possible events, accumulated costs and benefits expected to accrue from a combination of a given result and its impact on clinical practice. Probabilities showing the likelihood of a patient following each of the different pathways were obtained from the actual trial. Using a societal perspective, the results of the study revealed that undertaking research would result in £27,800 per additional QALY at most.

The authors[100] recognised that predicting relevant parameters—such as the scale of change in clinical practice following the trial—is challenging. To address this, they used estimates from different sources including expert opinion, and tested different assumptions in the model in a series of sensitivity analyses. Drummond and colleagues' model[100] combines principles found in earlier models[90;94] with methods commonly used in health economic evaluation. Although it does not avoid limitations seen in TAPSS-based models, most notably the need to quantify the impact of trial results on clinical

practice, it improves on existing methods by introducing and employing commonly-used decision analytic structures.

2.2.4.E. The welfare loss model

An alternative model, based on different underlying principles, was proposed by Phelps and Parente[95] and was subsequently updated by Phelps and Mooney[103]. The model is underpinned by the notion that variation in a health technology's rate of use across different regions within the same country not explained by socioeconomic factors can be attributed to disagreement about the technology's true efficacy. Owing to such disagreement, current use of a technology diverges from an optimal rate of use and this diversion results in a loss of welfare. Medical disagreement and welfare loss can be alleviated by evaluative research and technology assessment, as such studies can reveal a technology's true efficacy.

To illustrate their model, Phelps and colleagues[95;103] estimated the welfare loss associated with the current use of a technology by using the following index:

$$Welfare\ loss = 0.5 \times \sum_{i=1} m_{ij}\,(CoV_j)^2\ \times\ e_j$$

Here, m_{ij} represents the expenditure on technology i in region *j,* CoV_j is the coefficient of variations, in other words standard deviation of the current use of the technology divided by the average (assumed to be the optimal) use of the technology across regions, and e_j is the change in the marginal welfare gain or loss as the use of the technology changes.

The index can be seen as a representation of the expected gain (expressed in monetary units) from carrying out research to alleviate welfare loss.

The model was used to derive research priorities for research in the New York state[95]. The authors derived estimates of variations in the use of clinical interventions using Diagnostic Related Group in the state's hospitals and found that priority should be given to evaluative research on health technologies tackling psychosis, chronic obstructory pulmonary disease and circulatory diseases, as these are areas of considerable variation in clinical practice and, consequently, large loss of benefits[95;103].

The proposed model provides an effective way of ranking candidates for assessments according to a measure of the 'social need' for research. Although the model can be easily operationalised by using information from hospital admissions across regions, there are limitations associated with it, most importantly the assumption that the appropriate level of a technology's use can be approximated by the average rate across different regions[55] and the fact that the study does not account for variations in use between hospitals within the same region.

2.2.4.F. Townsend and Buxton's prospective evaluation of a trial's payback

Townsend and Buxton[92] developed a model for *ex ante* estimation of the potential returns of carrying out evaluative research and used it to estimate the possible implications of funding a large and costly clinical trial. Contrary to earlier retrospective work aimed to quantify the benefits from research and development spending[104;105], Townsend and Buxton's model was specifically tailored to assess the returns to research

prospectively, with a view to establishing whether a trial would be considered a good candidate for funding[92].

The model builds on ideas found in the work of Weinstein[94;94] and Eddy[90]. In particular, it is based on the premise that carrying out research to assess the effectiveness and cost-effectiveness of treatments is beneficial as it guides informed decisions about the use of technologies in clinical practice. Informed decisions bring about additional benefits to the population, in the form of health gains due to increased use of cost-effective treatments and reductions in the use of non-cost-effective treatments. These additional benefits are seen as a proxy for the value of conducting research, and can be measured as the difference between the benefits that would accrue if the proposed assessment took place and its results were used to inform clinical practice ('factual state'), and the benefits that would be experienced in the absence of research ('counterfactual state').

Carrying out a prospective assessment presents a main difficulty: the results of research and, consequently, the change in clinical practice following the results, cannot be known in advance. However, this is dealt with by specifying a number of possible research results, similar to Eddy's Delta results[90], termed "*exemplar outcomes*"[92[p.184]]. A wide variety of possible research results can be specified; however, for simplicity, Townsend and Buxton[92] defined three broad outcomes: 'positive', 'negative' and 'inconclusive'. On the premise that observing an exemplar outcome will trigger a change in policy, the authors[92] specified a 'positive' outcome to lead to wider adoption and provision of the technology in clinical practice, while 'inconclusive' and 'negative' outcomes would result in no change and decrease in the use of the technology, respectively.

Carrying out research, observing an exemplar outcome and changing practice accordingly, the authors[92] argue, brings about a stream of costs and benefits. For a given exemplar outcome, costs and benefits of the 'factual', (i.e. with research) state can be compared against the cost and benefits associated with the 'counterfactual' (i.e. without research) state. For each combination, the overall value of a study is given by the following formula:

$$CE_i = \frac{C_T + C_{i,pc} - C_{i,npc}}{B_{i,pc} - B_{i,npc}}$$

where CE_i is the 'cost-effectiveness' of the proposal associated with exemplar outcome i (i.e. 'positive', 'negative' or 'inconclusive'), C_T is the discounted costs of the trial *per se*, $C_{i,pc}$ is the discounted costs associated with policy change in the light of outcome i, $C_{i,npc}$ is the discounted costs associated with no change in policy, $B_{i,pc}$ represents the discounted benefits associated with policy change in the light of outcome i, and $B_{i,npc}$ is the discounted benefits associated with no change in policy.

To account for the fact that research results are not known in advance, each exemplar outcome was assigned a weight representing the likelihood of the particular outcome being observed. This allowed forming different combinations of outcomes, termed 'scenarios'. Values obtained for different scenarios gave a range within which the expected benefits of the study—expressed in terms of cost per unit of health benefit—is anticipated to fall.

Townsend and Buxton's[92] work was applied to a proposed trial of hormone replacement therapy (HRT), a promising treatment for alleviating menopausal symptoms in women which, as a downside, may be associated with greater risk of breast and endometrial cancer. A proposal for a trial was submitted for funding to the Medical Research Council (MRC). As the trial was expected to cost approximately £47 million, the MRC was interested in finding out whether this trial would be good use of the available resources.

Following the sequence of steps described above, Townsend and Buxton[92] estimated that conducting research and changing clinical practice according to its results would result in £240 (£1150 discounted) per additional QALY for an 'optimistic' scenario (i.e. greater weight given to 'positive' outcome), £160 (£770 discounted) per QALY for a 'neutral' scenario (i.e. equal weight to all scenarios) and £55 (£260 discounted) per QALY for a somewhat 'pessimistic' scenario (i.e. higher weight to 'negative' results).

Townsend and Buxton's model appears more comprehensive and refined than earlier models[90;94]. Nonetheless, similarly to previous work[90], the model requires a large amount of empirical data in order to be operationalised. As the authors[92] acknowledge, such data may not be readily available for many technologies, in which case empirical evidence will have to be substituted by expert opinion. Townsend and Buxton's model represents a useful addition to the existing analytic approaches, one that appears to strike a balance between practicality and comprehensiveness.

2.2.4.G. 'Value of information' analysis to inform research funding decisions

In the mid-1990s, Claxton and Posnett[97] proposed 'value of information' (VoI) as an alternative tool for assessing the need for, and the value of, conducting research. The framework follows the principles of statistical decision theory[106-108] and is applicable to situations where choices need to be made between alternative courses of action which are expected to result in uncertain payoffs.

The approach is directly applicable to treatment coverage and research funding decisions in health care. As Claxton[97] explains, decisions on treatment adoption are typically uncertain as the relevant payoff (here, health benefits accruing to the population) from adopting a treatment over another is not know with certainty. However, decisions are still needed and it has been argued that a treatment should be provided if, based on available evidence, it appears superior to its alternatives[97;109]. However, making decisions under uncertainty (or, put differently, under imperfect information) is expected to result in an opportunity loss: the loss of benefits due to adopting a decision which may turn out to be suboptimal and depriving the population of a more beneficial treatment. VoI analysis can be conducted to assess the expected loss of benefits due to making a decision under uncertainty and establish the expected value of acquiring better information through research, in order to minimise this loss[97].

Results of VoI analysis can be used as a criterion to establish whether conducting further research would be worthwhile[110], thus the approach has become prominent as a means of assessing the need for further research[111-114], with a number of economic evaluation

studies being complimented by VoI. The rationale, methods and concepts in VoI analysis are explained in more detail in Chapter 3.

2.2.4.H. Davies and colleagues' prioritisation model for the NHS HTA

Davies and colleagues[98] proposed their Economic Prioritisation Model to assist the NIHR Coordinating Centre for Health Technology Assessment (NCCHTA)(now NIHR HTA) select and prioritise health technology assessment studies. The model aimed to inform decisions at the stage of shortlisting candidates for assessment on the basis of evidence from 'vignettes', that is, summaries of available evidence on a specific topic. In addition to providing useful information to decision-makers, this work aimed to identify critical factors which determine the extent of the expected costs and benefits that might result from specific assessments.

The model is based on the principles introduced by Eddy[90] and builds upon earlier work by Drummond *et al.*[100] on the retrospective quantification of the costs and benefits expected to result from an assessment. The analysis involves a series of steps, including specification of possible Delta results and probabilities of the assessment showing a specific result, specification of the likely change in clinical practice in terms of future utilisation of the technology under assessment, and estimation of the costs and benefits for each Delta result. The extent of future utilisation, Davies *et al.*[98] point out, is affected by a number of factors, the most important of which are whether the assessment is exploratory (i.e. primary assessments aimed at answering questions not previously dealt with in rigorous studies) or confirmatory, (i.e. assessments for which a reasonably high

volume of evidence exists), the effectiveness of the result dissemination, the likelihood of new, competing technologies emerging, and the expected time horizon over which the technology under assessment is expected to be used.

The model[98] was used to estimate the expected costs and benefits of a series of assessments commissioned by the Pharmaceutical panel of the NCCHTA in 1997 and 1998. Analysis could not be undertaken for 20 percent and 40 percent of the assessments considered in 1997 and 1998, respectively, due to lack of complete clarity on the topic, interventions or patient groups of interest. However, for topics for which analyses could be conducted the available data were sufficient to give base case estimates of expected costs and benefits.

The analysis underlined the difficulty in making safe conclusions about the potential costs and benefits of an assessment; for approximately 58 percent of the assessments considered in 1997 and 73 percent of the assessments considered in 1998, the estimated results from the model were highly uncertain, spanning from net savings to net costs. This work[98] also showed that key determinants of the results were the additional cost of the assessed technology as compared to usual care, the likely levels of utilisation and the probability of the assessment finding a technology to be effective. The authors'[98] conclusions follow from the result of the study: at the stage where topics are considered for prioritisation, predicting the absolute values of the costs and benefits associated with the assessment is challenging. However, obtaining an estimate of the potential costs and benefits by carrying out an *ex ante* assessments is still plausible and can be of value to decision-makers.

2.2.4.I. The Preliminary Assessment of Technology for Health Services model

In an attempt to address limitations found in earlier models[90;92;97;100;101], Townsend and colleagues[67] developed and published their Preliminary Assessment of Technology for Health Services (PATHS) work. The PATHS model aimed to predict the additional benefits expected from conducting research on a specific topic, with a view to assisting research funding decisions. According to the authors[67], the model can be used at two levels: at the 'topic' level, to make recommendations on whether a specific technology or topic is worth evaluating, or at the 'research proposal level', to help with funding decisions around specific study proposals.

The model appears to expand and refine Townsend and Buxton's earlier work[92]. In brief, the model involves specifying alternative results that might be revealed in a clinical trial ('favourable', 'unfavourable' and 'inconclusive' for the treatments under assessment) and, for each of these results, it estimates the costs and benefits associated with two different states of the world—with and without evaluative research taking place. Possible research results are attached probabilities reflecting their likelihood of occurrence (weightings) and form a series of combinations.

For each combination, a measure of the expected benefits of the trial is obtained, in the form of the Expected Incremental Cost-Effectiveness Ratio (EICER), which shows the cost for an additional unit of benefit. EICER values can be used to provide a plausible interval within which the expected benefits of the proposal is expected to fall and can be compared against decision-makers' threshold values. According to the authors[67], by

providing estimates of the possible returns from different research proposals, the model can help research funding organisation to allocate the available budget in a more transparent and efficient way. This model is looked at more closely in the next chapter.

2.3. Main analytic frameworks

Analytic models identified in this review followed three distinct methodologies and, on this basis, they were classified into different frameworks. Models following the rationale found in Weinstein[94] and Eddy's[90] work have been collectively termed "*direct assessments of the cost-benefit of research*"[55[p.37]] or, more commonly, 'payback of research'[55;80;110]. Seven models are classified under this framework[67;90-92;94;98;100], all of which are underpinned by the notion that the value of evaluative research can be inferred by the benefits that this may bring about through changing clinical practice.

The second framework is 'value of information' (VoI). This method, adapted to the specific context of health care decision-making in the mid-1990s[97;115], is based on the notion that the value of research can be measured in term of reduction in uncertainty and improved decision-making. Last, the 'welfare loss' framework, laid out in Phelps and colleagues' work[95;103], aims to determine broad areas where research is needed. While this work can show which clinical areas would benefit from more research, its usefulness in identifying the benefits from conducting evaluative research in specific topics or proposals is deemed limited[80;110].

The interest in the present work is in the prospective 'payback' and 'value of information' (VoI) frameworks, which represent the two most prominent analytic frameworks

proposed to assist research priority-setting[55;80;116]. VoI and the latest 'payback' model, PATHS[67], are described in more detail in the following chapter.

2.4. Existing literature on 'payback' and 'value of information'

Two publications looking at 'payback' and VoI with a view to contrasting and assessing the approaches were identified. Chilcott *et al*.[55] published a review of the use of modelling in designing and prioritising clinical trials. As part of this review, the authors searched the literature for analytic approaches proposed to provide numerical estimates of the benefits of conducting clinical trials. Their work identified models published up to 1999, which were grouped into two categories: those based on Eddy's[90] direct assessment of the cost-effectiveness of research (including the models by Eddy[90], Detsky[91], Drummond *et al.*[117], Buxton *et al*.[104], Townsend and Buxton[92], and Davies *et al*.[118]), and those based on Bayesian expected value of information. At the time the review was published, the most recent model discussed[118] was a precursor of the Davies *et al*.[98] model, thus, for pragmatic reasons, the most comprehensive study following the 'payback' methodology—the PATHS model[67]—was not discussed or assessed.

As the purpose of Chilcott and colleagues[55] was to review the literature and give indications of the methodologies available, no practical application of 'value of information' and 'payback' on case studies was conducted. However, some conclusions were drawn, on the basis of existing applications of 'payback' and VoI in the literature. The authors found the methodologies underlying 'payback' and VoI to be intuitive, logical and coherent, but they also highlighted potential limitations of the frameworks. With regards to 'payback' based methods, the authors[55] stressed the subjectivity in the choice

of possible research outcomes and likelihoods for specified outcomes, and indicated that there is a need for the impact of such choices to be explored through sensitivity analysis. As far as VoI is concerned, the authors[55] praised the framework's potential to help with research design decisions, but also highlighted the fact that the framework appears to be more complex than 'payback'. In addition, Chilcott *et al*.[55] suggested more research to address the fact that, at the time the review was published, VoI results failed to account for the recognition that adoption decision are typically no implemented perfectly. Overall, the authors felt that analytic approaches have the potential to help in priority-setting and predicted a growing interest in analytic approaches in the future.

The second study[119], authored by Fleurence and published in 2007, aimed to *"provide a practical application between two different methods, 'payback' and EVI* [value of information] *that have been proposed to set priorities in research"*[119[p.1346]]. To this end, the application made use of case studies in the form of three trials in the area of osteoporosis and one trial on pressure ulcers. Interventions considered in these trials were provision of hip protectors and prescription of vitamin D and calcium against osteoporosis and provision of alternating pressure mattresses in pressure ulcers. The application of the methods was facilitated by two decision analytic model, which were constructed for each of these disease areas.

For the application of 'payback', Fleurence[119] used the latest available model (PATHS[67]). In doing so, and in line with previous applications of PATHS[67], the author specified three alternative scenarios: positive, inconclusive and negative. The clinical effectiveness of the interventions of interest (expressed as relative risk of experiencing an event (e.g.

fracture)), the change in clinical practice with and without the trial and the likelihood of observing the specified effectiveness in the trials were based on the author's assumptions. Results, in terms of net monetary benefits, were first expressed in a non-weighted form—effectively assuming that each outcome had the same probability of occurring—and they were subsequently weighted by different likelihood weights, forming an optimistic, neutral and pessimistic combination. Details on these practical applications, including a description of the specified possible results and hypothesised change in clinical practice are given in given in Appendix 2.B, Table 2.d.

For the 'value of information' analysis, Fleurence[80] used the simulated results of probabilistic sensitivity analyses undertaken as part of the osteoporosis and pressure ulcer models to look into the expected value of perfect information (EVPI) associated with further research in the investigated areas.

The results of the 'payback' application suggested that the assessed trials in osteoporosis (RECORD, Vitamin D/Calcium trial and Hip Protector trials) were expected to be cost-effective under the assumptions made about their possible outcomes, whereas the trial in pressure ulcers (Pressure trial) would be beneficial only if the assessed intervention (i.e. alternating pressure mattresses) was found to be effective (i.e. under a positive outcome). EVPI results suggested that all trials were associated with net monetary benefits in excess of the cost of the trial, making these trials potentially cost-effective investments. With regards to the methodologies *per se*, Fleurence[119[p.1350]] suggested that "*'Payback' provides a method to assess the cost-effectiveness of specific research designs, while EVI* [value of information] *makes inferences concerning a wider range of potentially*

relevant parameters, including interventions and study population that would be found in specific research designs, such as clinical trials". According to Fleurence[119], both 'payback' and VoI can be used by funding organisation to assist decisions on which proposed trials to fund while, in addition, VoI methods can help with research design and decisions about which interventions to research within a given disease area. The author highlighted potential problems around the feasibility of carrying out 'payback' and 'value of information', and drew attention to later method's inability to account for imperfect implementation.

2.5. Need for extension in existing evidence

Fleurence's study[119] is novel in that it looks into 'payback' and VoI by applying the approaches to case studies of proposed trials. As such, it serves as a starting point for additional work to build on this study, with a view to providing a more comprehensive assessment of the frameworks.

A shortcoming of Fleurence's study[119], and, at the same time, an opportunity for further work, is that 'value of information' analysis in the paper includes only an assessment of the expected value of perfect information (EVPI). While EVPI is an important concept, it is usually seen only as the starting point in VoI analysis, which is typically furthered to include expected value of perfect parameter information (EVPPI)[120;121] and, if possible, expected value of sample information (EVSI) analyses[115;122]. In a recent publication, Eckermann *et al*.[123] consider EVPI conducted on its own to provide only limited information, and stress that the full potentials of VoI can be only realised by conducting EVPPI and EVSI analysis. In addition, as Fleurence's[119] work was carried out and published

in 2007, it could not take into account recent advances related to the value of implementation. Recent work on value of implementation[124] is, nonetheless, important, as it aims to address an issue often portrayed as a 'soft spot' of VoI, and aspires to strengthen the case for its use in decision making.

With reference to the methodology for conducting 'payback' analysis, there also appears to be room for further work on aspects related to carrying out 'payback' analysis for trials assessing more than two treatments. In Fleurence's study[80], as explained in Chapter 6, the employed assumptions when comparing multiple treatments is that more than one treatment can be cost-effective at the same time. This assumption underpins the 'payback' calculations for the RECORD case study which compares four treatments, but it is nevertheless weak, as, in a comparative analysis, only one of the assessed treatments may represent the most cost-effective option[31;125].

These limitations have important consequences. First, without an assessment of all the relevant VoI concepts (including EVPPI, EVSI and value of implementation) a discussion around the strengths, weaknesses and potential of the framework is bound to be incomplete. In addition, by not undertaking more complex and potentially time consuming analyses (EVPPI and EVSI) (see Ch.8, section 8.5 for a discussion of time and expertise required for undertaking these analyses), it is difficult to draw reliable conclusions about the feasibility of undertaking 'payback' and VoI analyses.

Given the above, the present study seeks to extent the existing literature by undertaking a comprehensive analysis of 'payback' and VoI, one that a) accounts for recent advances in the methodologies, b) includes all the relevant concepts consisting the frameworks,

and c) looks into the value and potential role of the frameworks from different viewpoints that are considered relevant to research funding organisation. In this way, this work aims to provide a more spherical assessment of the frameworks and a more complete discussion of their potential role and value in research priority-setting.

2.6. Discussion

Two main approaches to priority-setting were identified and described in this chapter: deliberative and analytic. The main characteristics, similarities and differences of these approaches are summarised in Table 2.2.

A key difference between 'deliberative' and 'analytic' approaches relates to the principles underpinning each of them. In 'deliberative' approaches, decisions on whether a proposed piece of research should be funded are made by appointed experts through a process involving deliberations about the need for research and the value of the proposed research programme. On the other hand, 'analytic' approaches aim to introduce an element of objectivity, by advocating a framework where research funding decisions would be made on the basis of explicit information on the value of research. Given this, in 'deliberative' approaches the value of a proposal is inferred through discussion with the help of pre-determined criteria, whereas, in analytic approaches, the value of research is inferred directly from the benefits that the programme is expected to produce, which are calculated by following specific methodologies.

Table 2.2: Summary of main characteristics of deliberative and analytic approaches

	Deliberative approach	Analytic approaches	
		'Payback of research'	'Value of information'
Underlying principle	Judgements on the desirability of proposed research made by appointed experts on the basis of deliberations about the value of the proposed research programme.	Judgements on the desirability of proposed research should be made on the basis of explicit numerical calculation of the benefits expected from conducting a proposed research study.	
Estimation of potential benefits of research	Implicit, through discussion and deliberations. Discussion is typically guided by pre-determined criteria.	Explicit quantification of potential benefits, through calculations that link possible results of research to change in clinical practice and subsequent benefits to the population.	Explicit. Through the calculation of the expected gains from reduction in decision uncertainty.
Decision rules	Fund research which appears useful in light of expert opinion (panel and reviewers) with connection to specific criteria	Estimate the cost and benefits of research. Recommend research which results in net benefits (benefits obtained for a value lower than a hypothetical willingness to pay for these benefits)	
Current use	Widely used in decision making, by public research funding organisations and medical research charities.	Not used at present	
Proposed use		Proposed as an additional consideration for research funding decisions, alongside other relevant criteria[67].	Proposed as a 'criterion' for research funding decisions. Research proposals which do not meet the necessary conditions should be ruled out[120]

In relation to decision rules, 'deliberative' approaches advocate prioritising proposals that appointed experts consider to perform well against predetermined criteria. Typically, such criteria relate to the need for research in the area, the scientific rigour of the study and the ability of the research team proposing the study to complete it successfully. On the other hand, analytic approaches advocate decision rules which are underpinned by economic principles, suggesting that an activity—here, a programme of evaluative research—should be undertaken if it is expected to result in net benefits. In this context, net benefits arise when the monetary value of a unit of health benefit (such as a quality-

adjusted life year (QALY), exceeds the amount which society or decision-makers are willingness to pay for this benefit.

While 'deliberative' approaches are commonplace, analytic approaches are not formally used by funding organisations. Suggestions on the use of the most prominent analytic frameworks, 'payback' and VoI, appear to converge towards use of 'payback' and VoI results alongside other criteria, which may be deemed relevant by the organisations tasked with making funding decisions[67;120].

Deliberative and analytic approaches present specific strengths and limitations. The former have been seen as fairly uncomplicated to undertake[66] and appear to allow a quick turnover of results, which is an important strength when timely decisions on funding research are needed. In addition, deliberative approaches make use of informed views of panel members who typically have an expertise on the topic for which research is proposed as well as a thorough understanding of the research needs in this area. Last, decision-making based on criteria and deliberations gives funding bodies the flexibility to pursue objectives considered important to them. On the other hand, decisions reached using deliberative methods are, to a great extent, subjective as they depend largely on panel members' judgements. Indeed, under such arrangements decision-making may be affected by the synthesis of the panel and committees, which increases the risk that decision may be based on the views of panel members who are more vocal during the discussion, and raises the possibility that the views of members absent from the discussion may be overlooked[79].

These limitations, together with the acknowledgement that the allocation of research resources should be based on explicit evidence on research proposals' costs and benefits[90;100], have led researchers to develop analytic methods[90;100]. Such methods have been seen as a promising means of improving the transparency and explicitness in research priority-setting[71]. As Donaldson and Sox[64[p.67]] point out, analytic approaches are "*open to review and accountability, and are amendable to examination and adjustment not only of the results but of the methodology itself*". Given this, analytic models appear to have the potential to offer greater reassurance as to whether research funding is allocated in an efficient way[55;64;65].

However, analytic approaches also have limitations. First, they are usually resource-intensive, in the sense that they require more time and effort than deliberative approaches[55;64]. As a consequence, the additional time that analytic methods may require to give results may delay funding decisions and lead to delays in research. In addition, due to their nature, such approaches have been seen as 'mechanistic' and, because of this, decision-makers might be reluctant to use results from such approaches as a basis for decisions[64;110]. The extent to which the benefits from the use of analytic approaches, in terms of improved transparency and efficient decision-making, compensate for these drawbacks is unclear.

2.7. Chapter overview

This chapter reports the results of a review carried out to identify approaches to priority-setting for research. Two main approaches aimed at guiding research funding decisions

were distinguished: deliberative approaches, which involve discussing research topics and proposals on the basis of their merits according to predetermined criteria, and analytic approaches, which aim to infer the value of proposed research studies by quantifying the potential benefits that research may bring about. Strengths and limitation of each of the approaches were discussed drawing on published literature.

Nine analytic models were identified. With the exception of Phelps and colleagues' 'welfare loss' work[95;103], the models were classified in one of two overarching frameworks: 'payback of research' and 'value of information'. The principles, methods and identified applications of these frameworks are the focus of the following chapter.

CHAPTER 3. 'Payback of research' and 'value of information'

The previous chapter identified nine analytic models and classified them into two overarching frameworks: 'payback of research' and 'value of information' (VoI). Both 'payback' and VoI make extensive use of methods used in economic evaluation of health care technologies, thus the first part of this chapter introduces and describes central concepts in economic evaluation. The second part of the chapter aims to provide an in-depth description of 'payback' and VoI, by drawing on the existing literature.

3.1. Economic evaluation of health care technologies

In order to improve and maintain the health of the populations they serve, health care systems around the world provide a range of technologies and treatments. In publicly-funded health care systems with constrained budgets, decisions are often needed about which treatments should be covered financially and provided to the population.

Such decisions, it is acknowledged, need to be informed by evidence on treatments' costs and health benefits[8;31;126;127]. Such evidence is routinely produced by economic evaluations, which are systematic comparisons of alternative technologies, interventions and treatments in terms of their cost and consequences[31]. Depending on the nature of the treatment, costs may include expenditure for obtaining or implementing the treatment itself, related use of hospital and community services, related personal expenditures and, in some cases, costs due to productivity loss[31;128;129]. Consequences are

typically health-related and can be expressed in monetary, natural or quality-adjusted units[31;130].

The rationale for the use of economic evaluation to inform resource allocation is based on two paradigms: the 'welfarist' approach, under which evaluations determine the relative desirability of different resource allocations on the basis of individuals' utility[128;131], and the 'extra-welfarist' or 'social decision-maker' approach where appointed decision-makers decide what actions should be pursued in order to maximise socially desirable attributes—most importantly health[132;133]. The nature of the question to be addressed and the way consequences are measured determine the choice between separate—but arguably similar—techniques of economic evaluation[134;135].

3.1.1. Cost-benefit analysis

In line with principles of welfare economics, judgments on whether an activity (e.g. reimbursement and provision of a treatment) should be undertaken ought to be made with reference to whether its outcomes will contribute towards improving society's welfare[128]. Such judgements target allocative efficiency questions and can be informed by cost-benefit analysis (CBA), a technique that expresses the costs and benefits associated with an activity in monetary terms[132;136].

Valuing both costs and benefits in monetary terms allows a comparison between an activity's input (costs) and output (in the specific context, health benefits). If the health benefits expected to arise from the activity exceed its costs, the activity will result in 'net social benefits' and is worth undertaking. While costs can be easily measured in monetary

terms, obtaining a monetary valuation of health benefits is more complex and it requires translating health improvement into increased productivity according to wages (using the human capital approach)[128], or by eliciting peoples' willingness to pay for those benefits through observing their 'stated' and 'revealed' preferences[130;137].

Despite this technique's firm foundations in welfare economics, its use in economic evaluation has been limited, to a great extent due to practical and ethical difficulties with valuing health and health benefits in monetary terms[135;138].

3.1.2. Cost-effectiveness and cost-utility analyses

Whilst CBA can be used to answer questions on whether an activity is worth undertaking, cost-effectiveness analysis (CEA) can show which of competing activities should be undertaken if the aim is to achieve a given level of output (for example, health benefits) from the least input, or, equivalently, to achieve the greatest output out of a given level of input. In this respect, CEA is appropriate for addressing questions of technical efficiency[135]. Tackling such questions requires measuring each activity's costs in terms of monetary units and benefits in terms of relevant natural units, which, for instance, may be years of life saved, cases detected or episodes avoided[139].

CEA involves pair wise comparisons between two or more treatments and its results are commonly expressed as an incremental cost-effectiveness ratio (ICER), a measure which represents the difference in costs between two compared treatments, over the difference in benefits between these treatments. As an example, if the comparison is

between a commonly used treatment A and an alternative, more costly and more effective treatment B, CEA results can be expressed in terms of the following ICER:

$$ICER_{B-A} = \frac{Cost_B - Cost_A}{Effectiveness_B - Effectiveness_A} = \frac{\Delta C_{B-A}}{\Delta E_{B-A}}$$

In this case, the ICER represents the additional cost of obtaining an additional unit of health benefit. A special form of CEA is cost-utility analysis (CUA)[31;128;130]. The defining difference between CEA and CUA is that the latter measures benefits in terms of units which combine length of life in a health state with people's preference for this state. The most commonly used measure in CUA is the quality-adjusted life year (QALY). A QALY represents the equivalent of one year of life in full health and is calculated by weighting the period of time a person spends in a particular health state by the health-related quality of life (HRQoL) he or she experiences during this period. Typically an individual's HRQoL is estimated by obtaining a description of his/her health state and valuing this state according to the individual's[140] or society's[141] preferences for it. Cost-effectiveness and cost-utility analyses are widely used to inform treatment coverage decisions in a number of countries[14], including England and Wales[114], Australia[8] and Canada[127].

3.1.3. Cost-minimisation and cost-consequence analyses

A further form of economic evaluation is cost-minimisation analysis (CMA). The method is seen as a special case of CEA and has been considered appropriate in analyses where treatments are perceived to be of equal or equivalent effectiveness[31;142]. Given this, the comparison in CMA is in term of costs, with the less costly treatment being seen as

preferable. Although in theory CMA is a simpler version of CEA, in practice the situations where the use of CMA may be appropriate are rare. This is because, as Briggs and O'Brien[143] explain, the uncertainty surrounding estimates from sample studies such as clinical trials means that failure to detect a difference between treatments' effectiveness may be mistakenly perceived as a proof of equivalence, in which case CMA would lead to spurious conclusions[143].

Last, an economic evaluation can take the form of a cost-consequences analysis (CCA). This technique aims to address scepticism about combining results into a single number (e.g. cost per QALY) and its distinctive characteristic is that it presents results in a disaggregated way, typically in the form of a table documenting the costs and consequences associated with a treatment[31;144]. In CCA, no attempt is made to synthesise the available information into a unique number; instead decision-makers are left to interpret the information and assign their own values to the costs and consequences associated with the assessed treatment. This characteristic makes CCA an attractive approach as it is simple, consistent with the 'welfarist' roots of economic evaluation[144] and avoids assumption made in the process of constructing a single outcome[130]. However, the approach has been criticised on grounds that, by leaving interpretation and valuation of the results to decision-makers, it may allow room for unclear or arbitrary rules to enter the decision-making process[125].

3.1.4. Decision rules in economic evaluation

The primary aim of an economic evaluation is to inform resource allocation decisions. To this end, its results need to be interpreted clearly, using unambiguous decision rules. Decision rules for judging the results of economic evaluations differ across techniques. In CBA, the decision rule is relatively straightforward: if the benefits (valued in monetary terms) associated with a treatment exceed its cost, the treatment is expected to result in 'net benefits' and should be provided[128;135].

On the other hand, CEA and CUA results can determine whether a specific treatment compares favourably to its best alternative, that is, whether it provides the same benefits for lower cost or more benefits for the same cost. However, CEA results cannot determine whether a treatment is worth undertaking, especially when there is a trade-off between increased costs and increased health benefits (and *vice versa*). This issue can be addressed by attaching a monetary value to a unit of health benefit. Such a value, often called the 'cost-effectiveness threshold' or 'ceiling ratio' (commonly denoted by) has been seen as the decision-maker's or society's willingness to pay for an additional unit of benefit (or willingness to accept as compensation for a unit of benefit forgone)[128;139]. Both theoretical[145] and empirical ways[146] of establishing the value of have been suggested but no consensus exists over how the ceiling ratio should be set[147] and who should be responsible for setting such a value[148].

According to the decision rules used in CEA and CUA, if treatment B is more expensive and more effective than treatment A:

$$Treatment\ B\ is\ cost-effective\ if\ ICER = \ C_{B\text{-}A}/\ E_{B\text{-}A} <$$

Equivalently, if treatment B is less expensive and less effective than its comparator, the decision rule would be reversed:

$$Treatment\ B\ is\ cost-effective\ if\ ICER = \ C_{B\text{-}A}/\ E_{B\text{-}A} >$$

Here, ΔC_{B-A} and ΔE_{B-A} represent the difference in costs and effectiveness between treatments A and B, respectively. If treatment B is less effective and more costly, the treatment is deemed not to be cost-effective, while if treatment B is more effective and less costly that its comparator, treatment B is considered as cost-effective.

Despite the usefulness of the ICER as a summary measure of cost-effectiveness, being a ratio means that it has some undesirable mathematical and statistical properties which hinder analysis and presentation of results[115;149;150]. As a response, researchers have suggested a simple re-arrangement which incorporates the ceiling ratio and allows CEA and CUA results to be presented in a tractable, linear form[115;151]. In this, the standard measure of cost-effectiveness and the decision rule are combined to give the Incremental Net Monetary Benefit (INMB) formulation:

$$INMB_{B-A} = \ (\Delta E_{B-A} \times \ \ ) - \Delta C_{B-A}$$

Here, positive INMBs indicate that treatment B is cost-effective. Decomposing the above formula to consider the average net benefits associated with specific treatments gives:

$$NMB_A = \ (E_A \times \ \ ) - C_A$$

$$NMB_B = (E_B \times \quad) - C_B$$

Here, E_A and E_B represent the effectiveness associated with treatments A and B, and C_A and C_B are the costs of treatments A and B, respectively. Presenting the results of CEA and CUA analyses in this way offers significant advantages over ICERs: it provides an easier interpretation of the results of analyses where multiple treatments are compared[152], offers a tractable estimate of the variability in the results[150] and circumvents problems associated with the interpretation of ratios[153].

3.1.5. Decision analytic modelling for economic evaluation

In relation to the analytic methods they employ, cost-effectiveness and cost-utility analyses are often distinguished between trial-based and modelling studies[154;155]. The former studies are carried out to analyse evidence obtained from a single clinical trial[155;156], while the latter synthesise evidence from many sources by using analytic structures[157]. Limitations around evidence produced by a single study as well as the recognition that all available evidence needs to be taken into account have led health economists and decision-makers to regard modelling studies as the preferable framework for synthesising evidence and informing decision-making[158-160].

Decision modelling is an integral part of decision analysis[107] and, apart from in health care, it has been successfully used in different disciplines including engineering, environmental risk analysis and operational research[161-163]. Brennan *et al*.[164[p.1296]] define a model-based analysis as *"a formal comparison of health technologies, synthesising sources of evidence on costs and benefits, in order to identify the best option for decision-*

makers to adopt". In the context of health care, results of models are routinely used in treatment coverage decisions by the National Institute for Health and Clinical Excellence[160;165], which sees modelling as "*an important framework for synthesising available evidence and generating relevant estimates of clinical and cost-effectiveness*"[114[p.42]].

Different types of decision analytic models exist, the most common of which are decision trees, Markov models and individual sampling models[157;166;167]. The main stages in modelling are depicted in Figure 3.1.

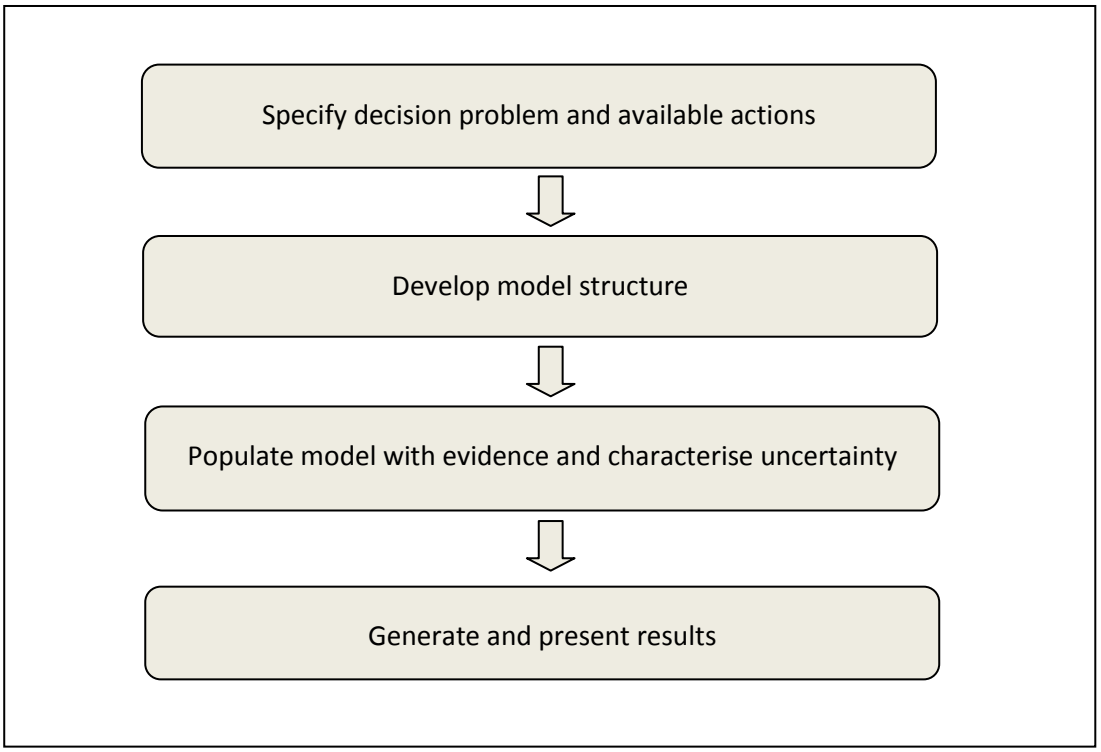

Figure 3.1: Steps in decision analytic modelling

The first step in modelling requires defining the question to be answered. This typically involves specifying the clinical area, the population of interest and the treatments to be

assessed. Once the decision problem and the comparators are specified, a choice is needed on the type and structure of the model. In choosing the appropriate type, pertinent considerations include, amongst other, the existence of interactions between affected individuals[168], the importance of time[164] and the availability of cohort information[164;166]. In addition, the model structure should allow a valid representation of the natural clinical progression of the disease in question, one that takes into account the particularities of the decision problem, that is whether the disease is acute or chronic, whether the options under comparison are diagnostic or therapeutic and whether the risk of experiencing an event changes over time[51;157].

The next step involves populating the model with available information. This process requires identifying and synthesising the available evidence, as well as converting this evidence into a form appropriate for use in the decision model. At this stage, uncertainty arising from the methods used in the model, from the model structure itself, as well as from the use of parameter values obtained from samples needs to be characterised and accounted for, preferably by using both probabilistic and deterministic sensitivity analysis techniques[169;170]. In deterministic sensitivity analysis, the impact of uncertainty is explored by recalculating the results for different plausible values of one or more uncertain parameters. On the other hand, in probabilistic sensitivity analysis, uncertain parameters are assigned probability distributions, rather than single values. Uncertainty in the results is then propagated by drawing a large number of values from each distribution in subsequent simulations—typically by using Monte Carlo methods[171;172].

The process gives a large number of cost and effect pairs together with the uncertainty associated with them[169].

The final stage in modelling involves presenting results numerically, in terms of ICERs and NMBs, and graphically, typically by plotting the generated cost and effect pairs on cost-effectiveness planes (CE planes)[173], cost-effectiveness acceptability curves (CEACs)[174] and, more recently, cost-effectiveness acceptability frontiers (CEAFs)[175]. In brief, a CE plane plots paired estimates of incremental costs and benefits obtained from Monte Carlo simulations on a four-quadrant plane. Depending on the quadrant on which cost-effectiveness results are located, a treatment may be more effective and more costly (North East quadrant), more effective and less costly (South East quadrant), less effective and less costly (South West quadrant) or less effective and more costly (North West quadrant), as compared to an alternative treatment.

In turn, CEACs can be used to represent the probability of a treatment being cost-effective at different ceiling ratio values[174;175]. For different ceiling ratios, a CEAC is produced by counting the proportion of incremental cost and effect pairs with a value less than the ceiling ratio[51]. Last, as there is a possibility that the intervention with the highest probability of being cost-effective may not result in the highest NMBs due to non-linearity in the model, results may also be presented as a CEAF. This is formed by a curve representing the interventions associated with the highest NMBs over a range of ceiling ratio values[175;176]. A description of the methods involved in structuring and analysing decision models for the case studies used in this project are given in Chapters 4 and 5.

3.2. The ‘payback of research’ framework

This part aims to give a detailed description of the ‘payback’ framework, drawing on relevant publications in the literature.

3.2.1. Rationale

The ‘payback’ framework is based on the notion that research is valuable because it generates evidence about treatments’ effectiveness and cost-effectiveness, which can lead to beneficial changes in clinical practice. Change in clinical practice (i.e. adoption and provision of cost-effective treatments, and termination of non-cost-effective care) is expected to result in additional benefits to the population. On this basis, the desirability of a research programme proposed to produce evidence on a treatment of interest can be inferred from the benefits likely to arise from a change in clinical practice informed by the produced evidence.

The framework’s rationale can be illustrated with an example. A hypothetical treatment A may be used by patients with a specific condition. A study (e.g. a clinical trial) is proposed to evaluate the effectiveness and cost-effectiveness of treatment A. If the study reveals the treatment to be beneficial, clinical practice would be expected to change towards greater use of treatment A. On the other hand, if treatment A is not cost-effective, its use will be limited. In either case, change in clinical practice is expected to result in additional benefits in the population, as more patients will be treated with a cost-effective treatment (or, in the case of ‘unfavourable’ results, less people will be using a non-cost-effective treatment). According to the idea underlying ‘payback’, those additional

benefits can be attributed to research (as research generated the information that induced a change in clinical practice) and can be taken into account in judging whether funding and conducting a research programme would be beneficial. The core steps in 'payback' analysis are given below:

a. Specification of hypothetical research results. These represent assumptions about the possible results that a proposed study may reveal for the treatment of interest;

b. Specification of possible change in clinical practice. Assuming that a particular result has been observed for a treatment of interest, this step specifies the change in the treatments' use;

c. Estimation of the stream of possible costs and benefits expected to arise under different research results and subsequent change in clinical practice, and

d. Calculation of expected costs and benefits by accounting for different likelihood of possible research results being observed.

Costs and benefits are usually calculated for two distinct 'states of the world', with research taking place and without research. The difference between them gives the additional benefits associated with funding and conducting the proposed study.

3.2.2. Main components of 'payback' models

The main components and characteristics of the 'payback' framework are discussed below with reference to identified 'payback' models[67;90-92;94;98;100]. A brief summary of these characteristics is given in Table 3.1.

3.2.2.A. Purpose

With regards to their purpose, some 'payback'-based models have an explicit focus on setting priorities for primary, evidence-generating research (typically clinical trials)[91;92;97;100] while other[67;90;94;95;98] can be potentially useful for funding decisions on both primary and secondary, evidence-synthesising research. Thus, depending on the variation used, the framework appears capable of informing funding decisions for either primary or secondary research.

3.2.2.B. Specification of possible research outcomes

The first important step in 'payback' models is the specification of possible research results. These represent hypothetical but plausible outcomes from research, typically expressed in terms of different measures of effectiveness. An important assumption is that a possible result represents the true effectiveness (or cost-effectiveness) of the treatments in question, which holds true irrespective of whether a study has been carried out and revealed it or not. Different terms are used for such results, including 'Delta results'[90] and 'exemplar outcomes'[92]. In terms of considerations to be taken into account when specifying possible research results, most of the identified studies agree that these should cover a range of eventualities and be meaningful to decision-makers[67;90;92;100].

Table 3.1: Characteristics of 'payback' models

Study	Aims	Methods			Results and conclusions	
	Prioritisation of primary or secondary research	**Specification of possible outcomes**	**Consideration of impact of outcomes on clinical practice**	**Quantification of costs and benefits**	**Measure of final results**	**Decision criterion**
Weinstein(1983)[94]	Primary research (clinical trials)	Not explicitly. Potentially plausible to incorporate different exemplar outcomes	Yes	Yes	Cost per year of life saved	Prioritise research programme with lowest cost per life-year saved
Eddy (1989)[90]	Secondary research (technology assessments)	Yes. Two possible exemplar outcomes (treatment cost-effective, treatment not cost-effective)	Yes	Yes	Number of severe retardations avoided per year	Estimate cost per benefit ratio for each proposed research programme. Rank research programmes in increasing order of cost per number of deaths avoided ratio Prioritise research programme with most desirable ratio
Detsky (1990)[91]	Primary research (clinical trials)	Yes. A number of possible results (risk reduction) were considered	Yes	Yes	Cost per year of life saved	Prioritise research programmes if they are considered cost-effective (in terms of cost of undertaking the assessment (sample size) per expected number of health benefits).

Study	Aims	Methods			Results and conclusions	
	Prioritisation of primary or secondary research	**Specification of possible outcomes**	**Consideration of impact of outcomes on clinical practice**	**Quantification of costs and benefits**	**Measure of final results**	**Decision criterion**
Drummond *et al.* (1992)[100]	Primary research (clinical trials)	Yes. Two possible exemplar outcomes: 'positive' (patient identified with disease) and 'negative' (patient is missed)	Yes	Yes	Cost per years retaining vision	Prioritise research programme if it results in acceptable cost per outcome ratio
Townsend and Buxton (1997)[92]	Primary research (clinical trials)	Yes. Three possible exemplar outcomes: 'positive' (treatment effective), 'negative' (treatment not effective) and 'inconclusive'	Yes	Yes	Cost per QALY	Prioritise research programme if it results in acceptable cost per outcome ratio
Davies *et al.* (2000)[98]	Secondary research (technology assessments)	Yes. Two possible exemplar outcomes: 'positive (treatment effective) and 'negative' (treatment not effective)	Yes	Yes	Cost per unit of outcome (various disease-specific measures for different case studies)	Prioritise research programme if it results in acceptable cost per outcome ratio
Townsend *et al.* (2003)[67]	Either primary or secondary research	Yes. Three possible exemplar outcomes: 'positive' (treatment effective), 'negative' (treatment not effective) and 'inconclusive'	Yes	Yes	Cost per unit of outcome (various disease specific measures for different studies)	Prioritise research programme if it results in acceptable cost per outcome ratio

In three of the identified 'payback' models[90;98;100], it is assumed that research will produce two kinds of results:

a. 'Positive', where research reveals that the treatment of interest is effective and cost-effective, and

b. 'Negative', where the treatment is shown to be inferior in terms of its effectiveness and cost-effectiveness.

The studies by Townsend and Buxton[92] and Townsend *et al*.[67] allow for the additional possibility of 'inconclusive' outcomes[67;92], while Detsky[91] represents possible results in the form of a distribution. Arguably, the latter resembles more closely the type of results that are likely to be observed in a clinical study.

3.2.2.C. Change in the use of the assessed treatment

Another important point in the 'payback' methodology relates to the impact of the results on clinical practice. Here, it is assumed that observing a specific research outcome will trigger a change in clinical practice. Trial results showing the treatment not to be effective are expected to result in a gradual decline in its use, while 'positive' research results (i.e. treatment is effective or cost-effective) are anticipated to lead to greater use of the treatment in clinical practice.

All the identified studies attempt to establish a link between the results and subsequent use of the treatment of interest. In general, clinical practice is assumed to change on the basis of a treatment's hypothesised effectiveness. The contemporary view that cost-effectiveness is

a key consideration in policy change is explicitly accounted for in the PATHS model[67]. Eddy[90] stresses that change in policy will depend on a number of factors, including the results of other assessments, existing policies that might affect the uptake of the assessed technologies, geographic regions and time periods. In the absence of further research (i.e. if the proposed research study is not undertaken), it is typically assumed that clinical practice will not change[90;92;100].

The concept of change in use is particularly important in 'payback', because it allows translating possible research results into potential benefits experienced by the population[67;91;98;100]. As obtaining hard evidence on the possible change in clinical practice is problematic, estimates for this change usually comes in the form of expert opinion[67].

3.2.2.D. Estimation of costs and benefits

All the 'payback' models attempted to quantify the costs and benefits following change in uptake of a treatment. Information for this is typically derived by combining existing evidence on costs of the treatment—typically obtained from the literature—with the hypothesised research results[67;92;100].

Costs and benefits are first estimated for an individual patient and they are subsequently extrapolated to the population that is expected to benefit from a decision informed by additional evidence over a number of years. Costs are typically measured from a provider's perspective. An exception is the Diabetic Retinopathy Study, where Drummond *et al*.[100] explored a variety of alternative perspectives, including those of government, health care sector, patient and the community. Benefits are typically expressed as disease specific

measures[67;90;98;100] and natural units such as premature death avoided[91;94], while in one study benefits were expressed as QALYs[92].

3.2.2.E. Weighted results and decision rules

Typically, the streams of costs and benefits associated with each possible research result are weighted by the likelihood of the result being observed. Such likelihoods weights were accounted for in all the identified models. Eddy[90] determined the probability of observing a 'positive' or 'negative' result to be 0.50; similarly, Davies *et al*.[98] specified the probability of 'positive' results at 0.67 on the premise that approximately two out of three treatments assessed in clinical trials show 'positive' findings. In two studies, the authors[91;102] followed a more sophisticated approach, where a probability distribution was assigned to the possible difference in effectiveness between assessed treatments. Likelihood weights have a sizeable effect on final 'payback' results and thus, it is important that chosen values are plausible and justified.

Different approaches have been followed in assigning likelihood weights to possible results. In early 'payback' models, each specified research result carries the same likelihood of occurrence[90;94;98;100]. A more sophisticated approach involving different combination of weights is employed by Townsend and colleagues[67;92]. In these studies, three possible research results ('negative', 'positive' and 'inconclusive') merge in 'optimistic', 'neutral' and 'pessimistic' combinations.

The area where 'payback' models present the greatest diversity is the interpretation of the generated results and the determination of rules for action. In general, models aim to be

consistent with decision rules used in economic evaluation and attempt, whenever possible, to calculate ratios of costs per expected benefits, with the difference that, in this context, the comparison is between a state with and without research[67;92;98;100].

It is often not clear how such ratios should be interpreted. In cases where ratios indicate that an assessment would result in cost savings and increased effectiveness, an obvious option would be to advocate carrying out research. Nonetheless, a clear rule is needed for the common situation where research is associated with additional benefits and higher costs. In such a case, Townsend and Buxton[92] compared the estimated ICER of the Hormone Replacement Treatment (HRT) trial against cost per QALY values of commonly-used interventions such as renal transplant or breast cancer screening, and pointed out that the HRT trial appears to represent value for money. As Davies *et al*.[98] point out, interpretation of results and decision rules are topics where further research would be particularly useful.

3.2.3. The Preliminary Assessment of Technology for Health Services (PATHS) model

As explained above, the 'payback' framework has been subject to constant development, with the methods introduced in early models[90;94] being refined and updated in subsequent studies[67;91;92;98;100]. The latest identified 'payback' model is the Preliminary Assessment of Technology for Health Services (PATHS)[67], published in 2003. This model combines the principles of early 'payback' work with methodological advances and current trends in health care decision-making (e.g. the introduction of economic consideration in treatment adoption decisions) and it is the focus of the present study.

Similarly to previous ‘payback’ models, PATHS aims to measure the additional benefits that would accrue from conducting research and changing clinical practice according to its results. As the results of research cannot be known in advance, this analysis is undertaken for a series of hypothetical, but possible, results.

The model is operationalised through a series of core steps. The first step involves specifying the possible results that a trial on a specific treatment of interest may give. For practical reasons, these are usually narrowed down to a ‘favourable’ result, where the assessed technology is shown to be effective, an ‘inconclusive’ result, where it is not clear whether the technology is more effective than its comparators, and an ‘unfavourable’ result, where the assessed technology appears to be inferior to its comparators.

Different research results are expected to have different impact on the prescription of the assessed treatments. For example, ‘favourable’ results might increase the use of the effective treatment in clinical practice; ‘unfavourable’ results would typically limit the use of the treatment, while ‘inconclusive’ results would typically not have an effect on current prescription patterns. On the other hand, if the trial is not conducted and no new evidence is generated, clinical practice is expected to remain largely unchanged.

The next step involves estimating the stream of costs and benefits (or net monetary benefits) that would arise for each possible result under two situations: a) with the trial taking place and b) without the trial taking place. In the former case, that is, with research being conducted, costs and benefits are those expected to accrue given the specified research results and the hypothesised change in prescription patterns in the light of these

results. Costs will involve the expenditure for conducting the trial, as well as the cost associated with providing the treatment in question. In the latter case, where no trial is carried out and clinical practice is expected to remain as it was, costs and benefits are those associated with a specified research result, but in this case, as the results will not be observed, no change in practice is expected to take place and no additional benefits will accrue from increasing the use of a cost-effective treatment (or restricting the use of the non-cost-effective treatment).

As some results are more likely than other, probabilities can be attached to them to represent the likelihood of each specific result occurring. Given this, different combinations can be formed, such as an 'optimistic' combination, where a 'favourable' result is more likely to be observed, a 'pessimistic' combination, where a greater likelihood is given to an 'unfavourable' result and a 'neutral' combination, where each result is given equal likelihood of occurrence.

For a given combination, the expected costs and benefits of the trial are expressed as an Expected Incremental Cost-Effectiveness Ratio (EICER), which is calculated as the difference in costs with and without the trial over the difference in effectiveness with and without the trial. EICERs can be subsequently compared to decision-makers' willingness-to-pay value for a unit of benefit. As Townsend *et al*.[67[p.17]] suggest "*the results of the application of the model would be presented in a report to a research-funding organisation, in which this EICER would be weighted with other payback considerations in the final decision about whether the research project should be funded or not*". The authors also point out that the results obtained from the model do not aim to replace existing processes of research prioritisation,

but rather to complement them, as the currently used process of peer review and criteria-based approaches are still needed to consider benefits that are not directly quantifiable and would not be reflected in the PATHS results[67].

3.2.4. Practical applications of the PATHS model

Separate searches were carried out to identify studies reporting practical applications of the PATHS model. Details on search strategies and numbers of articles identified are given in Appendix 2.A. The majority of the identified studies related to retrospective 'payback', a stream of research undertaken at Brunel University and focusing on assessing the returns to health services research by evaluating the extent to which research enhances knowledge and brings about health sector and broader economic benefits[93;177]. As such work looks into the benefits from research which has already taken place and reported results, retrospective 'payback' studies are not directly relevant to the aims of this thesis.

Two studies reporting practical applications of PATHS were found in the literature. The first was the study where the PATHS model was first presented[67]. There, Townsend *et al*.[67] applied the developed model to three case studies representing proposals for randomised controlled trials. The application aimed to illustrate the PATHS methodology as well as to assess the validity of its results. For this, the predicted results from the *ex ante* application of PATHS were compared to *ex post* estimates of the costs and benefits expected to accrue from the trial, the latter being estimated on the basis of observed results derived from trials which were funded and completed. Comparisons were possible in two of the three case

studies, in which the *ex ante* and *ex post* estimates were found to be in broad agreement. This was seen as a positive indication of the model's validity[67].

The second study was carried out by Fleurence[116] in 2007. The author used the model to explore whether three clinical trials of preventive interventions against fractures and one trial against pressure ulcers should take priority for funding. In this study, Fleurence[116] presented results in the form of INMBs, rather than EICERs. The study findings indicated positive NMBs associated with three of the four trials, suggesting that funding these trials would be 'cost-effective'. Details on the comparators, specified trial results and assumption on change in clinical practice following each possible result are given in Appendix 2.B, Table 2.c and Table 2.d.

3.2.5. Critique of 'payback of research' in the literature

The potential of 'payback' models to assist with research funding decisions has been often highlighted in the literature[55;65;98]. Such models present strengths that make them attractive. First, the core 'payback' methodology appears to be consistent with well-established principles of economic evaluation, in that it seeks to determine the additional costs and benefits from an action (here, funding and conducting research) against another (not conducting research). Because of this, as Chilcott and colleagues[55] explain, the framework can address the type of questions that decision-makers are ultimately interested in, that is, whether allocating resources to a specific research programme will be beneficial, given the costs and benefits associated with it. An additional strength of 'payback' is that it seeks to

identify ‘real benefits’ accruing to the population, by looking at how research results trigger changes in clinical practice[67].

Despite this, ‘payback’ models have been often criticised on the ground that, by judging the value of research according to its ability to trigger changes in clinical practice, the framework gives a greater weight to research on topics that have a potential to change clinical practice than other topics that should be researched because they are associated with significant uncertainty[80;110]. For instance, as Fleurence[80] explains, there may exist a widely used treatment about which there is little uncertainty that it is not cost-effective, but a great potential for gains from switching clinical practice away from this technology, towards a more cost-effective option. Thus, although existing evidence on the technology not being cost-effective may already be robust, the fact that a switch in practice will result in benefits makes this assessment, according to ‘payback’, a good candidate for research[80;110]. This, it has been argued[121], is in contrast with the primary aim of research, which should be to provide information which will reduce uncertainty.

3.3. The ‘value of information’ framework

The following section focuses on the second analytic framework of interest—’value of information’. The first part of the section gives an overview of the framework, by looking into its rationale and specific concepts. The section concludes with a discussion of prominent strengths and criticisms associated with the framework, as they appear in the literature.

3.3.1. Rationale

'Value of information' analysis is part of statistical decision theory—a collection of analytic techniques aimed to assist decision-making under conditions of uncertainty[108;163]. The methodology asserts that, in cases where a choice of action needs to be made and the consequences of alternative actions are not known with certainty, statistical and mathematical tools can be used to assess what the optimal action appears to be on the basis of current information, and in what cases obtaining more evidence to substantiate or retract this choice would be beneficial[106;107;178]. Such decision theoretic techniques appear to be directly applicable to health care; in this field decisions need to be made as to which of alternative treatments should be covered and provided to the population, while, at the same time, the consequences of these decisions are typically uncertain[179-181].

Within this context, it has been argued that there are two separate but related tasks that health care systems are called to fulfil. First, a decision is needed on which of alternative technologies should be adopted, taking into account existing evidence on their cost-effectiveness. Second, a separate decision is required on whether more information should be obtained through research to support the adoption-related decision[115;179].

The first task is accomplished with the help of economic evaluation studies, which indicate the optimal course of action (here, the optimal treatment adoption decision), given existing evidence[181;182]. If the objective of the decision-makers is to maximise health benefits out of the available resources, it has been argued, the choice to be made is the one with the greatest payoff given current information[106;115].

The second task is equally important. Decisions are needed on whether and to what extent resources should be devoted to further research, given that research provides valuable evidence towards ensuring that optimal adoption decisions are made. VoI can assist with this task by quantifying the potential benefits from pursuing further information through research and, on this basis, the framework has been often advocated as a useful tool for informing research funding decisions[80;109-112;179]. A discussion of main concepts in VoI is given below.

3.3.2. 'Value of information' concepts

'Value of information' analysis consists of a set of distinct, although closely associated, techniques. The rationale and aims of each of these techniques are given in turn below. A description of the techniques with a specific focus on methods and formulae for calculations is given in Chapter 7.

3.3.2.A. Expected value of perfect information

Existence of uncertainty implies that decisions which appear optimal in the light of current information may turn out to be 'wrong' when uncertainties resolve. For example, a decision to reimburse and provide a hypothetical treatment A which appears cost-effective compared to an alternative treatment B on the basis of existing (imperfect) evidence may prove wrong if treatment A is, in fact, suboptimal. In such a case there will be a cost associated with making the wrong decision (i.e. adopting treatment A), which can be measured in terms of forgone benefits due to providing the patients with a suboptimal treatment. Forgone benefits represent a cost to society and are seen as an opportunity loss (or opportunity cost)

due to making decisions under uncertainty. The expected opportunity loss can be quantified by weighting the probability of making the wrong decision by the opportunity loss due to making such a decision[97;115].

This expected opportunity loss can be minimised by conducting research and acquiring information that reduces uncertainty. In this way, the value of conducting further research can be seen as the reduction in the expected opportunity loss due to uncertainty. At the extreme, perfect information—that is, no uncertainty and perfect knowledge of which of treatments A and B is optimal—would eliminate the opportunity loss associated with uncertainty and ensure that the best treatment is adopted[97;115]. Given this, the expected value of acquiring perfect information (EVPI) is equal to the expected opportunity loss due to uncertainty, as perfect information would eliminate the cost of uncertainty altogether. In simple terms, the benefits from a decision made with perfect information are equal to the benefits one would be bound to forgo when making a decision under uncertain, imperfect information[97;107;108].

Expressed as net monetary benefits (NMBs), EVPI is simply the difference between the expected net benefits that would accrue from a decision with no uncertainty (perfect information) and the expected net benefits from the same decision, this time made on the basis of imperfect information. In this sense, the EVPI associated with an adoption decision represents the maximum expected benefits that could be ever realised by conducting research and informing the adoption decision. At first, EVPI is calculated for a single decision (i.e. an individual patient) but, as information can be disseminated freely, EVPI can be

extrapolated to the population of existing and future patients who are anticipated to be affected by the treatment decision, to give the population EVPI[51].

Expressing EVPI in monetary terms (such as NMBs) allows comparisons between the expected maximum benefits from research against the cost of research itself. If these gains exceed the expected costs of further research, conducting further research is potentially worthwhile. If, on the other hand, the gains are unlikely to exceed the cost of the research programme, conducting research would represent a waste of resources[115]. Given this, EVPI has been seen as the upper limit on the resources that a rational decision-maker should devote to acquiring further information through research and, in this sense, it consists the first criterion—or hurdle—for assessing whether carrying out evaluative research is justified[175;179].

3.3.2.B. Expected value of perfect information for parameters

In many situations, uncertainty around a treatment adoption decision is due to imperfect information around specific parameters affecting this decision (e.g. uncertain effectiveness of a treatment). In such cases, the EVPI can be extended to provide an estimate of the 'value of information' associated with conducting research on a specific parameter or groups of parameters.

The expected value of perfect information for specific uncertain parameters (i.e. the expected value of perfect parameter information (EVPPI)) is the difference between the expected benefits from a decision made with perfect information about specific uncertain parameter(s) and a decision made with current information about these parameters. The

obtained EVPPI is associated with an individual patient or decision and, as in the case of EVPI, it can be projected to the population affected by the decision to give the population EVPPI[183]. EVPPI can indicate parameters which contribute the most to decision uncertainty (e.g. quality of life) and, as a result, it can highlight the type of research that would be most beneficial (e.g. a non-experimental study to collect quality of life data)[184;185].

3.3.2.C. Value of information and implementation

EVPI and EVPPI results show the maximum benefits of pursuing better information, on the premise that any adoption decision made in the light of this information will be fully implemented. However, the implementation of recommendations is often imperfect and, in practice, the benefits from adopting the optimal treatment and making it available are unlikely to be realised in full[186]. This acknowledgement has led to the development of a conceptual framework which aims to estimate the value of undertaking strategies to enhance implementation and identify the situations where such strategies may be more valuable than pursuing better information[187;188].

This methodology, laid out by Fenwick and colleagues[187], evaluates the benefits expected to accrue from decisions made under different 'states of the world' regarding the availability of information about a treatment and the subsequent implementation, or uptake, of the treatment. Such states represent situations where decisions made under perfect information are followed by either perfect or 'current' (i.e. imperfect) implementation and states where decisions under current information are followed by perfect or current implementation. The framework aspires to address criticisms around the assumption typically used in VoI that is

that, once an adoption decision is made, this will be followed by perfect adherence[80;189]. Although the concept presents similarities with the 'payback' framework in that it aims to account for the impact of research in clinical practice, its developers have pointed out that they consider value of implementation as a distinct methodology[187].

3.3.2.D. Expected value of sample information

As explained above, EVPI and EVPPI represent the maximum benefits that could be realised from further research and establish a necessary condition: for a further study to be potentially worth conducting, the maximum benefits from research, expressed as EVPI and EVPPI, must outweigh the costs of conducting further research[97].

However, obtaining perfect information would require 'perfect' research, of an infinitely great sample size. This is, in practice, out of reach as typically further evidence is obtained from studies of limited samples. To decide whether a piece of evaluative research involving a sample n will be indeed worthwhile, one needs to estimate the actual benefits expected to arise from this specific study. Such an estimate is given by the expected value of sample information of a study with sample size n ($EVSI_n$)[115].

By analogy to EVPI, $EVSI_n$ shows the difference between the expected value (measured in NMBs) of a decision made under information from a study with sample n, and the value of a decision made with current information[190]. Population EVSI can be compared against the cost of conducting a study of sample size n. The difference between $EVSI_n$ and the cost of a study involving n patients is the Expected Net Benefit of Sampling ($ENBS_n$); a measure that represents the benefits expected to arise from a study of sample size n, net of the study's

cost[115;190]. In this way, the $ENBS_n$ is seen as the net payoff to a proposed study and it represents the 'sufficient' condition for conducting this study: if $ENBS_n$ is positive (i.e. EVSI exceeds the cost of research), further experimental research will be 'cost-effective'[115]. In addition, the $ENBS_n$ provides a condition for establishing a study's desirable design characteristics: the employed sample size and the allocation of participants between treatment arms will be optimal when the difference between the cost of a trial and EVSI is the greatest possible (or, equivalently, when $ENBS_n$ is maximum)[51;97;175].

3.3.3. Applications of 'value of information' in the literature

A search in the literature was undertaken to identify studies documenting practical applications of VoI in the area of health care. The review involved searches in major electronic bibliographic databases, additional searches to identify studies citing and cited in key studies, 'related articles' searches, and general searches on the internet (Appendix 2.C.)

Since the late 1990s VoI has been a well-researched topic, with a number of publications discussing the methods, proposing methodological advances or simply applying the method as an adjunct to economic evaluations existing in the literature. A review of the literature identified 128 VoI-related studies. The majority of these (77 studies) have a methodological focus, aiming to illustrate the method and describe methodological advances, while approximately one third of the identified studies (48) are applied economic evaluations incorporating EVPI and EVPPI calculations to underpin recommendations for future research. The majority of the identified literature is authored by researchers based in the UK, although a number of studies come from other countries, predominately the United States, the

Netherlands, Australia and Canada. A significant share of the papers is authored by a team at the University of York led by Professor K. Claxton, while other clusters of researchers publishing on VoI were identified at the Universities of Sheffield, Bristol, Oxford, Toronto and the Erasmus University in the Netherlands.

Apart from published articles, VoI applicable to health care decision-making is described in prominent textbooks on economic evaluation[31;51;139]. The approach has been also recommended as a systematic way to establish the value of further research in the latest guide to methods for technology assessment issued by NICE[114].

3.3.4. Critique of 'value of information' in the literature

The increasing attention given to 'value of information' has led researchers to highlight its strengths but also expose its limitations. A specific strength of the framework lies in its theoretical basis. VoI is part of statistical decision theory, a well-established methodology that makes use of Bayesian inference to determine the optimal course of action under uncertainty and the situations where further research is beneficial[55;180;189;191]. This is demonstrated in statements such as *"To our knowledge, this is the only method with a theoretically sound basis; therefore we foresee an important role for value of information analysis in guiding future research."*[192[p.424]]. In addition, the framework is compatible with the methods of economic evaluation, to the extent that it has been seen as a natural extension of these methods. Combined with the results of decision modelling, the approach has been suggested as a comprehensive framework for addressing treatment adoption and research funding decisions[110].

Moreover, the approach can be potentially useful in addressing both research funding and research design-related questions. As Briggs *et al*.[51] argue, EVPI, EVPPI and EVSI analyses can establish which research programmes should take priority, identify comparators and endpoints to be included in further research and indicate subgroups of patients for which research would be valuable. The approach has also been proposed as an alternative to traditional approaches for trial research design, being able to determine the sample size and allocation of patients between treatments that would maximise the returns to its cost[115;189].

Despite this, VoI methods are also computationally demanding, especially when carrying out EVSI computations, where the value of sampling information needs to be evaluated for each possible trial result and different sample sizes and allocation of participants[51]. In addition, as VoI results quantify the uncertainty around an adoption decision, inappropriate representation of this uncertainty and flaws in an economic evaluation will inevitably lead to inaccurate—or potentially misleading—results about the value of further research.

3.4. Main characteristics of 'payback of research' and 'value of information'

Although they represent distinct frameworks, 'payback' and VoI share similarities. Main characteristics of the frameworks, in terms of the principles underpinning them, rationale and methods, are summarised in Table 3.2.

The frameworks recognise that the aim of evaluative research is to provide evidence to inform decision making, and express the 'value' of a research programme in terms of the benefits expected to accrue to the population through improved decision making. For both

the frameworks, the ultimate aim is to obtain explicit numerical estimates of the expected benefits of research, which can be useful in informing research priority-setting.

Table 3.2: Table summarising main characteristics of 'payback of research' and 'value of information'

	Payback of research	Value of information
Rationale and aim	Calculation of the potential health benefits associated with proposed research as a means of assisting priority setting for research.	Calculation of the potential health benefits associated with proposed research as a means of assisting priority setting for research.
Underpinning methods	Cost-effectiveness analysis. Comparison of costs and benefits associated with two distinct 'states of the world' (i.e. with research taking place and without research).	Statistical decision theory. Comparison between the expected benefits of obtaining perfect or improved information and the maximum cost of acquiring perfect/improved information (i.e. cost of clinical trial).
Specification of possible outcomes of research	Explicitly specified (Delta results, exemplar outcomes).	Not explicitly specified, but possible outcomes are expressed through a large number of possible realisations of results produced through probabilistic sensitivity analysis.
Specification of change in clinical practice	Direction and magnitude of change in clinical practice is accounted for explicitly for each possible outcome.	Direction of expected change in clinical practice (i.e. adoption/rejection in light of different simulated results from probabilistic sensitivity analysis) accounted for implicitly. Perfect implementation of the most optimal strategy is assumed.
Estimation of costs and benefits	Estimation of cost and benefits associated with each possible outcome.	Estimation of costs and benefits for different possible outcomes (i.e. different possible realisation of results in probabilistic sensitivity analysis).
Calculation of results	Results are calculated as the difference between the costs and benefits (i.e. net benefits) between a state with research and a state without research	Results are calculated as the net monetary benefits expected to accrue due to reduction in, or elimination of, uncertainty
Decision rules	Value of a research proposal is inferred by the calculated 'payback' results. Positive net monetary benefits indicate that research is potentially beneficial.	Value of a research proposal is inferred by the calculated VoI measures (e.g. EVPI, EVSI, EVPPI) net of the cost of research. Positive net monetary benefits indicate that research is potentially beneficial.

Divergence in the approaches can be seen when looking at the methods used in predicting the expected benefits of research. In 'payback', possible research outcomes are specified

explicitly (i.e. Delta results[90] or exemplar outcomes[92]) and are the starting point for the analysis. Under the assumption that research will provide valid and unbiased estimates of treatments' effectiveness and cost-effectiveness, possible outcomes are used as a proxy for the true effectiveness and cost-effectiveness of the treatments of interest to the population. On the other hand, VoI methods do not explicitly specify possible research results, but, instead, express the possible effectiveness and cost-effectiveness of the compared treatments in the form of a large number of output results generated through probabilistic sensitivity analysis. Each of these results represents a possible realisation of treatments' cost-effectiveness and is obtained externally, through decision modelling, rather than internally, through the VoI analysis itself. In general, though, the fact that both frameworks account for possible results implicitly or explicitly, can be seen as a noteworthy commonality between the approaches.

A notable aspect in 'payback' is the specification of the direction and magnitude of change in clinical practice in light of each possible result. For example, the approach seeks to specify not only how the practice is expected to change (e.g. adoption of a treatment associated with 'favourable' outcomes) but it also requires determining explicitly the extent of change in the use of the treatment (i.e. treatment uptake). In VoI, direction of change in practice is implicitly specified by assuming that for a given realisation there will be a complete switch to the optimal treatment. Unlike 'payback', main VoI concepts assume that the chosen treatment will be implemented perfectly (i.e. all eligible patient will receive the treatment that appears optimal), although some allowance for this assumption is made in published work looking into the value of implementation[124](see Chapter 7).

The aspect where the methods differ the most is the calculation of final results. To obtain an estimate of the value of a proposed research programme, 'payback' methods typically compare the benefits accruing from a 'state of the world' where research has taken place, and a state where research has not been conducted. This comparison gives the gain (or loss) associated with research and it is interpreted as the value of undertaking the proposed study. The interpretation of VoI results is less intuitive; results show the expected gains in benefits due to reduction or elimination of uncertainty. VoI is typically expressed in terms of NMBs and it is compared against the cost of research itself. Decision rules for 'payback' and VoI stipulate that the positive net benefits indicate that carrying out a proposed piece of research is anticipated to be beneficial.

3.5. Chapter overview

The chapter aimed to provide a detailed description of the most prominent analytic frameworks—'value of information' and 'payback of research'. The frameworks have a common starting point: they are both underpinned by the notion that the aim of evaluative research, such as a clinical trial, is to provide evidence to inform better adoption decisions, and thus the desirability of conducting the trial can be inferred by the benefits expected from improved decision-making. Nonetheless, the frameworks present differences in the way they identify and measure those benefits. In 'payback' the value of the trial is inferred by the benefits expected to accrue from a change in practice induced by information generated from the study, while VoI values evaluative research according to the reduction in uncertainty—and, thus, the reduction in the expected loss of benefits—that the trial is expected to bring about. Despite these methodological differences, there exist conceptual

similarities in the approaches. Strengths and limitation of the frameworks, as seen in the literature, were also outlined.

This and the previous chapter sought to introduce the aims of the project, describe the context of research priority-setting and look into proposed approaches, with a focus on analytic approaches. The next part of the study describes the application of 'value of information' and 'payback of research' analysis to case studies.

PART II. Practical application of ‘payback’ and ‘value of information’

The first part (Chapters 1, 2 and 3) set the background and identified two main analytic frameworks proposed to assist with funding decisions in research—‘payback of research’ and ‘value of information’. This part, Part II, focuses on the practical application of these frameworks to case studies. The primary aim of this application is to provide an insight into the frameworks’ capabilities, strengths and limitations. Observations made in this part formed the basis for assessing the frameworks and drawing conclusions on their usefulness, practicality and potential value in the last part of the thesis (Part III).

Case studies used in this application represent proposals for clinical trials (BTOG-2 and TRAPEZE phase III trials) aiming to give primary evidence for specific treatment adoption decisions in non-small cell lung cancer (NSCLC) and hormone-refractory prostate cancer (HRPC). The case studies gave access to the trials’ protocols and, in the case of HRPC, provided patient-level data from the early stage TRAPEZE phase II trial.

The application of ‘payback’ and ‘value of information’ to the case studies was carried out iteratively, in a two-stage process. The first, preliminary stage required assessing the extent of the existing, pre-trial evidence around the treatment adoption decisions that each trial aimed to inform, while the second, main stage, involved using ‘payback’ (PATHS model) and ‘value of information’ (VoI) to determine whether obtaining further evidence through conducting the proposed BTOG-2 and TRAPEZE phase III trials would be beneficial.

For each case study, the first task involved specifying the relevant adoption-related decision (i.e. which of specific treatments to adopt and provide to patients) and research-related decision (i.e. whether evaluative research should be carried out to generate more evidence to facilitate the adoption-related decision). As a next step, the existing evidence around the adoption-related decisions was identified and synthesised through decision analytic models. Work undertaken in this stage for the NSCLC and HRPC case studies is detailed in Chapters 4 and 5, respectively.

In the main stage of the study, the focus turns on the research-related decisions, that is, whether funding and carrying out additional research in the form of the proposed BTOG-2 and TRAPEZE phase III trials would be worthwhile. For this, 'payback of research' and 'value of information' were applied to each case study with the help of the constructed models. The application of 'payback' to case studies involved specifying possible results that the proposed studies may show, translating these results into expected costs and health benefits through the decision models, evaluating the possible change in clinical practice in the light of these outcomes and estimating the additional benefits that would be expected to arise across the population. On the other hand, the application of VoI required translating the uncertainty around the adoption-related decisions into estimates of the expected benefits from conducting further research through different measures of perfect and sample information. 'Payback' and VoI analyses are reported in Chapters 6 and 7, respectively.

At the time that this doctoral project started (September 2009), both the BTOG-2 and TRAPEZE trials were funded and on-going. To account for this fact, the analyses were undertaken in a retrospective manner, by making use of evidence that would have been

available at the time the funding decisions were considered (2004 and 2006 for NSCLC and HRPC, respectively).

CHAPTER 4. Adoption-related decision in non-small cell lung cancer

This chapter describes the methods used in, and the results obtained from, a decision analytic modelling exercise aimed to summarise the current evidence around a treatment adoption-related decision in the area of non-small cell lung cancer (NSCLC). The developed model and the result of this work aimed to assist with the application of 'payback of research' and 'value of information' to the NSCLC case study. The first part of the chapter reports a review of the existing evidence on the effectiveness and cost-effectiveness of chemotherapy agents for NSCLC, while the second part focuses on the methods and the results of the model-based economic evaluation.

4.1. Background

Lung cancer is one of the most common types of cancer and the leading cause of cancer-related deaths worldwide and in the UK[193-196]. According to the World Health Organization[197], in 2004, lung cancer was responsible for the death of approximately 1.3 million people around the world. In the UK, using 2008 data, Cancer Research UK estimated that lung cancer is responsible for approximately 35,300 deaths per year and accounts for 24 percent and 21 percent of all male and female cancer deaths in the country, respectively[194] (Figure 4.1).

A number of risk factors appear to contribute to lung cancer, most importantly tobacco smoking, exposure to carcinogens and air pollution[198;199]. Lung cancers are classified into two

types: non-small cell lung cancer (NSCLC), which includes squamous cell carcinoma, adenocarcinoma and large cell carcinoma, and small cell lung cancer[198]. Non-small cell lung cancer is the predominant type of lung cancer, accounting for approximately 80 percent of all lung cancer cases[200;201].

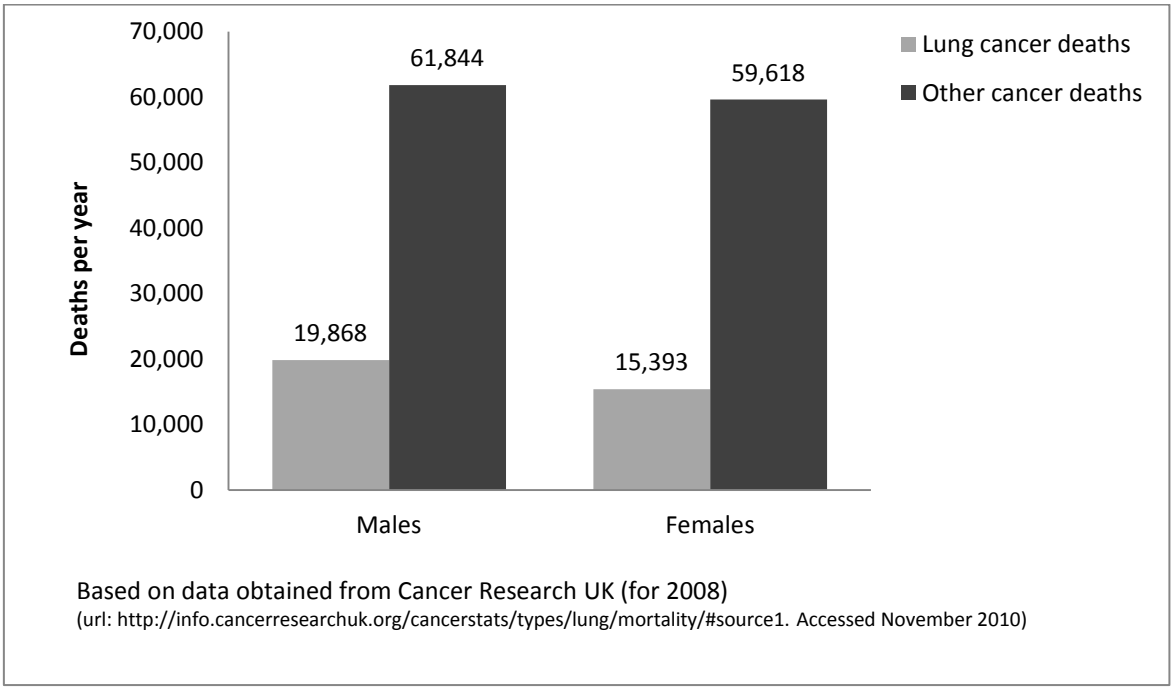

Figure 4.1: Number of lung-cancer-related and other cancer deaths in the UK

Only a small proportion (about 20 percent) of NSCLC cases are potentially curable and these are usually patients with non-advanced disease eligible for curative surgery[202]. Thus, patients with advanced disease (i.e. NSCLC stage IIIB where cancer has spread to the mediastinum (i.e. heart, oesophagus and great vessels) and NSLCC stage IV where distant metastases are present[203]) are incurable at presentation and candidates for treatments that extend life or palliate symptoms. Common treatment options for this patient group include 'best

supportive care', which aims to control symptoms and usually comprises radiotherapy and non-cytotoxic drugs, or chemotherapy, which is usually given to extend survival[204;205].

A number of chemotherapy regimens are available, including paclitaxel, docetaxel and gemcitabine; the last is known to offer a valuable, though modest, increase in life expectancy and quality of life when compared to radiotherapy alone or best supportive care[198;202;206]. Gemcitabine (Gem) is considered the bedrock of chemotherapy and it is typically used in combination with platinum analogues which aim to trigger cell death. The combination of gemcitabine with a particular platinum analogue, cisplatin (Cisp) has been shown to prolong survival in advanced cancer patients and is recommended by the National Institute for Health and Clinical Excellence (NICE)[198;207] as the standard first-line treatment for advanced NSCLC.

Apart from cisplatin, gemcitabine is routinely given with carboplatin (Carb), another platinum analogue which acts similarly to cisplatin but it is considered to be less toxic and more convenient to administer[208]. As gemcitabine plus carboplatin (Gem+Carb) may also be beneficial, there is uncertainty around the optimal first-line treatment for the specific patient group[208;209].

4.2. Adoption-related and research-related decisions

Given the above, the relevant adoption-related question relates to which of Gem+Carb or Gem+Cisp should be adopted and provided to patients with advanced NSCLC. In turn, stemming from this, the relevant research-related question is whether public funds should

be committed to a clinical trial aiming to provide additional evidence on the effectiveness and costs associated with Gem+Cisp and Gem+Carb.

A phase III randomised controlled trial, the British Thoracic Oncology Group 2 (BTOG-2) trial, was proposed to provide primary evidence on the effectiveness and costs of chemotherapy treatments for NSCLC. The proposal was submitted for funding to the CRUK Phase III Clinical Trial Grants stream, secured funding in 2004 and started recruitment in 2005. Results of the trial were expected to become available in late 2011. The cost of the trial, as given in the research proposal, was £336,721 (in 2004 prices). This includes research-related costs (i.e. cost of researchers' salaries, facilities and indirect costs) of £134,221 and service support costs (i.e. cost related to obtaining informed consent, additional clinic time and collection of quality of life questionnaires) of £202,500. Since the investigated treatments are routinely available in clinical practice, no funds for excess treatment costs were requested.

4.3. Existing clinical and cost-effectiveness evidence

A systematic review was undertaken to summarise the existing clinical and economic evidence, with the additional aim of retrieving information that could be used in the planned decision model. The review involved searches in bibliographic databases (MEDLINE, NHS CRD, ISI Web of Knowledge), searches in relevant portals (NICE Clinical Guidelines and Cochrane Collaboration Reviews), general searches on the internet through Google Scholar®, as well as searches in the reference lists of identified articles. Details on searches and retrieved articles are given in Appendix 3.A. Identified studies were included if they:

a. focused on patients with advanced NSCLC (stage IIIB or IV) eligible for chemotherapy;

b. involved an assessment of gemcitabine, cisplatin or carboplatin alone or in combination with other chemotherapeutic agents as first-line treatment for advanced NSCLC;

c. reported cost-effectiveness results, or clinical effectiveness results on at least one of the following outcomes: tumour response, survival, time-to-progressive disease, quality of life, toxicity, treatment discontinuation rates.

In order to use only evidence existing up to the point when the research-related decision was considered, the review included studies available before July 2004. Twenty-one studies were included in this review. Of these, thirteen assessed treatments in terms of their clinical effectiveness, seven were economic analyses, while one study[201] looked at both the clinical effectiveness and cost-effectiveness of first-line chemotherapy treatments for NSCLC.

4.3.1. Clinical effectiveness evidence

Fourteen publications assessing the clinical effectiveness of NSCLC treatments were found. Of these studies, thirteen reported the results of single trials, while the study by Clegg *et al.*[201] summarised the results of multiple studies. Most of the identified clinical studies compared either gemcitabine plus cisplatin (Gem+Cisp) or gemcitabine plus carboplatin (Gem+Carb) against different combinations involving etoposide, mitomycin, vinblastine, vinorelbine, paclitaxel and ifosfamide, with or without cisplatin or carboplatin. A summary of the identified studies is given in Appendix 3.B, Table 3.b. In these studies, Gem+Cisp and

Gem+Carb appeared, in general, more effective than their comparators, resulting in small gains in survival and quality of life, but showing no marked difference in toxicity and rates of adverse events.

A head-to-head comparisons of Gem+Cisp against Gem+Carb for NSCLC patients was reported in one study, by Zatloukal and colleagues[208]. This study[208] gave the results of a phase III trial involving 89 and 87 patients in the Gem+Cisp and Gem+Carb arms. Findings showed the treatments to be of comparable effectiveness (in terms of response and overall survival) and toxicity; thus the authors concluded that Gem+Carb may be a sound alternative to Gem+Cisp, especially for patients unable to receive cisplatin. The main characteristics and conclusions of the Zatloukal *et al.*[208] study are given in Table 4.1.

Table 4.1: Summary of main effectiveness studies (NSCLC)

Study	Study type	Intervention	Comparator	Outcome measures	Authors' conclusions
Zatloukal *et al.* (2003)[208]	Phase III RCT	Gem+Cisp (89 patients in arm)	Gem+Carb (87 patients in arm)	Response Survival Toxicity Time-to-progression	Both Gem+Cisp and Gem+Carb are effective and comparable in efficacy and toxicity Gem+Carb may be acceptable for patients who cannot receive Gem+Cisp.

4.3.2. Cost-effectiveness evidence

Eight studies (including the study by Clegg *et al.*[201]) looking at the costs or cost-effectiveness of Gem+Cisp and/or Gem+Carb were identified. Of these, two were cost-minimisation analyses[210;211], one was a cost analysis[212], three cost-effectiveness analyses[201;213;214] and two were reviews of existing economic studies[215;216]. The majority of these studies assessed

Gem+Cisp or Gem+Carb against other chemotherapeutic regimens. The main characteristics of the identified economic studies are given in Appendix 3.B, Table 3.c.

Only one study reported a direct comparison of Gem+Cisp against Gem+Carb was found. This was a cost-minimisation analysis carried out in the United States by Khan and colleagues[210]. Assuming equivalent effectiveness across treatments, the authors investigated the costs per patient and cost per treatment course for cisplatin and carboplatin in combination with different agents, including gemcitabine, in the areas of non-small cell lung cancer, small cell lung cancer and ovarian cancer. The study results suggested that, in NSCLC, the use of carboplatin was associated with an extra US$9200 (£7180 in 2004 prices) and US$2100 (£1640 in 2004 prices) per patient and treatment course respectively[210]. A summary of the study by Khan *et al.*[210] is given in Table 4.2.

Table 4.2: Summary of main economic studies (NSCLC)

Study	Study type	Compared treatments	Authors' conclusion
Khan *et al*. (1999)[210]	Cost-minimisation analysis	Cisplatin in combination with other agents (gemcitabine, paclitaxel, vinorelbine) Carboplatin in combination other agents (gemcitabine, paclitaxel, vinorelbine)	In NSCLC, carboplatin-based regimens appear more costly than cisplatin-based regimens

4.4. Decision modelling for NSCLC

A decision model was built with the aim to summarise the existing evidence and provide a comprehensive assessment of the effectiveness and cost-effectiveness of Gem+Cisp and Gem+Carb. The methods for constructing and populating the model are given below.

4.4.1. Model structure

The model followed the progression of a cohort of 1000 NSCLC patients through three health states:

a. 'Progression-free' (PGF), where patient remain on stable disease with no signs of disease progression;

b. 'Progression' (PG), where disease has progressed, and

c. 'Death' (D).

The cohort enters the model at the 'Progression-free' state, where patients are scheduled to receive a four-cycle course of chemotherapy, with each cycle lasting 21 days. Patients stay in this state until they experience disease progression. Patients who experience disease progression, either during or after the four-cycle treatment period, move to 'Progression' and eventually to 'Death'. Owing to the advanced stage of the disease and the short time period between the onset of advanced disease and death, the majority of the deaths are due to the disease[208;217]. Thus, moves from the 'Progression-free' state to 'Death' were not modelled. A graphical representation of the decision model can be seen in Figure 4.2.

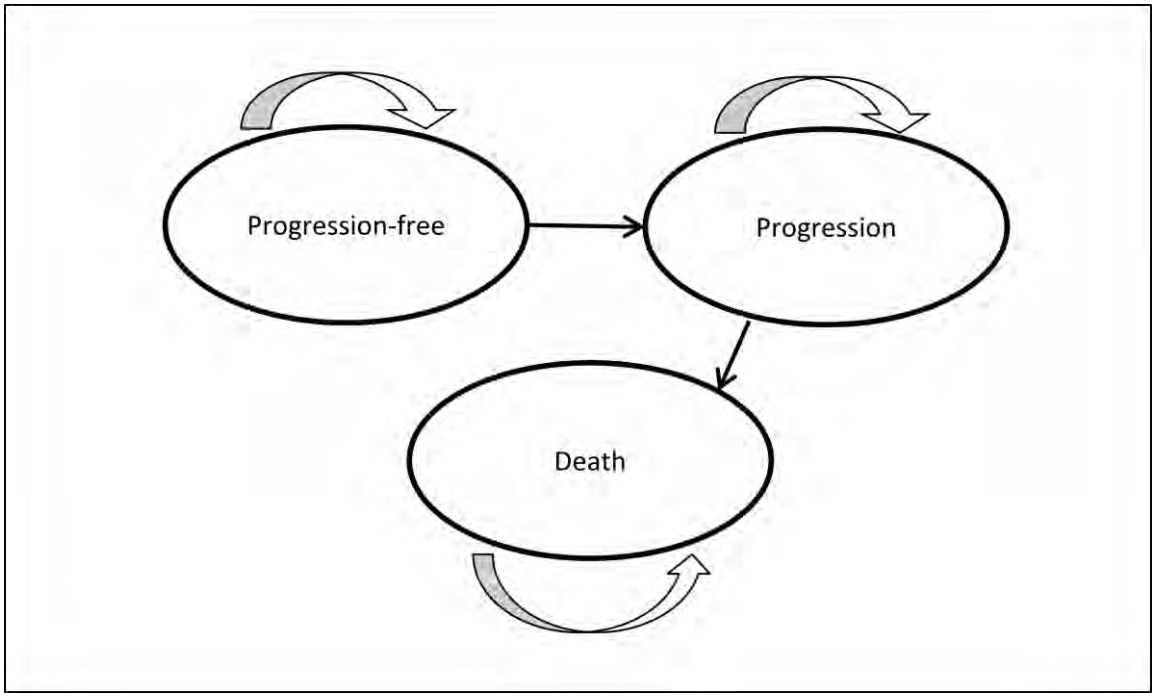

Figure 4.2: Model structure for NSCLC

The length of a model cycle was set to 21 days, equal to the treatment cycle. The cohort was followed for 3 years (53 cycles), by which time the vast majority of the patients are expected to have died.

4.4.2. Model input

Input parameters for the model were obtained from the published literature. When published evidence was not available, relevant information was obtained from expert opinion. Parameters and values used in the model, in terms of disease progression, costs and quality of life are described below.

4.4.2.A. Disease progression

Information on disease progression, in the form of Kaplan-Meier survival and time-to-progression curves for each treatment was obtained from the only study comparing Gem+Cisp against Gem+Carb[208]. These curves formed the basis for fitting Weibull models, which were used to derive estimates of the probability of a patient being at in any of the three model states at any given cycle. Weibull models are commonly used to describe the distribution of time-dependent events such as cancer survival and progression and are characterised by two parameters: the shape parameter alpha (α) and the scale parameter beta (β)[218;219].

In fitting a Weibull model to existing survival curves, the task is to determine the values of the parameters α and β. This can be done by manipulating the survival function $S(t)$ of the Weibull distribution

$$S(t) = exp\left[-\left(\frac{t}{\beta}\right)^{\alpha}\right]$$

where t represents time (here, in discrete 21-day cycles) and α, β represent the parameters of the Weibull model, to give

$$ln[-lnS(t)] = \alpha\, ln(t) - \alpha\, ln\, \beta$$

This expression shows that the $ln[-ln\, S(t)]$ is a linear function of $ln(t)$. Regressing $ln[-ln\, S(t)]$ against $ln(t)$ gives ordinary least square estimates of the model intercept and coefficients, which can be used to obtain the α and β parameters for the Weibull model, as

$$\alpha = regression\ coefficient\ of\ ln(t)\ and$$

$$\beta = exp\ (-model\ intercept\ /\ regression\ coefficient\ of\ ln(t))$$

Once the parameters of the Weibull models for disease progression and death were estimated, the resulting progression-free survival and survival functions were used to calculate the probability of a patient staying at the 'Progression-free' state, being at the 'Progression' state or being dead. This gave the number of patients in the 'Progression-free' and 'Death' states throughout the model. In turn, the number of patients in the 'Progression' state was calculated as the difference between the total number of patients in the cohort (i.e. 1000) and the sum of the patients in the remaining states (i.e. Patients in the 'Progression' state = total number of patients in the cohort – (patients in 'Progression-free' state + patients in 'Death' state). For example, at cycle 10, 468 out of 1000 patients had not progressed and were in the 'Progression-free' state and 400 patients had died, leaving 132 patients in the 'Progression' state.

In the first two cycles for Gem+Cisp, the calculated sum of patients in the 'Progression-free' and 'Death' states exceeded the number of patients in the cohort (by ten patients in cycle 1 and six patients in cycle 2). This resulted in observing negative numbers of patients in 'Progression' state at cycles 1 and 2. In order to correct for this minor inconsistency, the number of patients in each of the states at cycles 1 and 2 were re-calculated by scaling the values down to 1000. Tables with the output of the regression model for disease progression and survival, the estimated α and β parameters as well as graphs of the fitted Weibull

models for each treatment are given in Appendix 3.C, Table 3.d to Table 3.k and Figure 3.b to Figure 3.e.

4.4.2.B. Resource use and costs

Costs associated with each treatment were estimated according to use of health care resource due to:

a. drug acquisition and administration;

b. adverse events;

c. other medical resources, and

d. terminal care.

In estimating these costs, health care resource use derived from published studies was multiplied by unit cost estimates taken from national published sources[220-222]. Costs were converted to 2004 prices using the Hospital and Community Health Services pay and price inflation indices[220], to reflect the relevant values in the year when the research funding decision was considered.

Drug acquisition costs were calculated according to the standard treatment schedule of two administrations of gemcitabine on days 1 and 8 of a 21-day cycle, and one administration of platinum analogue (either cisplatin or carboplatin) on day 1 of the treatment cycle[198]. For these calculations, chemotherapy doses were multiplied by unit costs of drugs published in the British National Formulary[222]. The body surface area of an 'average' NSCLC patient (1.7

square metres) was taken from the literature[201]. Required doses, constituent parts and unit costs can be found in Table 4.3.

Table 4.3: Unit costs of drug acquisition and administration (NSCLC)

Resource	Dose (assuming body surface area 1.7 square metres)	Cost (2004 prices)	Source
Gemcitabine (1250mg/m^2)	2125mg	£295	British National Formulary[222]
Cisplatin 80mg/m^2	136mg	£61	
Carboplatin AUC 6	680mg	£252	
Dexomethasone	8mg	£4	
Mannitol	200ml	£2	
Chemotherapy administration (outpatient)	£142		NHS Reference Cost Schedules 2009-2010[221]

According to the literature, chemotherapy administrations for NSCLC typically take place in an outpatient setting[213;223;224]. Outpatient chemotherapy administration cost was obtained from the NHS Reference Costs Schedules[221]. Different possibilities regarding the split between patients receiving chemotherapy in an inpatient and outpatient setting were explored in sensitivity analyses. The total cost of drug acquisition and administration per treatment cycle for Gem+Cisp and Gem+Carb are given in Table 4.4.

Table 4.4: Cost of drug acquisition and administration for a treatment cycle (NSCLC)

Treatment	Day 1 of treatment cycle				Day 8 of treatment cycle			Total cost per treatment cycle
	Gem	Platinum analogue	Supple-mentary agents	Admini-stration	Gem	Supple-mentary agents	Admini-stration	
Gem+Cisp	£295	£61	£7	£142	£295	£4	£142	£946
Gem+Carb	£295	£252	£2	£142	£295	£4	£142	£1133

Separate calculations were carried out to obtain estimates of the expected cost of adverse events. In line with the literature, the focus was on significant toxicities which would typically lead to hospitalisation[214;224]. The probability of a patient in the cohort experiencing a serious adverse event was estimated from data on adverse event occurrence reported in the Zatloukal *et al.*[208] study. This probability was combined with the unit cost of resolving each type of adverse event taken from the NHS Reference Costs Schedules[221], using the formula:

$$Expected\ cost\ of\ adverse\ events_k = \sum P(AE_{i,k}) \times UC_i$$

Here, $P(AE_{i,k})$ is the probability of a patient on treatment k experiencing adverse event i and UC_i is the unit cost for resolving the adverse event i. An alternative, fixed value for the cost of adverse events (£544 in 2004 prices) obtained from the literature[201] was used in deterministic sensitivity analysis. The expected cost of experiencing an adverse event is given in Table 4.5.

Table 4.5: Expected cost of adverse events (NSCLC)

Adverse event	Unit cost of resolving adverse event	Gem+Cisp		Gem + Carb	
		Proportion of patients with adverse events*	Expected cost	Proportion of patients with adverse events*	Expected cost
Anaemia	£465	0.126	£59	0.18	£84
Thrombocytopenia	£446	0.164	£73	0.326	£145
Neutropenia	£390	0.095	£37	0.146	£57
Granulocytopenia	£390	0.23	£90	0.303	£118
Total expected cost	-	**£258**		**£404**	

*** Data from Zatloukal *et al.*[208]**

An estimate of the expected cost of other medical resources (additional outpatient visits and examinations) associated with Gem+Cisp was taken from Schiller and colleagues[212] (£728 in 2004 prices). In the absence of estimates of other medical cost specific to Gem+Carb, and in view of the fact that such costs are not expected to differ significantly between treatments[225], the above value was used for both treatments. Last, an estimate of the costs associated with terminal care for cancer patients was obtained from Clegg *et al*.[201] (£1460 in 2004 prices).

4.4.2.C. Health-related quality of life

Although estimates of health-related quality of life (HRQoL) were reported in most of the trials identified in the search for effectiveness and cost-effectiveness evidence, these were non-preference-based measures, such as the EORTC-QLQ C30 and LC13[206;226-231]. In view of the fact that such instruments have limited applicability to economic evaluations[130;232], a separate search for preference-based quality of life values was carried out. No preference-based quality of life (utility) scores for advanced NSCLC were identified in the pre-2004 literature. Therefore, utility scores were obtained from expert opinion (Professor L. Billingham, Professor of Biostatistics, University of Birmingham, 10-05-2011). The employed utility scores are given in Table 4.6.

Table 4.6: Preference-based health-related quality of life scores by health state (NSCLC)

Health state	Mean	Standard error	Source
Progression-free	0.65	0.08	Expert opinion (Professor L. Billingham, University of Birmingham)
Progression	0.45	-	
Utility increment (i.e. Difference between 'Progression-free' and 'Progression' scores)	0.20	0.04	

4.4.3. Analysis of uncertainty

Uncertainty in the model was propagated through probabilistic sensitivity analysis. Key parameters were represented by probability distributions, from which 5000 sets of values were drawn through Monte Carlo simulations[169;171] to give 5000 estimates of the costs and effects associated with each treatment.

As mentioned earlier, transition probabilities for the model were drawn from fitted Weibull time-to-progression and survival curves. Each curve is characterised by a shape parameter a and a scale parameter β, which are derived from the intercepts and coefficients obtained from the fitted linear regression model. Thus varying the transition probabilities required varying a and β, through varying the coefficients of the linear regression model. The latter were given normal distributions with mean and standard errors taken directly from the linear model. Parameter values and assigned distributions are given in Appendix 3.D, Table 3.I.

Cost parameters were also varied in probabilistic sensitivity analysis. Drug acquisition and administration costs were assigned gamma distributions using the method of moments[51]. Owing to lack of information on the standard error of the cost of drug acquisition and

administration, and to avoid holding this cost fixed, the standard error was set at ten percent of the mean value. This value was chosen on the grounds that drug acquisition and administration costs are expected to be uniform across patients and are relatively well established in the British National Formulary[222] and the NHS Reference Costs Schedules[221]. Making the cost of adverse events probabilistic involved varying the probability of a patient experiencing an adverse event, by fitting beta distributions to proportions of patients with different adverse events. Costs of other medical resources and terminal care were assigned gamma distributions. As no estimates of the variability around the latter values were available, a standard error of 25 percent of the mean value was used to reflect the greater variability in the use of other medical resources between patients. Details on assigned distributions and sources of employed values are given in Appendix 3.D, Table 3.m.

Preference-based quality of life (utility) values were also subject to probabilistic sensitivity analysis, by varying the utility values for the 'Progression-free' and the difference in utility (i.e. utility increment) between the 'Progression-free' and 'Progression' states. The varied values (i.e. utility of 'Progression-free' state and utility increment) were assigned normal distributions, with estimates of means and standard errors obtained from expert opinion (Professor L. Billingham, Professor of Biostatistics, University of Birmingham, 10-05-2011). In this way, the utility score for the 'Progression' state was varied by varying the utility increment, which ensured that randomly drawn utility values for this state in probabilistic sensitivity analysis are not higher than those of the less severe 'Progression-free' state. A table with utility scores and assigned distributions is given in Appendix 3.D, Table 3.n.

In addition to probabilistic sensitivity analysis, a series of deterministic analyses were carried out, where input parameters were given different plausible values. Parameters targeted in deterministic sensitivity analysis were the discount rate, the average body surface area, the split between patients receiving chemotherapy in an outpatient and inpatient setting, and the expected cost of adverse events.

4.4.4. Cost-effectiveness analysis

The analysis adopted the perspective of the National Health Service (NHS). In agreement with existing recommendations[114], both costs and health benefits (QALYs and life-years gained (LYG)) were discounted at an annual rate of 3.5 percent per year. The model was built and analysed in Microsoft Excel® (Microsoft Corporation, v.2007, Redmond, Washington, US) while STATA® (StataCorp, v. 11, College Station, Texas, US) and Parameter Solver (MD. Anderson Cancer Centre, University of Texas, Houston, Texas, US) were used for statistical analysis and distribution fitting tasks.

4.5. Results

Cost-effectiveness results, in terms of cost per QALY and cost per LYG, were obtained from 5000 Monte Carlo simulations of the probabilistic model. These were expressed as point estimate ICERs and were presented graphically in cost-effectiveness planes (CE planes)[173]. Results were also plotted as cost-effectiveness acceptability curves (CEACs)[174;175] and cost-effectiveness acceptability frontiers (CEAFs)[176].

4.5.1. Point estimate results

Cost-effectiveness results of the model, expressed in terms of point estimate ICERs, are shown in Table 4.7. Gem+Cisp appeared less costly than Gem+Carb, resulting in cost savings of approximately £740. This difference is mainly due to higher Gem+Carb costs of drug acquisition and administration and higher expected costs of adverse events. In terms of QALYs, Gem+Cisp appeared to be slightly more effective than Gem+Carb, resulting in a gain of 0.015 QALYs. As Gem+Cisp is less costly and it is associated with additional QALYs, this treatment dominates Gem+Carb. In terms of cost per life-years gained, Gem+Carb appears to be marginally superior to Gem+Cisp, offering an additional 0.016 LYG (approximately 6 days of additional survival). Given this, the cost per LYG of Gem+Carb compared to Gem+Cisp is approximately £45,000.

Table 4.7: Point estimate cost-effectiveness results (NSCLC)

Treatment	Total cost	Total QALYs	Total LYG	ICER (£ per QALY)	ICER (£ per LYG)
Gem+Cisp	£5830	0.583	1.000	-	-
Gem+Carb	£6568	0.568	1.016	Dominated	£45,030

4.5.2. Results of deterministic sensitivity analyses

Deterministic sensitivity analyses were carried out to assess the impact of different assumptions on the cost-effectiveness results (Table 4.8). Alternative assumptions about a patients' average body surface area and setting of chemotherapy administration had a minimal impact on the acquisition and administration costs of Gem+Cisp and Gem+Carb, and did not change the overall conclusion. Similarly, using alternative estimates of the cost

associated with adverse events inflated the total cost of both treatments by roughly the same amount, which had little impact on the difference in cost between Gem+Cisp and Gem+Carb. Last, different discount rates had a small impact on costs and benefits of Gem+Cisp or Gem+Carb, which did not affect the general conclusion of Gem+Carb being more costly and less effective than Gem+Cisp.

Table 4.8: Results of deterministic sensitivity analysis (NSCLC)

Parameter	Gem+Cisp		Gem+Carb		ICER (£ per QALY)
	Total costs	Total QALYs	Total costs (£)	Total QALYs	
Base case	£5830	0.583	£6568	0.568	Gem+Carb dominated
Average body surface area 1.6m^2	£5760	0.583	£6482	0.568	Gem+Carb dominated
Average body surface area 1.8m^2	£5835	0.583	£6590	0.568	Gem+Carb dominated
Chemotherapy administration- 90% outpatient, 10% inpatient	£5844	0.583	£6579	0.568	Gem+Carb dominated
Chemotherapy administration- 80% outpatient, 20% inpatient	£5857	0.583	£6590	0.568	Gem+Carb dominated
Chemotherapy administration- 70% outpatient, 30% inpatient	£5871	0.583	£6601	0.568	Gem+Carb dominated
Chemotherapy administration- 60% outpatient, 40% inpatient	£5884	0.583	£6611	0.568	Gem+Carb dominated
Chemotherapy administration- 50% outpatient, 50% inpatient	£5898	0.583	£6622	0.568	Gem+Carb dominated
Expected cost of adverse events at £544 as in Clegg *et al*. (2001)[201]	£6169	0.583	£6731	0.568	Gem+Carb dominated
No discounting	£5877	0.598	£6618	0.582	Gem+Carb dominated
Discount rates- 6% p.a. for both costs and benefits	£5799	0.573	£6534	0.559	Gem+Carb dominated

4.5.3. Results of the probabilistic sensitivity analysis

Probabilistic sensitivity analysis results, in the form of estimated cost and QALY pairs generated through 5000 Monte Carlo simulations, are plotted in the CE plane shown in

Figure 4.3. In general, the generated estimates—appearing as points on the plane—are scattered across all four quadrants and appear clustered around the origin.

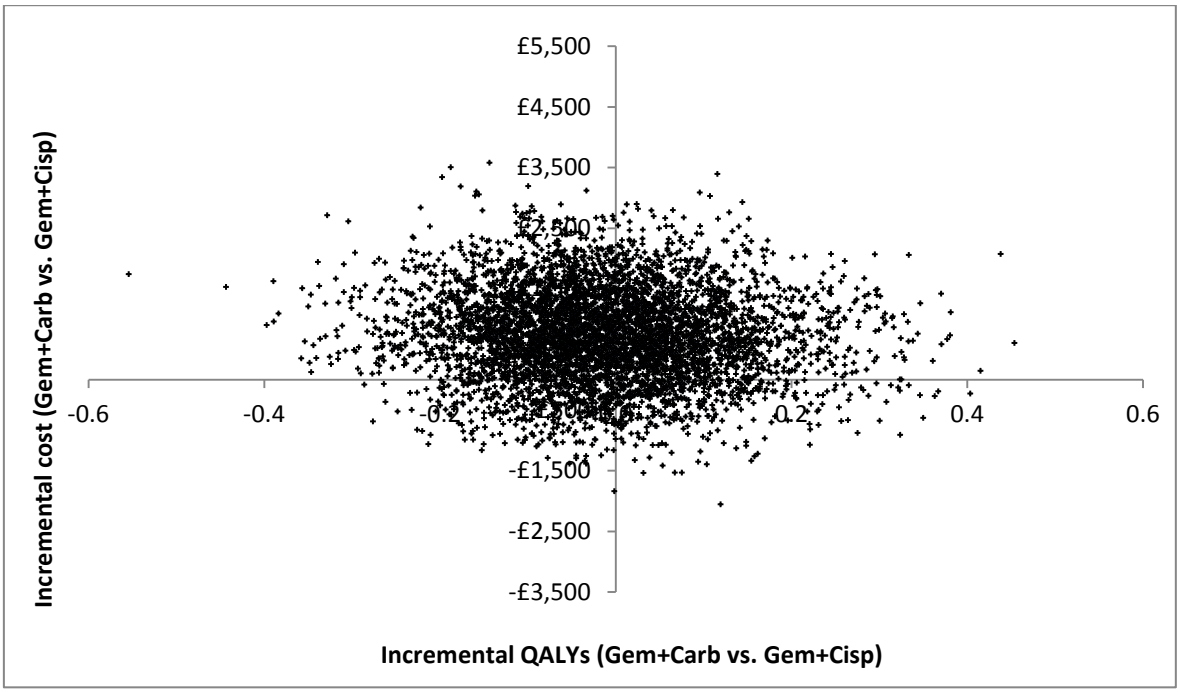

Figure 4.3: Cost-effectiveness plane for comparison between Gem+Carb and Gem+Cisp (NSCLC)

Proportions of the 5000 points on each of the quadrants in the plane are given in Table 4.9. Almost half of the points are located on the North West quadrant indicating that Gem+Carb is likely to be more costly and less effective than Gem+Cisp. Approximately one-third of the points are in the North East quadrant, suggesting that Gem+Carb may be more effective and more costly than Gem+Cisp. The remaining points are approximately split between the South East and South West quadrants.

Table 4.9: Proportion of incremental cost and QALY pairs on cost-effectiveness plane (NSCLC)

Quadrant	Proportion of points
North East	0.35
South East	0.08
South West	0.09
North West	0.48

The CEACs for Gem+Cisp and Gem+Carb are given in Figure 4.4. In a situation where the decision-maker is not prepared to pay any amount for additional health benefits (i.e. the ceiling ratio is zero), the probabilities of Gem+Cisp and Gem+Carb being cost-effective are 0.83 and 0.17, respectively.

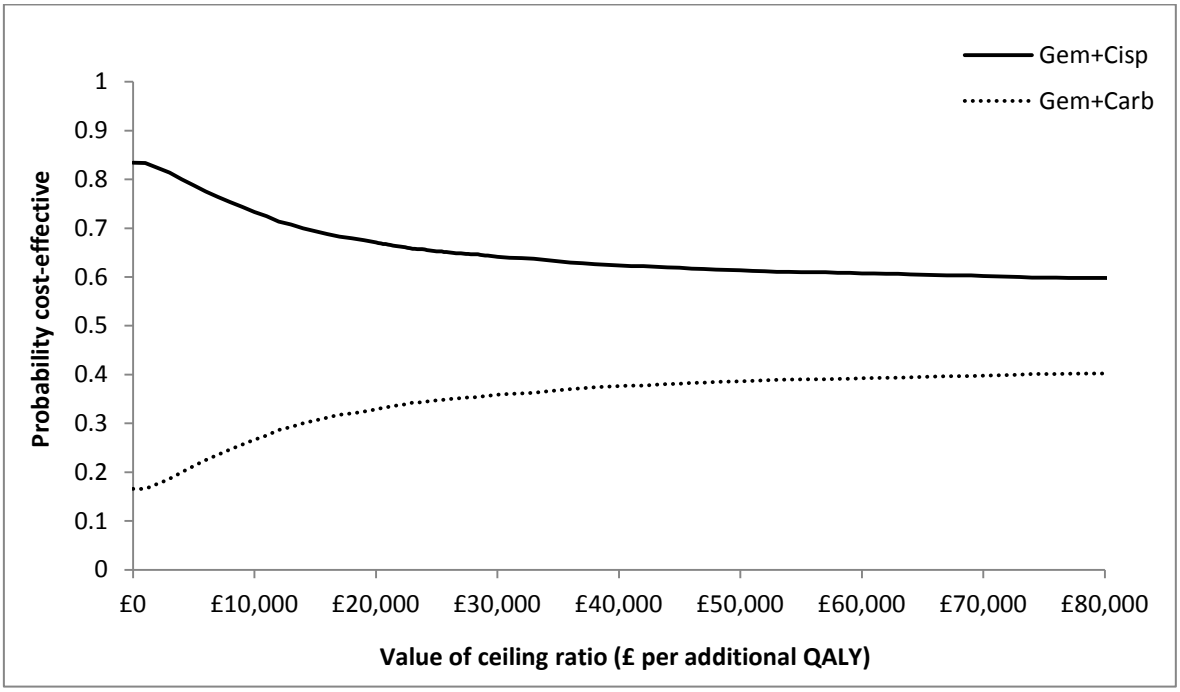

Figure 4.4: Cost-effectiveness acceptability curves (NSCLC)

Assuming a ceiling ratio of £30,000 per QALY gained, the probability that Gem+Cisp is more cost-effective than Gem+Carb is approximately 0.64. In other words, if society is willing to pay £30,000 for an additional QALY, there is a 64 percent chance that Gem+Cisp is the optimal treatment. At a high ceiling ratio of £80,000, the probability of Gem+Cisp being the most cost-effective option is approximately 0.60. As Gem+Carb does not result in the highest average net monetary benefits (NMBs) for values of the ceiling ratio at least as high as about £250,000 (not shown in this graph), the relevant CEAF corresponds to the CEAC for Gem+Cisp (Figure 4.5).

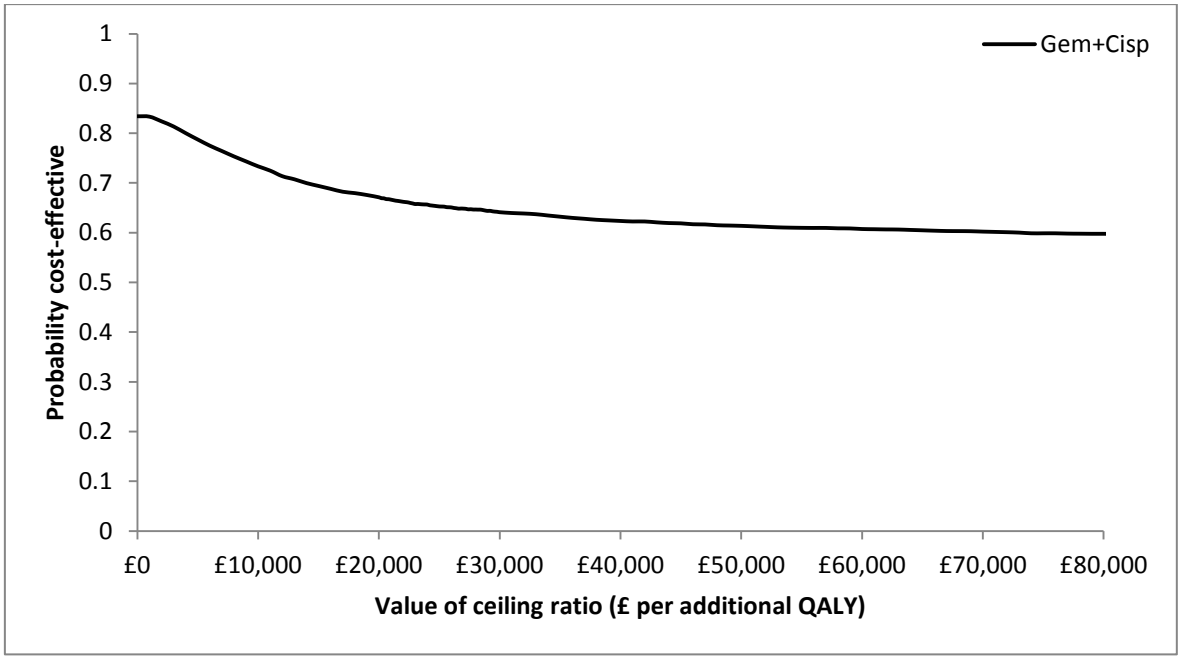

Figure 4.5: Cost-effectiveness acceptability frontier (NSCLC)

4.6. Discussion

The chapter describes an economic evaluation of Gem+Cisp and Gem+Carb as treatment options for NSCLC. In summary, results suggest that Gem+Cisp is less costly and more

effective than Gem+Carb; nonetheless, there is uncertainty surrounding the comparison, with the probability that Gem+Cisp is the preferred option at £30,000 per QALY being estimated at 0.64.

As explained earlier, the purpose of this analysis is to assess the treatment adoption-related decision associated with this case study and to provide an analytic structure that can be used as a vehicle for the application of the analytic approaches to the NSCLC case study. Although this analysis was carried out for the purposes of the present doctoral project and does not aim to inform 'real world' treatment adoption decisions, care has been taken to ensure that the employed methods are appropriate and justifiable.

Given this, the study presents certain strengths. In particular, the cost-effectiveness analysis makes use of a decision analytic model that traces the progression of NSCLC patients from a progression-free state to death. Key parameters used in the model, such as rates of disease progression, were identified through systematic searches in major bibliographic databases and were extrapolated using well-established methods[232;233]. In addition, the uncertainty around model parameters was accounted for by conducting probabilistic and deterministic sensitivity analyses. In particular, distributions representing the likelihood of each parameter to assume certain values were attached to all key parameters in the analysis and, when alternative values were found, these were also explored in deterministic sensitivity analyses. Last, results were presented as point-estimates and were plotted in CEACs and CEAFs, which show the probability of treatments being cost-effective and resulting in the highest net monetary benefits at different values of the ceiling ratio .

Despite this, the analysis has a number of limitations. First, a more comprehensive model which would have taken into account treatment discontinuation might have offered better accuracy, nonetheless no evidence on the prognosis of people who discontinue treatment (i.e. transition probabilities from a discontinuation state to progression and death) was found, and using assumptions for such parameters would introduce considerable bias. Second, survival and progression data were taken from a relatively small study[208]. However, this was the only study reporting a head-to-head comparison between Gem+Cisp and Gem+Carb identified in the literature. In some cases, no estimates of the variance around specific parameters such as cost of drug acquisition and administration and cost of adverse events were available. In order to account for the uncertainty around such estimates, assumptions had to be made about their variance. This was deemed preferable to holding the parameters fixed at their mean values. Last, no estimates of utility scores for the specific patient group were found in the pre-2004 literature and thus values were obtained from expert opinion.

4.7. Chapter overview

The chapter reports a model-based analysis aimed to establish the effectiveness and cost-effectiveness of Gem+Carb and Gem+Cisp for NSCLC given the information existing prior to a decision on funding a phase III trial on these treatments. In doing so, costs and QALYs associated with each treatment were estimated through a three-state decision analytic model. Overall, Gem+Cisp appeared superior to Gem+Carb, being less costly and resulting in additional QALYs. However, it must be noted that there is considerable uncertainty around the results, due to uncertainties around parameters such as transition probabilities,

probabilities of adverse events and preference-based quality of life scores. At a ceiling ratio of £30,000 per QALY there is a 0.36 probability that Gem+Cisp is not the optimal choice. The NSCLC model and the obtained results form the basis for addressing the research-related decision in subsequent chapters.

CHAPTER 5. Adoption-related decision in hormone-refractory prostate cancer

This chapter reports the methods employed in building a decision model to synthesise the available evidence on the treatment adoption decision for hormone-refractory prostate cancer (HRPC). As in the NSCLC case study, the decision model and its results aim to facilitate the application of the 'payback' and 'value of information' analyses to the relevant case study. The first part of the chapter gives the background to HRPC and specifies the pertinent adoption-related and research-related decisions. The chapter continues with the methods used in this model and the results of the cost-effectiveness analysis.

5.1. Background

Prostate cancer is the most common cancer in men and one of the most common causes of cancer-related deaths in the UK, second only to lung cancer. In 2009, there were about 41,000 new cases of prostate cancer, approximately a quarter of which are anticipated to lead to death from the cancer[234].

In the early stages, prostate cancer is limited to the prostate gland and treatments for localised disease include early surgical resection, external beam radiotherapy or androgen suppression[235]. Although such treatments are initially beneficial, an estimated 10 to 50 percent of patients eventually become metastatic, with cancer spreading to the pelvic lymph nodes and bone[236;237]. Metastatic disease is commonly treated with hormone ablation therapies, either by bilateral orchiectomy or medical castration. Hormone ablation is

associated with high initial response rates, but recurrence is almost unavoidable. Over time, the majority of patients stop responding and become refractory to hormone therapy[238;239]. Treatments for hormone-refractory prostate cancer (HRPC) are, in essence, palliative. Options typically include chemotherapy, further hormone manipulations as well as traditional palliative therapies such as radiotherapy and surgery for obstructive syndromes and bone problems[240].

Standard practice with reference to chemotherapy treatment has emerged to be docetaxel. This is routinely combined with prednisolone (or prednisone, which is metabolised to prednisolone by the patient's organism), which has immunosuppressive and anti-inflammatory properties. In England and Wales, docetaxel plus prednisone/prednisolone (DP) is recommended by NICE as the standard treatment for patients with hormone-refractory prostate cancer[241;242], largely on the basis of favourable survival and quality of life detected in the TAX327 randomised controlled trial[243;244].

Advanced HRPC typically spreads to bones, which in turn results in severe skeletal pain. Bone metastases are observed in about 80 percent of the HRPC patients[245]; therefore it is common for HRPC treatments to combine chemotherapy with agents capable of addressing bone-related problems. Two agents which have been proven beneficial in skeletal-related problems are zoledronic acid (ZA) and strontium-89 (Sr89)[240;246;247]. ZA is an intravenously administered bisphosphonate that aims to inhibit the loss of bone mass and has demonstrated anti-cancer activity[248;249], while Sr89 is a bone-seeking radionuclide that acts by delivering therapeutic radiation to affected areas on bones. Both the treatments have

been proven effective in palliating bone metastases, and combining DP with ZA, Sr89 or both has been seen as potentially beneficial for patients suffering from advanced HPPC[250;251].

5.2. Adoption-related and research-related decisions

In light of the above, a treatment adoption-related question arises as to which chemotherapy combination is the most beneficial and should be provided to HRPC patients with bone-related problems. Given this, the relevant research-related decision focuses on whether funding should be devoted to a clinical trial aiming to give further primary evidence on the effectiveness of the following combinations:

a. docetaxel plus prednisolone (DP),

b. DP plus zoledronic acid (DP+ZA),

c. DP plus strontium-89 (DP+Sr89), and

d. DP plus zoledronic acid plus strontium-89 (DP+ZA+Sr89).

An early-stage trial, the Taxane Radioisotope Zoledronic Acid (TRAPEZE) phase II trial investigating these treatments had been already carried out. Continuation of this study to a phase III RCT was proposed to provide more robust evidence on the assessed treatments. The proposal was submitted for funding to the NIHR Health Technology Assessment programme through the programme's 'researcher-led' stream and secured funds in 2006. The total cost of the trial was £2.54 million and involved research-related cost (i.e. researchers' salaries, cost of facilities and trial organisation), excess treatment costs (i.e. cost

incurred due to participants in experimental arms receiving treatments other than standard care) and support costs (i.e. costs due to extra clinic time and patient recruitment).

5.3. Existing clinical and economic evidence

A literature review was carried out to retrieve existing evidence on the clinical effectiveness and cost-effectiveness of the chemotherapy combinations of interest. Searches for relevant literature were conducted in bibliographic databases (MEDLINE, NHS CRD, ISI Web of Knowledge), websites of institutions offering clinical guidance and systematic reviews (NICE Clinical Guidelines and Cochrane Collaboration Reviews), through Google Scholar®, as well as in the reference lists of relevant publications. Search term and methods used in this review are given in Appendix 4.A.

Identified studies were selected and their full text was retrieved if they:

a. had a focus on chemotherapy-eligible patients with advanced HRPC;

b. investigated chemotherapy combinations including either of DP, DP+ZA, DP+Sr89 or DP+ZA+Sr89 aimed as first-line treatment for HRPC;

c. reported cost-effectiveness results or clinical results on at least one of the following outcomes: tumour response, survival, time-to-progressive disease, quality of life, toxicity or treatment discontinuation rates.

The review aimed to retrieve information available up to the point when the decision to fund the TRAPEZE phase III trial was considered, thus it included articles published before

the end of 2006. Only one study[243] on clinical effectiveness and one economic study[244] were identified; these studies are described below.

5.3.1. Clinical effectiveness evidence

Much of the existing evidence on the effectiveness of DP comes from a single study, the TAX327 phase III randomised controlled trial[243]. The trial involved 1006 patients, who were randomised to three chemotherapy arms: (a) docetaxel at 75mg/m^2 plus 5mg prednisone or prednisolone administered every three weeks, (b) docetaxel at 30mg/m^2 plus 5mg prednisone or prednisolone administered weekly, and (c) mitoxantrone 12mg/m^2 plus 5mg prednisone or prednisolone every three weeks. The trial showed DP administered every three weeks to result in a statistically significant improvement in survival (HR: 0.76, CIs: 0.62 to 0.94), pain and quality of life compared to mitoxantrone[243]. A summary of this study is given in Table 5.1.

Table 5.1: Summary of main effectiveness studies (HRPC)

Study	Study type	Intervention	Comparator	Outcome measures	Authors' conclusions
Tannock *et al*. (2004)[243]	Phase III RCT	Mitoxantrone + prednisone or prednisolone (Patients in arm=335)	Docetaxel 75mg/m^2 every 3 weeks + prednisolone or prednisone (Patients in arm=332) Docetaxel 30mg/m^2 weekly + prednisolone or prednisone (Patients in arm=330)	Survival Pain Prostate-specific antigen levels Adverse events Quality of life	Docetaxel and prednisolone provided every three weeks is associated with superior survival and improved outcomes in terms of pain, prostate-specific antigen level, and quality of life, as compared with mitoxantrone plus prednisone.

Evidence from the TAX327 study was used in the NICE Technology Assessment Report commissioned to investigate the effectiveness and cost-effectiveness of DP[244;252] and had a significant impact on the Institute's decision to recommend DP as the preferable treatment for HRPC patients who are generally able to care for themselves[242]. No studies comparing DP and DP+ZA, DP+Sr89 and DP+ZA+Sr89 were identified in the literature.

5.3.2. Cost-effectiveness evidence

The only relevant cost-effectiveness study identified in the literature was the Technology Assessment Report mentioned above, by Collins and colleagues[244]. The study aimed to explore the effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone compared to a series of alternative treatments including mitoxantrone and estramustine[244]. To investigate this, Collins *et al*.[244] constructed a two-state (alive and dead) Markov model and populated it with evidence from the case for adoption of Taxotere® (trade name of docetaxel) to NICE. The study found two treatments not to be dominated: DP given every three weeks, and mitoxantrone plus prednisone/prednisolone. In the comparison between them, DP resulted in an ICER of approximately £32,700 per additional QALY (£33,750 in 2006 prices). A summary of the study is given in Table 5.2 below.

Table 5.2: Summary of main economic studies (HRPC)

Study	Study type	Compared treatments	Authors' conclusion
Collins *et al*. (2005)[244]	Systematic review, cost-effectiveness analysis and cost-utility analysis	Docetaxel + prednisone/prednisolone (3-weekly) Mitoxantrone + prednisone/prednisolone Prednisolone/prednisone alone Docetaxel + prednisone/prednisolone (weekly) Docetaxel + estramustine Docetaxel + estramustine + prednisone/prednisolone (70mg/m^2 every 3 weeks) Docetaxel + estramustine + prednisone/prednisolone (35 mg/m^2 twice every 3 weeks) Mitoxantrone + prednisone/prednisolone + clodronate	Docetaxel plus prednisone/prednisolone (3-weekly) appears cost-effective in comparison to other chemotherapy and non-chemotherapy regimens, only if the NHS is willing to pay £32,706 per QALY.

5.3.3. Evidence from the TRAPEZE phase II clinical trial

As noted earlier, patient-level evidence on the specific treatments was available from the TRAPEZE phase II trial. This was a randomised controlled trial involving 200 HRPC participants allocated into four treatment arms: DP (50 participants), DP+ZA (49 participants), DP+Sr89 (51 participants) and DP+ZA+Sr89 (50 participants). The study collected data on progression-free survival, adverse events, health care resource use and health-related quality of life.

5.4. Decision modelling for HRPC

Decision analytic models are used widely to synthesise evidence and provide summary results on the cost and effectiveness of health technologies[51;157]. In the context of this study, a decision model was constructed to investigate the clinical and cost-effectiveness of DP, DP+ZA, DP+Sr89 and DP+ZA+Sr89 on the basis of available evidence. Methods involved in building, populating and analysing the model are given in the following section.

5.4.1. Model structure

A Markov model was developed to assess and compare the cost-effectiveness of the treatments of interest. Markov models are commonly used to represent disease pathways in areas where events occur at discrete time periods[253;254]. The model consists of four discrete health states:

a. 'Progression-free, on treatment' (PGF-OT) where advanced HRPC patients with stable disease receive one of the compared chemotherapy treatments;

b. 'Progression-free, not on treatment' (PGF), reflecting the state in which patients have not shown signs of progression, but they have stopped receiving treatment, either because they completed the course or because they discontinued before the end of the scheduled treatment period;

c. 'Progression' (PG), where patients have developed progressive disease, and

d. 'Death' (D).

A cohort of HRPC patients in stable disease enter the model in the PGF-OT state, where they are scheduled to receive six cycles of chemotherapy, with each cycle lasting three weeks. Patients stay in this state for six cycles, unless they discontinue treatment due to intolerable toxicity (in which case they move to the state PGF), discontinue due to disease progression (in which case they move to the state PG), or die.

At the end of the treatment course, patients who have completed all six cycles move to the PGF state. Upon progression, patients move to the state PG and, eventually, to the absorbing state D. The structure of the decision model is given in Figure 5.1.

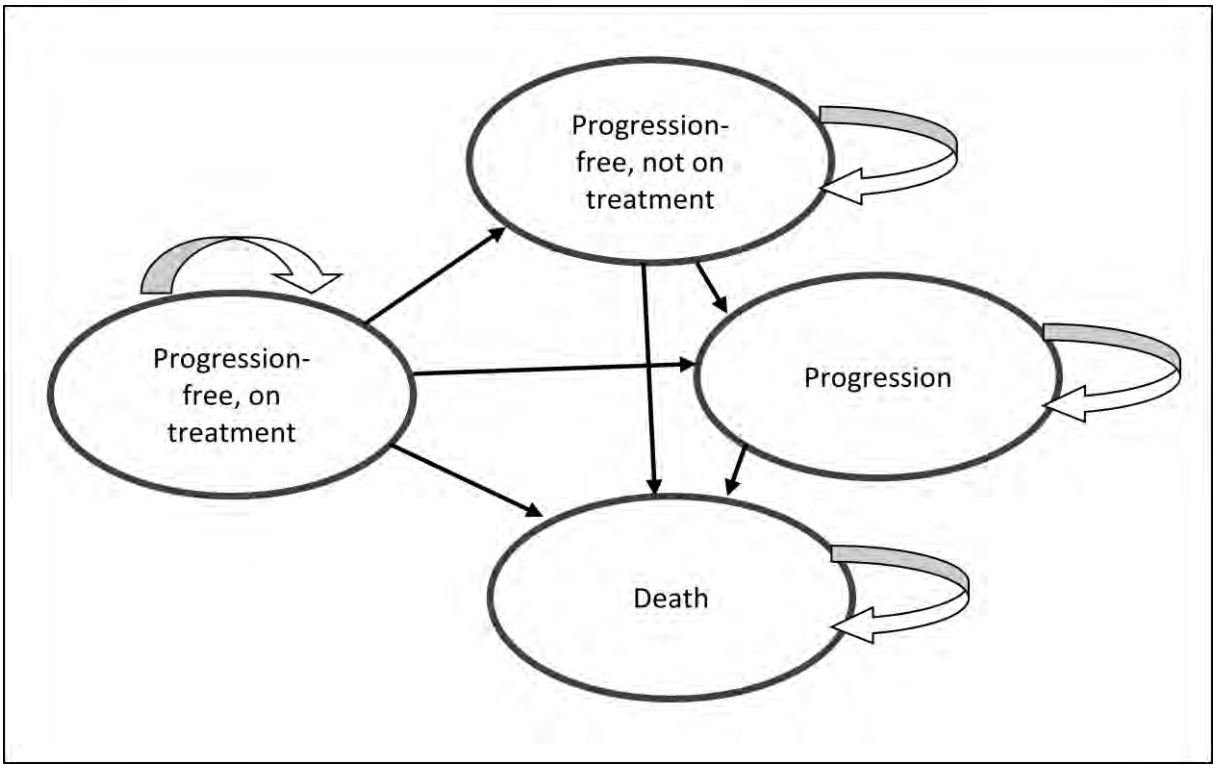

Figure 5.1: Model structure for HRPC

The length of each discrete model cycle was set at three weeks to concur with a treatment cycle. Patients were followed for four and a half years (80 cycles), by which time nearly the entire cohort of patients had died. Evidence on patients' transitions between model states was based on available patient-level data from the TRAPEZE phase II trial. When patient-level data was not available, estimates were drawn from the published literature and expert opinion.

5.4.1.A. Transition probabilities

Probabilities of transitions between model states were sought for two main stages in the model: during the period of scheduled treatment (cycles 1 to 6), where transitions reflected treatment discontinuation, and thereafter (cycles 7 to 80), where transitions represented disease progression after a patient had gone past the treatment period.

As mentioned earlier, the model was designed to take into account the fact that HRPC patients may discontinue treatment at any point before completion of the scheduled six cycles of chemotherapy. Reasons for discontinuations are toxicity, disease progression or death. In the case of discontinuation due to intolerable toxicity, a patient would move from PGF-OT to PGF; in case of discontinuation due to progression, a patient would move from PGF-OT to PG, while in the event of death, a patient would move from PGF-OT to D. Discontinuation rates were calculated using phase II trial data on numbers of patients who withdrew from treatment at different points during cycles 1 to 6. For example, the probability of a patient discontinuing DP due to progressive disease and thus moving from PGF-OT to PG was found by dividing the observed five transitions to PG by the total of 219

patient-cycles at risk of progression, to give a value of 0.023. For each treatment arm, numbers of patient-cycles spent at each of the four states during the first six cycles are given in Appendix 4.B, Table 4.b.

Probabilities of transition taking place after the scheduled treatment period (i.e. after cycle 6) were calculated in a similar way, by dividing the number of observed transitions by the total number of patient-cycles at risk. For example, the probability of a patient on DP treatment moving from PG to D was calculated from the observed number of 29 deaths, divided by the total of 596 patient-cycles at risk to give a probability of 0.049. Transition matrices showing counts of cycles spent at each health state in the period beyond cycle 6 can be found in Appendix 4.B, Table 4.c.

To test whether the use of non-time-dependent transition probabilities is appropriate for representing the available data, two Kaplan-Meier survival curves were generated for each treatment, the first depicting the raw patient-level survival data obtained from the TRAPEZE phase II trial and the second showing survival calculated from entering the estimated non-time-dependent transition probabilities in the model. A close match between the curves was observed (see Appendix 4.C) suggesting that the calculated non-time dependent transition probabilities were a good representation of the empirical data and thus they were appropriate for use in the model.

5.4.1.B. Resource use and costs

Costs to the NHS were calculated by taking into account health care resource use accruing under four main categories:

a. cost of drug acquisition and administration,

b. cost of serious adverse events,

c. cost associated with second-line treatment, and

d. cost of terminal care.

The cost of drug acquisition was calculated according to dosages reported in the TRAPEZE trial protocol for a male patient of average body surface area of 1.9 square metres[244]. According to the study's treatment schedule, both DP and ZA administrations—relevant to the DP, DP+ZA and DP+ZA+Sr89 arms—would take place on the first day of each 3-week cycle for a total of six chemotherapy cycles, while Sr89—relevant to DP+Sr89 and DP+ZA+Sr89 arms—would be administered once, at the end of the sixth cycle. Unit costs of drugs were taken from the British National Formulary[222]. As the unit cost of strontium-89 was not available, an estimate was obtained from a NHS hospital [personal communication with Dr C. Boivin, Head of Nuclear Medicine, Queen Elizabeth Hospital Birmingham, 03-06-2011]. Costs are reported in 2006 prices, in order to relate to the point in time when funding for the TRAPEZE phase III trial was considered. Information on doses, constituent parts and unit costs is given in Table 5.3.

Table 5.3: Unit costs of drug acquisition and administration (HRPC)

Resource	Dose (assuming body surface area 1.9 square metres)	Cost (2006 prices)	Source
Docetacel (75mg)	142.5mg	£986	British National Formulary[222]
Prednisolone (10mg daily)	10mg (daily, for 21 days per cycle)	£12	
Dexomethasone (antiemetics)	8mg	£5	
Zoledronic acid (4mg)	4mg	£169	
Strontium-89	150MBq	£1576	Personal communication: Dr C. Boivin, Department of Nuclear Medicine, Queen Elizabeth Hospital Birmingham
Chemotherapy administration (outpatient)	£158		NHS Reference Cost Schedules[221]

In accordance with the literature, chemotherapy was considered to be administered in an outpatient setting[243;244]. The cost of chemotherapy administration was taken from published NHS Reference Costs Schedules[221] and represents the national average cost for an outpatient chemotherapy delivery. The total cost per treatment cycle for DP, DP+ZA, DP+Sr89 and DP+ZA+Sr89 is given in Table 5.4.

Table 5.4: Total cost of drug acquisition and administration for a treatment cycle (HRPC)

Treatment	DP	Zoledronic acid	Strontium-89	Antiemetics	Administra-tion	Total cost per treatment cycle
DP	£998	-	-	£5	£158	£1160
DP+ZA	£998	£169	-	£5	£158	£1329
DP+Sr89	£998	-	£263*	£5	£158	£1423
DP+ZA+Sr89	£998	£169	£263*	£5	£158	£1592
*A single fraction of strontium-89 is given in a six-cycle course. For presentation purposes, the cost of strontium-89 (£1576) has been amortised to one treatment cycle.						

The cost of serious adverse events leading to hospitalisation was calculated by estimating the probability of a patient experiencing an adverse event on the basis of events observed in the phase II trial and multiplying this by the unit cost of resolving the adverse event, using the formula:

$$Expected\ cost\ of\ adverse\ events_k = \sum P(AE_{i,k}) \times UC_i$$

Here, $P(AE_{i,k})$ is the probability of experiencing an adverse event i while on treatment k and UC_i is the unit cost of resolving the event i. Unit costs for resolving the most common adverse events (diarrhoea, febrile neutropenia, haemoglobin, infection, neutrophils/granulocytes, pain and urinary retention) were obtained from the NHS Reference Cost Schedules[221]. Other, less common adverse events were grouped together under 'other adverse events' and were assigned the national average cost of a non-elective inpatient hospital stay taken from the Unit Cost of Health and Social Care report[220]. The expected cost of adverse events is given in Table 5.5. Counts of patients experiencing different adverse events and the unit cost for treating these events are given in Appendix 4.D, Table 4.d and Table 4.e, respectively.

Table 5.5: Expected cost of adverse events (HRPC)

Adverse event	Expected cost of adverse events			
	DP	**DP+ZA**	**DP+Sr89**	**DP+ZA+Sr89**
Diarrhoea	£6	£6	£6	£12
Febrile neutropenia	£26	£27	£51	£17
Haemoglobin	£10	£11	£0	£21
Infection	£37	£30	£14	£15
Neutrophils/granulocytes	£35	£0	£17	£0
Pain	£90	£55	£124	£54
Urinary retention	£0	£34	£0	£8
Other	£810	£537	£357	£972
Total expected cost	**£1014**	**£699**	**£569**	**£1099**

Depending on their physical condition, patients who do not respond to chemotherapy may be eligible for second-line palliative care, typically in the form of radiotherapy, further chemotherapy or further treatment with radioisotopes. For each treatment, the expected cost of second-line care was calculated by weighting the probability of a patient receiving further chemotherapy, radiotherapy or radioisotopes treatment by the cost of these palliative options, according to the formula:

$$Expected\ cost\ of\ second\text{-}line\ treatment\ _k = \sum P(T_{i,k})\ \times UC_i$$

Here, $P(T_{i,k})$ is the probability of a patient on treatment k to receive second-line treatment i, obtained from phase II trial data and UC_i is the unit cost of second-line treatment i. The total expected cost of second-line treatment is given in Table 5.6, while counts of patients

who received second-line treatment and unit costs of each of the provided treatments are given in Appendix 4.D, Table 4.f and Table 4.g, respectively. The expected cost was calculated on the basis of the fact that patients receiving further chemotherapy are typically given an average of four cycles of DP, while those who are treated with radiotherapy or further radioisotopes (strontium-89) typically receive a single fraction of radiation [personal communication with Professor N. James, Professor of Oncology, University of Birmingham, 06-03-2011]. Last, the cost of terminal care was taken from the literature[201] (£1532 in 2006 prices).

Table 5.6: Expected cost of second-line treatment (HRPC)

	DP	DP+ZA	DP+Sr89	DP+ZA+Sr89
Expected cost of second-line treatment	£606	£889	£743	£475

5.4.1.C. Health-related quality of life

Data on patients' health-related quality of life were collected throughout the trial by using the EQ-5D health status classification instrument. Responses to EQ-5D were translated to preference-based (utility) scores by using a UK-specific tariff[255;256]. These data allowed estimating the mean utility scores by treatment and health state given in Table 5.7 below. The utility score for the state 'Death' is zero.

Table 5.7: Preference-based quality of life scores by health state and treatment (HRPC)

Treatments	'Progression-free, on treatment' state		'Progression-free, not on treatment' state		'Progression' state	
	Mean	Standard error	Mean	Standard error	Mean	Standard error
DP	0.625	0.04	0.605	0.05	0.500	0.08
DP+ZA	0.712	0.03	0.706	0.03	0.569	0.05
DP+Sr89	0.714	0.04	0.693	0.04	0.503	0.09
DP+ZA+Sr89	0.724	0.03	0.624	0.05	0.558	0.08

5.4.2. Analysis of uncertainty

Probabilistic sensitivity analysis was carried out to account for the uncertainty around key parameters in the model. Five thousand sets of values were drawn from distributions attached to parameters. For each set of values, results were recalculated to give 5000 estimates of each treatment's total costs and effects[169;171].

With reference to transitions probabilities, the move from state 'Progression' to 'Death' was informed by binomial data and was represented by a beta distribution. Beta distributions are characterised by parameters α and β, where

$$\alpha = number\ of\ observed\ events$$

$$\beta = sample\ size - number\ of\ observed\ events$$

Following on from the earlier example, the probability of a patient in the DP arm moving from 'Progression' to 'Death' can be represented by a beta distribution with parameter α

equal to the number of observed transitions (29) and parameter β equal to the number of patient-cycles at risk of transition (596) minus the observed transitions (beta(29,567)).

Patients in states other than 'Progression' (i.e. PGF-OT and PGF) may move to more than one different states (e.g. from PGF-OT to PGF or PG or D), that is, the data informing these transitions are multinomial. Thus, uncertainty around the transition probabilities from 'Progression-free, on treatment' and 'Progression-free, not on treatment' to other states can be represented by Dirichlet distributions. Dirichlet distributions are multivariate extensions of beta distributions, where the number of parameters is equal to the number of categories (here, transitions). For example, the uncertainty around the probability for a patient on DP moving from 'Progression-free, on treatment' to the rest of the states is represented by a Dirichlet(204,5,5,5). Details on parameters and their respective distributions are given in Appendix 4.E, Table 4.h.

Similarly, estimated costs of drug acquisition and administration were assigned gamma distributions. No information on the variance of these estimates was available. Therefore, standard errors were hypothesised to be 10 percent of the mean values, on the premise that drug acquisition and administration costs are typically well specified in national formularies[221;222] and they do not tend to vary greatly.

The expected cost of adverse events was made probabilistic by varying the probability of patients on each treatment experiencing different serious adverse events. To do so, beta distributions were attached to counts of adverse events taken from the TRAPEZE phase II trial. Similarly, the cost of second-line treatment was varied through assigning beta

distributions to the probability of patients in each treatment receiving further chemotherapy, radiotherapy or radioisotopes as second-line care. The cost of terminal care was assigned a gamma distribution, with the standard error set at a value of 25 percent of the mean cost on the assumption that there is greater variability around such costs than for drug acquisition and administration costs. Details on distributions attached to cost-related parameters are given in Appendix 4.E, Table 4.i.

Last, uncertainty around utility scores for each health state and treatment was propagated by calculating utility increments. Here, rather than varying utility scores of different health states, probability distributions were attached to differences in mean utility scores between states (PGF-OT and PGF; PGF-OT and PG), to ensure that utility scores for more severe health states do not assume values higher than those of 'milder' states. Utility increments were attached normal distributions with mean and standard errors estimated from the available data. These can be found in Appendix 4.E, Table 4.j.

In addition to probabilistic analysis, deterministic sensitivity analyses were carried out to explore the impact of different assumptions around values of discount rates, patients' mean body surface area and setting of chemotherapy administration on results.

5.4.3. Cost-effectiveness analysis

The perspective adopted in the analysis was that of the NHS. Costs and benefits(QALYs and life-years gained) were discounted at an annual rate of 3.5 percent, in line with current NICE recommendations[114]. The decision model was built and analysed in Microsoft Excel® (Microsoft Corporation, v.2007, Redmond, Washington, US), while statistical analysis tasks

were also carried out in STATA® (StataCorp, v. 11, College Station, Texas, US) and Parameter Solver (MD. Anderson Cancer Centre, University of Texas, Houston, Texas, US).

5.5. Results

Cost-effectiveness results obtained from the model are expressed as point estimate ICERs (in terms of cost per QALY and cost per LYG) and are plotted in cost-effectiveness planes (CE planes), cost-effectiveness acceptability curves (CEACs) and cost-effectiveness acceptability frontiers (CEAFs).

5.5.1. Point-estimate results

Mean total cost, QALYs and life-years gained for each treatment calculated through the model are given in Table 5.8. As the comparison involved more than two mutually exclusive treatments, ICERs were calculated by sorting the treatments in order of increasing cost, eliminating dominated treatments and comparing the remaining options[139;257].

Table 5.8: Point estimate cost-effectiveness results (HRPC)

Treatment	Total cost	Total QALYs	Total LYG	ICER (£ per QALY)	ICER (£ per LYG)
DP	£8,949	0.806	1.383	-	-
DP+ZA	£9,855	0.829	1.251	Weakly dominated (£39,726)	Dominated
DP+Sr89	£10,172	0.957	1.481	£8113(against DP)	£12,491
DP+ZA+Sr89	£11,436	0.874	1.341	Dominated	Dominated

The analysis showed DP to be the least costly treatment, followed by DP+ZA, DP+Sr89 and DP+ZA+Sr89. In terms of QALYs, DP was the least effective treatment. Although DP+ZA is

more effective than DP, this treatment is weakly dominated by a combination of DP and DP+Sr89. DP+ZA+Sr89 is more costly and less effective than DP+Sr89 and thus it is dominated by the latter. In the comparison between non-dominated treatments, DP+Sr89 is associated with a greater cost and additional QALYs and an ICER of about £8100 per additional QALY. In terms of cost per life-years gained, DP+ZA and DP+ZA+Sr89 are strongly dominated by DP and DP+Sr89, respectively. Here, too, the remaining comparison is between DP and DP+Sr89, with the latter being associated with an ICER of about £12,500 per life-year gained.

5.5.2. Results of deterministic sensitivity analyses

Deterministic sensitivity analyses were conducted to investigate the effect of various plausible assumptions on results. The results of these analyses are given in Table 5.9. Overall, the impact of different assumptions on the magnitude of the results was limited. Alternative dosages of chemotherapy treatments resulting from different assumptions about patients' average body surface area did not result in additional or fewer drug vials used and thus these assumptions had no impact on total costs. The setting of chemotherapy administration had a minimal impact on costs; assuming that half of the patients received chemotherapy in an inpatient setting raised the total cost by approximately £18. A somewhat larger change in costs and benefits resulted by employing different discount rates—0 percent (i.e. no discounting) and 6 percent.

Table 5.9: Results of deterministic sensitivity analysis (HRPC)

Parameter	DP		DP+ZA		DP+Sr89		DP+ZA+Sr89		ICER (£ per QALY) (DP vs. DP+Sr89)
	Total cost	Total QALYs	Total cost	Total QALYs	Total costs	Total QALYs	Total costs	Total QALYs	
Base case	£8949	0.806	£9855	0.829	£10,172	0.957	£11,436	0.874	£8113
Average body surface area 2m^2 (base case 1.9 m^2)	No changes in total costs due to greater/smaller average body surface area, as higher/lower active treatment dosage is covered by the same 8ml vial.								
Average body surface area 1.8 m^2 (base case 1.9 m^2)									
Chemotherapy administration- 90% outpatient, 10% inpatient	£8966	0.806	£9872	0.829	£10,190	0.957	£11,454	0.874	£8105
Chemotherapy administration- 80% outpatient, 20% inpatient	£8983	0.806	£9889	0.829	£10,207	0.957	£11,472	0.874	£8112
Chemotherapy administration- 70% outpatient, 30% inpatient	£8999	0.806	£9906	0.829	£10,225	0.957	£11,490	0.874	£8118
Chemotherapy administration- 60% outpatient, 40% inpatient	£9016	0.806	£9923	0.829	£10,243	0.957	£11,507	0.874	£8124
Chemotherapy administration- 50% outpatient, 50% inpatient	£9033	0.806	£9940	0.829	£10,261	0.957	£11,525	0.874	£8130
No discounting	£9046	0.837	£9952	0.858	£10,290	0.992	£11,553	0.904	£8006
Discount rates- 6% p.a. for both costs and benefits	£8885	0.786	£9791	0.810	£10,094	0.934	£11,358	0.854	£8186

For different assumptions, incremental analysis of DP versus DP+Sr89 (i.e. the non-dominated treatments) resulted in ICERs broadly similar to that in the base case. The additional cost arising from the hypothesis that up to half of the treatment administrations may take place at an inpatient setting had an approximately equal effect to all treatments and it was cancelled out in incremental analysis, resulting in ICER values similar to that of the base case analysis (range between £8105 and £8130). Compared to the base case value of £8113 per QALY, the ICER calculated from undiscounted costs and benefits was slightly more 'favourable' to DP+Sr89, while discounting at 6 percent per annum led to an increase in the ICER of DP+Sr89 of approximately £70.

5.5.3. Results of the probabilistic sensitivity analysis

Cost and QALY estimates obtained from the probabilistic sensitivity analysis were plotted in CE planes. These show the comparisons between DP and DP+ZA (Figure 5.2), DP+Sr89 (Figure 5.3) and DP+ZA+Sr89 (Figure 5.4). Each point on the planes represents a simulated pair of incremental costs and effects (QALYs).

The proportions of points in each quadrant of the CE planes are shown in Table 5.10. With respect to the comparison between DP and DP+ZA, approximately 79 percent of the 5000 points appear in the north half of the plane, indicating that this treatment is likely to result in higher costs than DP. On the other hand, the points appear split evenly between the west and east halves of the plain (56 percent and 44 percent respectively), suggesting that there is considerable uncertainty on whether DP+ZA is more effective than DP.

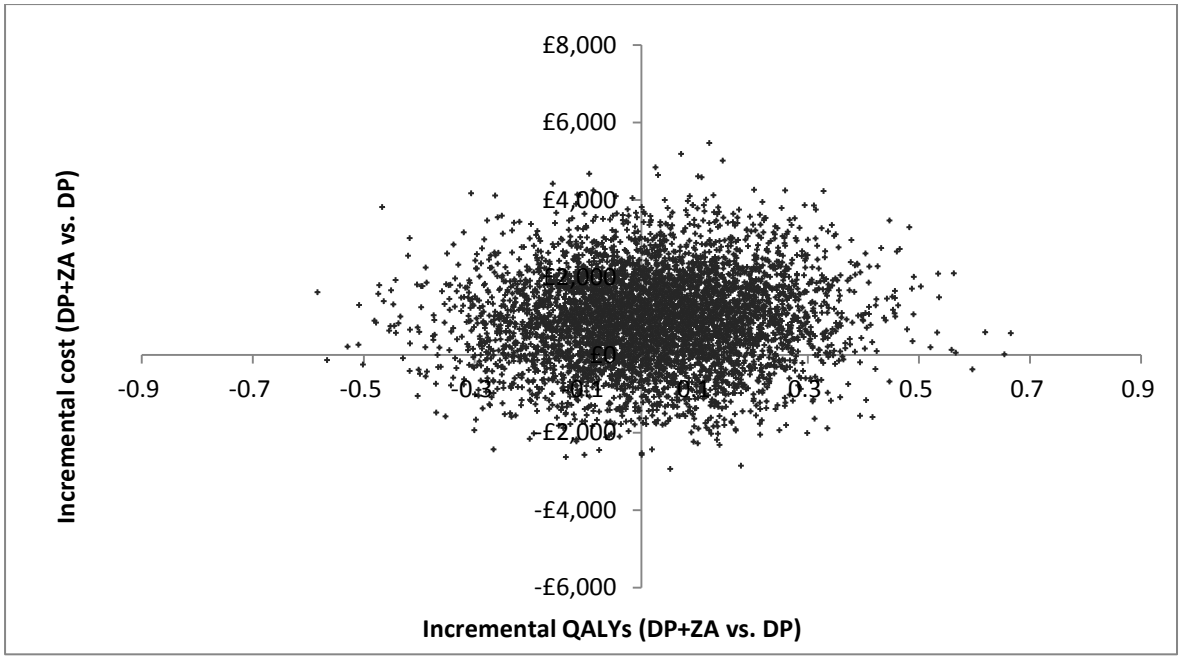

Figure 5.2: Cost-effectiveness plane for comparison between DP+ZA and DP (HRPC)

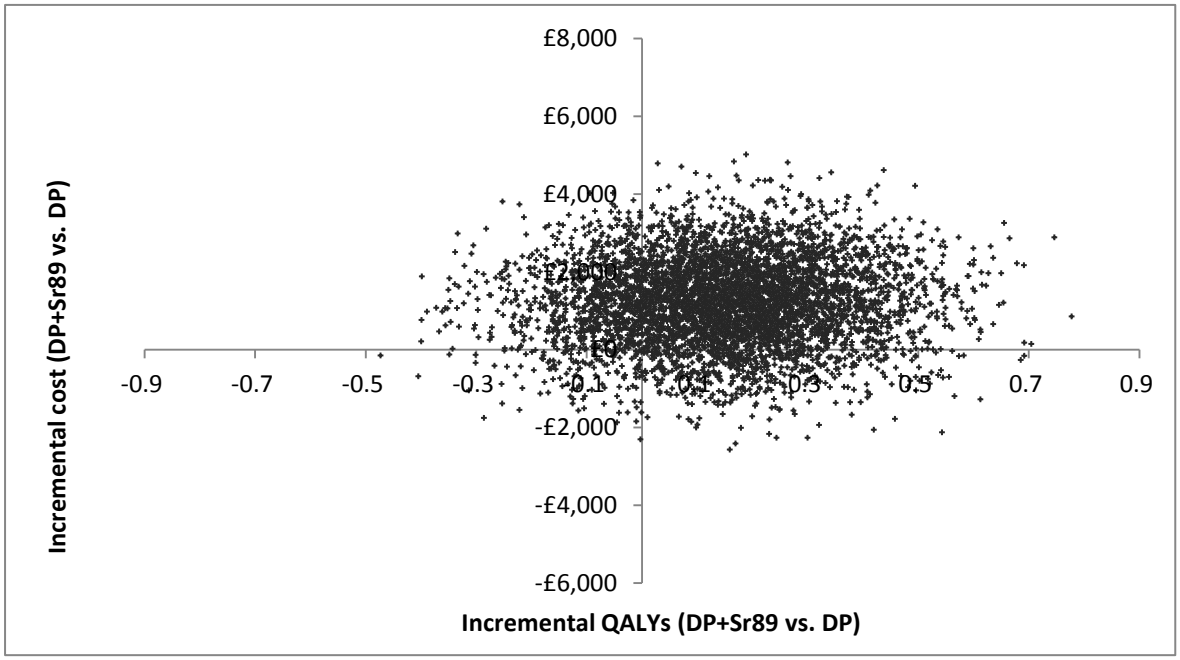

Figure 5.3: Cost-effectiveness plane for comparison between DP+Sr89 and DP (HRPC)

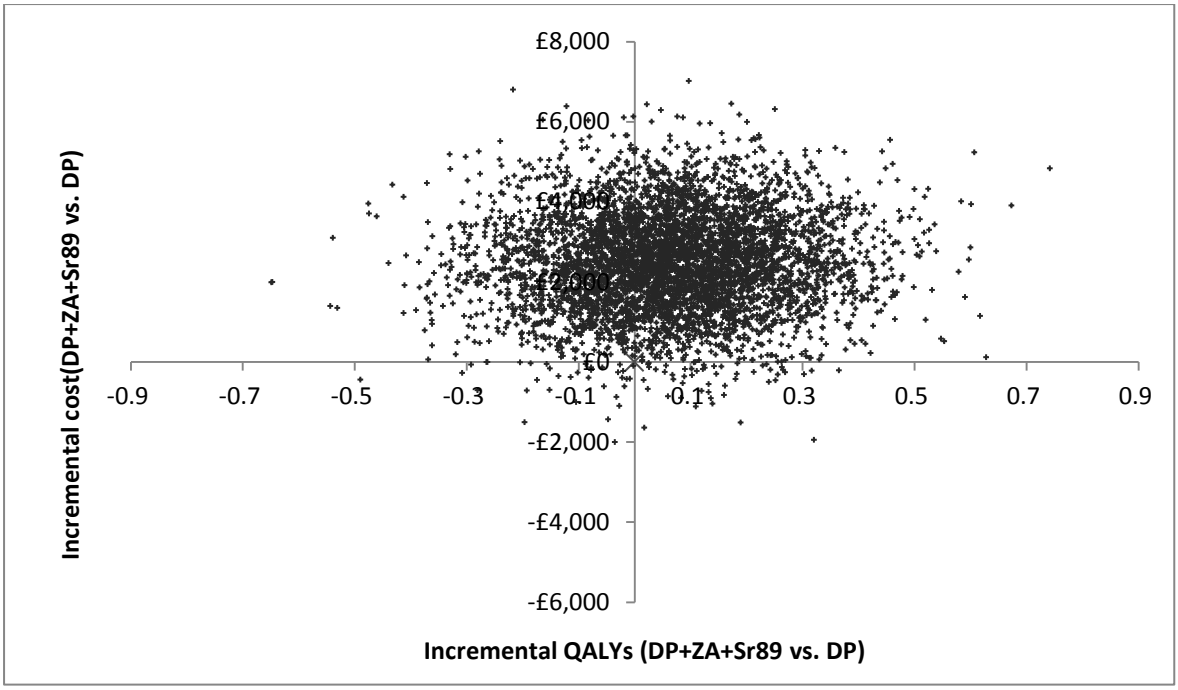

Figure 5.4: Cost-effectiveness plane for comparison between DP+ZA+Sr89 and DP (HRPC)

Table 5.10: Proportion of cost and QALY pairs on cost-effectiveness planes (HRPC)

Quadrant	Proportion of points in each quadrant of the CE planes		
	DP+ZA vs. DP (Figure 5.2)	**DP+Sr89 vs. DP** (Figure 5.3)	**DP+ZA+Sr89 vs. DP** (Figure 5.4)
North East	0.45	0.70	0.66
South East	0.11	0.10	0.01
South West	0.10	0.03	0.01
North West	0.34	0.17	0.32

In the comparison between DP and DP+Sr89, the majority of the points (about 70 percent) appear in the north east quadrant, indicating that DP+Sr89 is more costly and more effective than DP. Last, with respect to the comparison between DP and DP+ZA+Sr89, the latter

treatment is highly likely to be more costly (98 percent of points in the north half), while there is a reasonably high likelihood that this treatment is also of superior effectiveness (66 percent of points in the west half).

CEACs representing the probability of each of the compared treatment being cost-effective for ceiling ratio values ranging from £0 to £80,000 are given in Figure 5.5. If a decision-maker is not willing to pay any additional amount for extra health benefits (i.e. the ceiling ratio is zero), the most cost-effective option is DP, with a probability of 0.72. For a ceiling ratio of £30,000 per QALY, the treatment with the highest probability of being cost-effective switches to DP+Sr89 (probability of 0.53). At a high ceiling ratio value of £80,000 per QALY, the probability of DP+Sr89 being the optimal treatment is about 0.55 while DP+ZA+Sr89 has the second highest probability, of approximately 0.19.

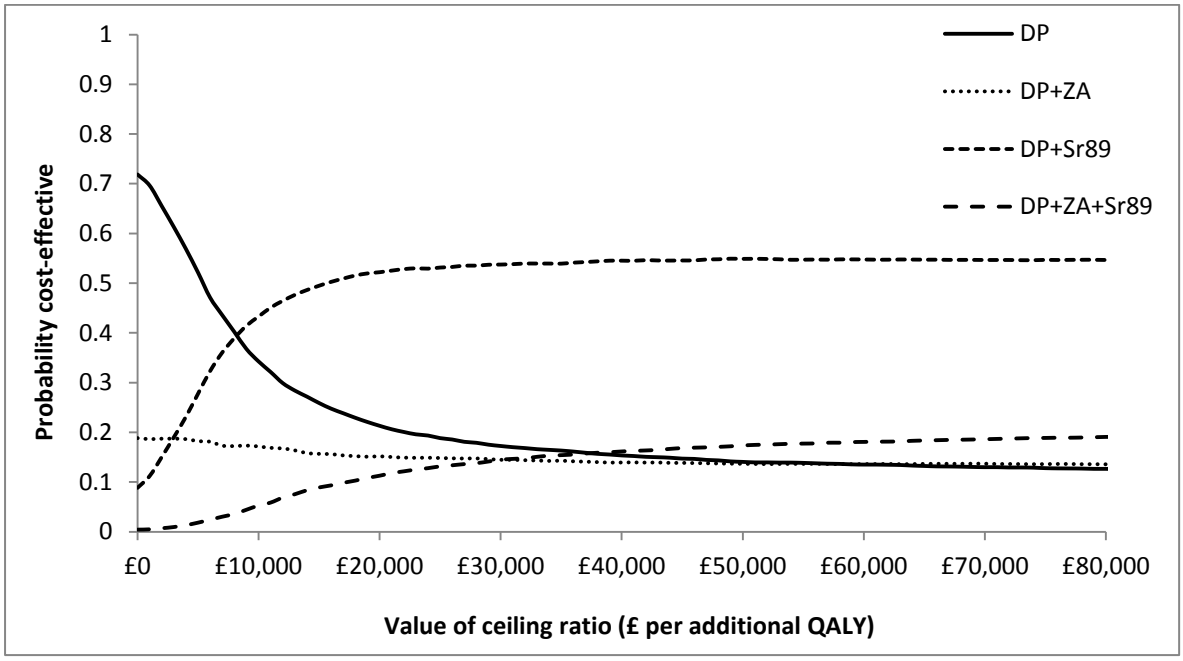

Figure 5.5: Cost-effectiveness acceptability curves (HRPC)

The results were also plotted as a cost-effectiveness acceptability frontier (CEAF), which plots the treatment resulting in the highest net monetary benefits (NMBs) at different values of the ceiling ratio. The frontier showed that DP is cost-effective at ceiling ratios up to £8100, with DP+Sr89 becoming cost-effective at ceiling ratios over this value (Figure 5.6).

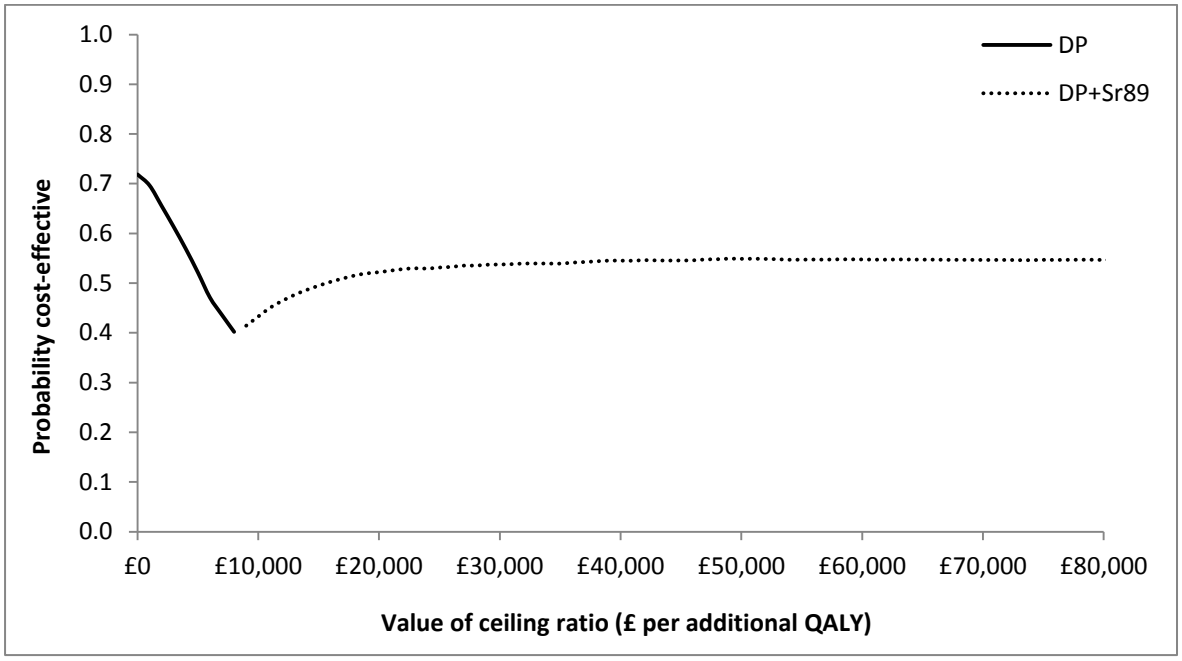

Figure 5.6: Cost-effectiveness acceptability frontier (HRPC)

5.6. Discussion

The chapter presented a cost-effectiveness analysis of chemotherapy treatments for HRPC. This analysis sought to answer the adoption-related decision around HRPC given existing evidence and aimed to serve as the basis for assessing the need for further evidence in the following chapters. Overall, DP+Sr89 appears to be a cost-effective alternative to DP,

resulting in a relatively low ICER of approximately £8100 per additional QALY. Nonetheless, the existence of uncertainty in the results suggests that, at a ceiling ratio of £30,000 per QALY, the probability that DP+Sr89 is the optimal choice is low, at 0.53.

Although the generated results do not aim to inform 'real world' treatment adoption decisions, the work aimed to be as comprehensive as possible. Given this, the analysis presents certain strengths. A particular strength arises from the use of a Markov model. Such models trace patients from the point they receive treatment until death and measure the accumulated survival, costs and health outcomes at discrete points in time[253;258]. In addition, the model is structured in a way that takes into account treatment discontinuation, a factor which has a sizeable impact on treatments' cost and effectiveness. Moreover, key parameters in the model, such as transition probabilities and probability of experiencing different adverse events were calculated on the basis of patient-level data obtained from a phase II study involving the very same population and very same treatment options that the analysis is concerned with. Last, uncertainty around model parameters was accounted for in deterministic as well as probabilistic sensitivity analyses.

However, the analysis is not without limitations. First, the phase II trial from which estimates of treatment effect, adverse events and health-related quality of life are taken was relatively small, involving approximately 50 patients per treatment arm. However, the uncertainty around these was made explicit by assigning distributions to these parameters. An additional limitation is associated with the fact that entering the 'Progression-free, not on treatment' state are patients who either completed the treatment course, or

discontinued due to toxicity. These two subgroups may differ in the quality of life they experience, but this difference is expected to be small, given that toxicity events are typically resolved quickly and are not expected to result in prolonged hospitalization or chronic conditions which could have a notable impact on quality of life[239]. Last, to avoid ignoring uncertainty around parameters for which estimates of variance were not available (e.g. cost of a strontium-89 fraction, cost of terminal care), employed standard errors were based on assumptions.

5.7. Chapter overview

The chapter reports the methods and results obtained from a cost-effectiveness analysis aimed to address the adoption-related decision in the HRPC case study. Results showed that two of the assessed treatments (DP+ZA, DP+ZA+Sr89) were dominated. Of the remaining treatments, DP+Sr89 appeared cost-effective as compared to DP, resulting in a low ICER of approximately £8100 per QALY. This conclusion, however, should be viewed with caution, as the probability of DP+Sr89 being cost-effectiveness at a ceiling ratio of £30,000 is approximately 0.53, suggesting that there is considerable uncertainty around the results and a 0.47 probability that DP+Sr89 is not the optimal treatment. The constructed model and its results are key to the application of 'payback' and 'value of information' to the HRPC case study, which is reported in the following chapters.

CHAPTER 6. Practical application of 'payback of research' to case studies

This chapter reports the practical application of the 'payback of research' framework to the NSCLC and HRPC case studies. The application was carried out by using the Preliminary Assessment of Technology for Health Services (PATHS)[67], the most recent 'payback' model identified in the literature. Rather than to provide results for an actual decision on funding further research, the application aimed to identify strengths and limitations of this framework and suggest areas where improvements may be needed. Observations made throughout this analysis form the basis for assessing and discussing 'payback' in Chapters 8 and 9.

The chapter begins with a brief description of the rationale and methods underpinning the PATHS model and continues with the application of the model to each of the case studies. The final part of the chapter discusses points relevant to this application.

6.1. Description of methods in PATHS

As explained earlier in Chapter 3, 'payback' models are based on the notion that evidence generated through evaluative research is valuable because it enables informed choices about treatment provision, which, in turn, lead to additional benefits arising to the population. On this basis, the approach assumes that the value of conducting a piece of research—for example, a trial investigating two treatments—can be inferred by looking at the difference in costs and benefits (or, for simplicity, the difference in net monetary

benefits (NMBs)) expected to accrue to the population under two distinct situations: with and without the trial taking place. In the first situation, where a trial takes place, it is expected that it will generate evidence on the treatments' effectiveness (for example, disease progression rates), this evidence will be used as input in an economic evaluation study to give estimates of the treatments' per-patient cost and benefits, and, once these estimates become available, there will be a change in clinical practice. In the second situation, where research does not take place, clinical practice is expected to remain largely as it is. In each of these situations, the total sums of NMBs accruing to the population of patients eligible for the treatments in question depend on two factors: a) the costs and benefits associated with the treatments, and b) the use of these treatments in clinical practice, in terms of the proportion of patients receiving each of them.

In a prospective situation, that is, before a trial has taken place, results cannot be known in advance and need to be hypothesised. Therefore, hypothetical outcomes need to be specified so that they cover different eventualities. In general, outcomes for an assessed treatment may be 'favourable' (i.e. the trial shows results according to which a treatment appears cost-effective), 'inconclusive' (i.e. trial results are such that the treatment appears of inconclusive cost-effectiveness) or 'unfavourable' (i.e. the trial reveals results such that the treatment does not appear cost-effective).

An important assumption in the approach is that each specified possible outcome (i.e. 'favourable', 'inconclusive' and 'unfavourable') taken one at a time represents the actual underlying cost and benefits of the treatment, which hold true no matter whether research

has been conducted and revealed them or not. This assumption underpins the calculation of benefits with and without research. In simple terms, if research takes place, the specified (considered 'true') effectiveness of a treatment will be revealed, its cost-effectiveness will be subsequently established and this will trigger a beneficial change in clinical practice. For example, if a trial assessing treatments A and B takes place and reveals A to be more cost-effective than B, practice will change so that more patients will be given the cost-effective treatment A, while the use of the non-cost-effective treatment B will be limited. On the other hand, in the absence of research the specified (considered 'true') cost-effectiveness will still stand (in the example, A will still be the most cost-effective treatment), although this will not be revealed and will not induce a change in practice (i.e. the opportunity to increase the use of the cost-effective treatment A and restrict the use of the non-cost-effective treatment B will be missed).

In this way, an estimate of the 'cost-effectiveness' of a research study given each broad possible outcome and the change in clinical practice associated with it can be obtained by comparing costs and benefits under the 'with research' and 'without research' states using the formula:

$$ICER_i = \frac{C_{st} + C_{r,i} - C_{nr,i}}{B_{r,i} - B_{nr,i}}$$

Here, i is an indicator for a possible outcome; r and nr index the 'with research' and 'without research' situations, respectively; C_{st} represents the cost of the proposed research study; $C_{r,i}$ and $C_{nr,i}$ are the costs associated with outcome i in the 'with research' and

'without research' situations and, similarly, $B_{r,i}$ and $B_{nr,i}$ are the benefits (e.g. QALYs) under outcome i, with and without research, respectively. These cost and benefits can be projected to the population of current and future patients by multiplying them by the number of eligible patients per year and the number of years the information produced by research is expected to be useful, that is, the 'time horizon' of the study. Expressed in terms of incremental net monetary benefits (INMBs), the formula can be written as:

$$INMB_i = \lambda \times \left(B_{r,i} - B_{nr,i}\right) - \left(C_{st} + C_{r,i} - C_{nr,i}\right)$$

Here, λ stands for the ceiling ratio, a hypothetical value of a decision-maker's willingness to pay for an additional unit of benefit. The above formulae give the benefits accruing from each possible outcome i, but, in practice, only one of these outcomes may come true.

Although it is not known in advance which of the possible outcomes represents the true cost-effectiveness of the treatments, it is possible to obtain a summary measure of the proposed study's payoff by forming 'combinations'. In doing so, specified possible outcomes are attached 'likelihood weights' representing the probability of a particular outcome to transpire[67]. For example, an 'optimistic' combination may assume that a 'favourable' result for a treatment is more likely to be observed, while a 'pessimistic' combination may attach greater likelihood to an 'unfavourable' outcome.

Following the above notation, the weighted cost per QALY associated with a combination is given by:

$$Weighted\ ICER_k = \frac{\sum_i (p_i \times (C_{st} + C_{r,i} - C_{nr,i}))}{\sum_i (p_i \times (B_{r,i} - B_{nr,i}))}$$

or, equivalently, expressed in INMBs, by:

$$INMB_k = \lambda \times \sum_i p_i \times (B_{r,i} - B_{nr,i}) - \sum_i p_i \times (C_{st} + C_{r,i} - C_{nr,i})$$

where k is an index for combinations and p_i is the probability of observing study outcome i. Costs and benefits expected to accrue over a specified time horizon after research results are disseminated need to be discounted in order to account for positive time preference.

6.2. 'Payback' analysis for NSCLC

The section reports the application of PATHS to the NSCLC case study, with a view to estimating the costs and benefits expected to arise from a trial comparing Gem+Cisp against Gem+Carb. In line with the core methodology of 'payback', the analysis involves: a) specifying possible research outcomes, b) determining the possible change in clinical practice in the light of each possible outcome, and c) estimating the stream of costs and benefits expected to arise under each possible outcome and the change in practice associated with it. The analysis assumes a ceiling ratio value of £30,000 per additional QALY. This is seen as the cost per QALY value above which NICE will need an increasingly stronger case in order to adopt a treatment[114]. Different ceiling ratios are equally applicable. A base case analysis is reported first, followed by additional analyses.

6.2.1. Base case analysis

This section reports the results obtained from the base case analysis, which follows the core steps of the PATHS model[67].

6.2.1.A. Required information

For this application, information was needed on the discounted number of patients eligible for the treatments of interest (gemcitabine plus cisplatin (Gem+Cisp) and gemcitabine plus carboplatin (Gem+Carb) as first-line treatments in NSCLC) for the period over which information produced from further research would be expected to be useful. The number of new cases of lung cancer per years has been estimated at approximately 41,500[259]. Of those cases, about 9 percent (about 3830 patients per year) are patients with advanced NSCLC eligible for first-line platinum-based chemotherapy[260]. The period over which information from the trial was assumed to be useful, that is, the relevant time horizon, was set at five years, starting from the time that results are expected to be disseminated. According to the BTOG-2 trial protocol, results were expected to become available in 2011. The number of eligible patients was discounted at an annual rate of 3.5 percent[114] over five years to give a total number of approximately 13,800 eligible patients.

In addition, information is needed on the proportion of patients treated with Gem+Cisp and Gem+Carb in the year when the trial results are expected to become available. No such estimates were identified in the literature, but expert opinion suggested that, in 2011, the treatments were used in equal proportions [personal communication with Mr Andrew

Stanley, pharmacist, NHS Pan Birmingham Cancer Network, 11-11-2011], with half of the eligible patients receiving Gem+Cisp and the rest receiving Gem+Carb. Last, the cost of the proposed trial itself was £336,721 and it was taken from the BTOG-2 trial proposal submitted for funding.

6.2.1.B. Specification of possible outcomes

In a 'real world' situation, a clinical trial would give clinical evidence on a treatment's effectiveness (e.g. probability of cancer progression at one-year follow up). This evidence would be expected to be used as input in a cost-effectiveness analysis, which would establish the cost-effectiveness of the treatment in question. The outcome of the cost-effectiveness analysis would be expected to have an impact on clinical practice.

In a prospective assessment, different possible outcomes of the cost-effectiveness analysis need to be specified in advance. Thus, the first step in this analysis involved specifying such possible outcomes, which should be broad enough to cover different eventualities and should be associated with an impact on clinical practice[90;98]. To determine the link between possible outcomes and direction of change in clinical practice, the analysis adopted a stance similar to that of NICE, assuming that a treatment is likely to be recommended and used if it is associated with additional benefits at a cost lower than a presumed 'threshold' value—here, £30,000 per QALY. With these considerations in mind, the specified outcomes were:

a. ‘Favourable’: The BTOG-2 trial shows results (disease progression rates) such that, when these results are entered in the NSCLC decision model, Gem+Carb is cost-effective compared to Gem+Cisp (i.e. $ICER_{Gem+Carb}$ well below £30,000 per additional QALY);

b. ‘Inconclusive’: The BTOG-2 trial shows results (disease progression rates) such that, when these results are entered in the NSCLC decision model, Gem+Carb appears non-conclusively more (or less) cost-effective than Gem+Cisp (i.e. $ICER_{Gem+Carb}$ near £30,000 per additional QALY), and

c. ‘Unfavourable’: The BTOG-2 trial shows results (disease progression rates) such that, when these results are entered in the NSCLC decision model, Gem+Carb appears not cost-effective compared to Gem+Cisp (i.e. $ICER_{Gem+Carb}$ well above £30,000 per additional QALY).

The second step involved specifying the trial results, in terms of disease progression rates, that a trial would show if the specified possible outcomes were to transpire. Hypothetical trial results of the probability of progression at one-year follow up for Gem+Carb were specified in order to match the ‘favourable’, ‘inconclusive’ and ‘unfavourable’ possible outcomes. In the NSCLC model, the probability of Gem+Carb progression at each model cycle was drawn from a Weibull distributions with shape parameter α and scale parameter β. To adjust the progression probabilities so that they match the specified possible outcomes (i.e. ‘favourable’, ‘inconclusive’, ‘unfavourable’), the scale parameter of the Weibull time-to-progression model for Gem+Carb was modified accordingly. The scale and

shape parameters for the Weibull models for disease progression are given in Table 6.1, with the parameters changed to match each possible outcome appearing in bold. The new scale parameters of the Weibull model translate to Gem+Carb progression probability at one-year follow up of 0.64, 0.72 and 0.87 for the 'favourable', 'inconclusive' and 'unfavourable' research outcome, respectively.

Table 6.1: Parameters of Weibull distribution for disease progression for different possible outcomes (NSCLC)

	Favourable outcome		**Inconclusive outcome**		**Unfavourable outcome**	
	Alpha	Beta	Alpha	Beta	Alpha	Beta
Progression (Gem+Cisp)	1.404	12.17	1.404	12.17	1.404	12.17
Progression (Gem+Carb)	1.287	**17.00**	1.287	**14.26**	1.287	**9.85**

The values of costs and QALYs for Gem+Carb resulting from the specified transition probabilities appear in Table 6.2 in bold. In must be noted that, for the 'unfavourable' scenario, transitions may be specified so that Gem+Carb could be either more effective than Gem+Cisp but not cost-effective (i.e. resulting in an ICER higher that the assumed value of the ceiling ratio) or less effective and not cost-effective (i.e. dominated by Gem+Cisp). For the base case analysis, a choice of transition probabilities that make Gem+Carb dominated was made, as this situation reflects more closely the existing evidence, that is, the fact that in the NSCLC model Gem+Carb was shown to be dominated.

Table 6.2: Cost-effectiveness results based on 'favourable', 'inconclusive' and 'unfavourable' outcomes (NSCLC)

Outcome	Gem+Cisp		Gem+Carb		ICER (Gem+Carb vs. Gem+Cisp)
	Total cost	Total QALYs	Total cost	Total QALYs	
Favourable to Gem+Carb	£5830	0.583	**£6762**	**0.639**	£16,700
Inconclusive	£5830	0.583	**£6710**	**0.612**	£30,280
Unfavourable to Gem+Carb	£5830	0.583	**£6560**	**0.567**	Dominated

6.2.1.C. Change in clinical practice

In a situation where research has been carried out and given evidence on a treatment's effectiveness and cost-effectiveness, this evidence is expected to be disseminated and affect clinical practice. The exact magnitude of the change in clinical practice, in terms of changes in the prescription share (i.e. uptake) of the treatment of interest is difficult to predict. In this study, guesses on treatments' future uptake in the light of different possible outcomes were formed through discussion with experts in cancer services commissioning based at the NHS Pan Birmingham Cancer Network. These were Mrs K. Metcalf (network director, responsible for the network's overall commissioning strategy), Dr D. Ford (network clinical lead, oncologist) and Mr A. Stanley (network pharmacist, responsible for the network's formulary and commissioning of chemotherapy treatments).

As noted above, under current (2011) practice, Gem+Cisp and Gem+Carb are prescribed in equal proportions (50 percent each). In the absence of further evidence from a trial, the experts felt that prescription shares were unlikely to change in the future. In the event of an outcome 'favourable' to Gem+Carb and appropriate dissemination of results, it was agreed

that the treatment's use would be likely to increase to a level of approximately 75 percent, with Gem+Cisp being prescribed to the rest 25 percent of the eligible population. Such a marked increase was thought to be possible given the fact that no significant barriers exist in switching between these two treatments. Following an 'inconclusive' outcome, the experts were in agreement that there would be no change in clinical practice and the current prescription patterns would continue to hold. Last, observing an 'unfavourable' outcome for Gem+Carb (that is, a 'favourable' outcome for Gem+Cisp) would be expected to lead to a decrease in the use of Gem+Carb to 25 percent of the eligible population. The hypothesised values of treatment uptake for the situations with and without research are summarised in Table 6.3.

Table 6.3: Treatment uptake following the specified research outcomes (NSCLC)

Possible outcome	Prescription in 'with research' state	Prescription in 'without research' state
Favourable to Gem+Carb (i.e. Gem+Carb cost-effective)	Gem+Cisp: 25% Gem+Carb: 75%	Gem+Cisp: 50% Gem+Carb: 50%
Inconclusive (i.e. Gem+Carb not clearly cost-effective)	Gem+Cisp: 50% Gem+Carb: 50%	
Unfavourable to Gem+Carb (i.e. Gem+Carb not cost-effective)	Gem+Cisp: 75% Gem+Carb: 25%	

6.2.2. Results of the base case analysis

Costs and benefits expected to arise to the population of eligible treatments under different possible outcomes in the 'with research' state can be calculated by multiplying the costs and QALYs under each outcome by each treatment's hypothesised prescription share. Here, the likelihood of observing a possible outcome is the same across these outcomes; in other

words, no likelihood weights are attached to indicate that any of these outcomes is more likely to occur. Results of this non-weighted analysis are given in Table 6.4.

Table 6.4: Non-weighted costs and benefits of research in base case analysis (NSCLC)

	Favourable outcome	**Inconclusive outcome**	**Unfavourable outcome**
With research			
Cost	£90,081,664	£86,511,060	£82,960,067
Trial cost	£336,721	£336,721	£336,721
QALYs	8617	8241	7985
Without research			
Cost	£86,868,565	£86,511,060	£85,477,767
QALYs	8425	8241	7931
Net implications			
Net cost	£3,549,820	£336,721	-£2,180,979
Net QALYs	192	0	55
Cost per QALY	**£18,450 per additional QALY**	**Costs for no additional QALYs**	**Cost savings for additional QALYs**
$NMB_{\text{with research}}$	£168,103,680	£160,370,345	£156,267,944
$NMB_{\text{without research}}$	£165,881,317	£160,707,066	£152,446,181
INMBs (£30,000 per QALY)	**£2,222,363**	**-£336,721**	**£3,821,763**

With research and under a specific outcome, the cost (or QALYs) are found by multiplying the discounted eligible population by the sum resulting from the combination of costs (or QALYs) associated with the outcome and the hypothesised uptake in clinical practice. For example, under a 'favourable' outcome, the cost and QALYs 'with research' (r) are found as:

$$Cost_{r,favourable}$$

$$= Eligible\ population$$

$$\times \left(Cost_{Gem+Cisp,favourable} \times Prescription_{Gem+Cisp,favourable}\right.$$

$$\left. +\ Cost_{Gem+Carb,favourable} \times Prescription_{Gem+Carb,favourable}\right)$$

$$= 13{,}797\ \times (£5830\ \times 0.25 + £6762\ \times 0.75) = £90{,}081{,}664$$

$$QALYs_{r,favourable}$$

$$= Eligible\ population$$

$$\times \left(QALYs_{Gem+Cisp,favourable} \times Prescription_{Gem+Cisp,favourable}\right.$$

$$+\ QALYs_{Gem+Carb,favourable}$$

$$\left. \times Prescription_{Gem+Carb,favourable}\right)$$

$$= 13{,}797\ \times (0.58\ \times 0.25 + 0.64\ \times 0.75) = 8617$$

Under the same outcome, in the absence of research (nr), clinical practice is assumed to remain as it is, with half of the patients taking Gem+Carb and the rest taking Gem+Cisp. In this situation, the costs and QALYs are estimated as follows:

$$Cost_{nr,favourable}$$

$$= Eligible\ population$$

$$\times \left(Cost_{Gem+Cisp,favourable} \times Prescription_{Gem+Cisp,current\ prescription}\right.$$

$$\left. +\ Cost_{Gem+Carb,favourable} \times Prescription_{Gem+Carb,current\ prescription}\right)$$

$$= 13{,}797\ \times (£5830\ \times 0.5 + £6762\ \times 0.5) = £86{,}868{,}565$$

$$QALYs_{nr,favourable}$$

$$= Eligible\ population$$

$$\times \left(QALYs_{Gem+Cisp,favourable} \times Prescription_{Gem+Cisp,current\ prescription}\right.$$

$$\left. +\ QALYs_{Gem+Carb,favourable} \times Prescription_{Gem+Carb,current\ prescription}\right)$$

$$= 13{,}797\ \times (0.58\ \times 0.5 +\ 0.64 \times 0.5) = 8425$$

In this way, the costs and QALYs with and without further research (here, the proposed BTOG-2 trial) can be estimated for each of the three possible outcomes. On the basis of these costs and QALYs, the ICER and INMBs for the 'favourable' outcome are shown below:

$$ICER_{favourable} = \frac{£336{,}721\ +\ £90{,}081{,}664\ -\ £86{,}868{,}565}{8617\ -\ 8425} = £18{,}450\ per\ QALY$$

and

$$INMB_{favourable} = £30{,}000\ \times\ 192\ -\ £3{,}549{,}820\ =\ £2{,}222{,}363$$

The cost per QALY and the INMBs associated with each possible outcome have been calculated in the same way.

Under the 'favourable' outcome for Gem+Carb, carrying out the BTOG-2 trial is estimated to result in greater costs and more QALYs than in a situation without the trial (Table 6.4). This is because under this outcome more patients will be offered Gem+Carb, which, in this case, is assumed to be more costly and more effective than Gem+Cisp. As obtaining an additional QALY in this case is expected to cost less than the ceiling ratio of £30,000, conducting

research is expected to result in positive INMBs of about £2.22 million (ICER of £18,450 per QALY). Given an 'inconclusive' outcome, conducting the trial would result in no additional QALYs (as no change in prescription is expected to take place) for an extra cost equal to the cost of the trial. As a result, this outcome is associated with negative INMBs of £336,700. Last, under the 'unfavourable' outcome, conducting the trial is associated with an increase in QALYs and cost savings. This is because, under the specific outcome, clinical practice will move towards restricting the use of the more costly and less effective Gem+Carb. In this situation, carrying out research appears particularly appealing, leading to additional NMBs of approximately £3.82 million.

To reflect the fact that different outcomes are associated with different likelihoods of occurrence, each possible outcome was assigned a likelihood weight and formed different combinations. In line with methods used in previous applications of the PATHS model[67;116], three combinations were formed:

a. 'Optimistic', where the probability of observing a 'favourable', 'inconclusive' and 'unfavourable' outcome is 0.5, 0.25 and 0.25, respectively;

b. 'Neutral', where each outcome has an one-third probability of being observed, and

c. 'Pessimistic', where the probability of observing a 'favourable', 'inconclusive' and 'unfavourable' outcome is 0.25, 0.25 and 0.5, respectively.

The weighted net costs, net QALYs and INMBs for each of these combinations are given in Table 6.5, Table 6.6 and Table 6.7. In each of these tables—and in all tables of the same form throughout this chapter—the second column gives the attached likelihood weight, while the third and fourth columns give the weighted costs and QALYs, respectively. For each possible outcome, the weighted cost (or, equivalently, the weighted QALYs) is calculated by multiplying the relevant net cost (net QALYs) as found in the non-weighted analysis (Table 6.4) by the respective likelihood weight (e.g. weighted cost of 'favourable' outcome in 'optimistic' combination = 3,549,820 × 0.5 = 1,774,910).

The last column in each table gives the weighted INMBs calculated as weighted QALYs multiplied by the ceiling ratio, minus weighted costs (e.g. weighted INMBs of 'favourable' outcome in 'optimistic' combination = 96 × £30,000 − £1,774,910 = £1,111,181. Last, the total weighted INMBs for a specific combination are calculated as the sum of the weighted INMBs of each outcome and are given in the bottom-right cell of each table. According to the results, carrying out the proposed BTOG-2 trial is estimated to lead to positive INMBs of £1.98 million, £1.88 million and £2.38 million under the 'optimistic', 'neutral' and 'pessimistic' combinations, respectively.

Table 6.5: Weighted costs and benefits of research in base case analysis ('optimistic' combination) (NSCLC)

Optimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted $INMB_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.50	£1,774,910	96	£1,111,181
Inconclusive	0.25	£84,180	0	-£84,180
Unfavourable	0.25	-£545,245	14	£955,441
Sum of weighted net costs and QALYs	-	£1,313,846	110	-
Weighted ICER and INMBs	£11,958 per additional QALY		£1,982,442	

Table 6.6: Weighted costs and benefits of research in base case analysis ('neutral' combination) (NSCLC)

Neutral combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted $INMB_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.33	£1,171,441	63	£733,380
Inconclusive	0.33	£111,118	0	-£111,118
Unfavourable	0.33	-£719,723	18	£1,261,182
Sum of weighted net costs and QALYs	-	£562,836	82	-
Weighted ICER and INMBs	£6,902 per additional QALY		£1,883,444	

Table 6.7: Weighted costs and benefits of research in base case analysis ('optimistic' combination) (NSCLC)

Pessimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted $INMB_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.25	£887,455	48	£555,591
Inconclusive	0.25	£84,180	0	-£84,180
Unfavourable	0.50	-£1,090,489	27	£1,910,882
Sum of weighted net costs and QALYs	-	-£118,854	75	-
Weighted ICER and INMBs	Cost savings of £118,854 and 75 additional QALYs		£2,382,292	

In summary, the results of this analysis suggest that conducting the trial and changing clinical practice according to its results would result in additional health benefits at a value lower than society (or a decision-maker) is assumed to be willing to pay for these benefits, resulting in positive NMBs. In this respect, according to the interpretation used in 'payback' studies[67;116], the BTOG-2 trial represents a 'cost-effective' use of resources.

6.2.3. Additional analyses

Highly uncertain parameters, such as hypothesised prescription shares and likelihood weights attached to outcomes, are expected to have an effect on 'payback' results. Different assumptions about these parameters were tested in sensitivity analyses. Probabilistic analysis was also carried out, where uncertain parameters were characterised by probability distributions. These analyses are described below.

6.2.3.A. Deterministic sensitivity analyses

One of the possibilities investigated in sensitivity analyses is that clinical practice may be more responsive to research results than it was assumed in the base case analysis. According to this scenario, following a 'favourable' outcome for Gem+Carb, there would be a greater uptake of the treatment in clinical practice (80 and 85 percent of the eligible patients would use Gem+Carb, as opposed to 75 percent in the base case) and, equivalently, 'unfavourable' results would lead to a greater decrease in the number of patients using Gem+Carb (20 and 15 percent, as opposed to 25 percent). This was estimated to lead to increased INMBs as compared to the base case results for all combinations, suggesting that,

if such an implementation pattern was prevalent, the option of carrying out research would be even more appealing (Table 6.8 and Table 6.9).

Table 6.8: Deterministic sensitivity analysis assuming higher uptake (80 percent) following 'favourable' results (NSCLC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
Gem+Carb current prescription share	50%		
Gem+Carb future prescription share	80%	50%	20%
INMBs optimistic combination	£2,446,275		
INMBs neutral combination	£2,326,803		
INMBs pessimistic combination	£2,926,095		

Table 6.9: Deterministic sensitivity analysis assuming higher uptake (85 percent) following 'favourable' results (NSCLC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
Gem+Carb current prescription share	50%		
Gem+Carb future prescription share	85%	50%	15%
INMBs optimistic combination	£2,910,107		
INMBs neutral combination	£2,770,163		
INMBs pessimistic combination	£3,469,897		

On the other hand, it is possible that clinical practice may change in a less marked way than that assumed in the base case analysis. In such a case, it would be expected that, following 'favourable' outcomes, there would be a lower rate of Gem+Carb uptake than in base case and, following 'unfavourable' outcomes, there would be a smaller decrease in the number of patients using Gem+Carb. The results suggested that carrying out research in this case

would still be beneficial, but it would result in slightly lower INMBs than those observed in the base case results (Table 6.10 and Table 6.11).

Table 6.10: Deterministic sensitivity analysis assuming lower uptake (70 percent) following 'favourable' results (NSCLC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
Gem+Carb current prescription share	50%		
Gem+Carb future prescription share	70%	50%	30%
INMBs optimistic combination	£1,518,609		
INMBs neutral combination	£1,440,084		
INMBs pessimistic combination	£1,838,489		

Table 6.11: Deterministic sensitivity analysis assuming lower uptake (65 percent) following 'favourable' results (NSCLC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
Gem+Carb current prescription share	50%		
Gem+Carb future prescription share	65%	50%	35%
INMBs optimistic combination	£1,054,777		
INMBs neutral combination	£996,725		
INMBs pessimistic combination	£1,294,687		

A further possibility explored in deterministic sensitivity analyses relates to the current prescription share of Gem+Carb. Assuming that this share is actually higher than the base case estimate (i.e. 60 percent of the eligible patients receive Gem+Carb as opposed to 50 percent in the base case analysis) resulted in higher INMBs than in the base case for the 'neutral' and 'pessimistic' combinations, and in slightly lower INMBs for the 'optimistic'

combination. On the other hand, setting the current prescription percentage of Gem+Carb to a lower value than that in the base case (40 percent) led to lower, although still positive, INMBs for the 'neutral' and 'pessimistic' combinations, and higher INMBs for the 'optimistic' combination (Table 6.12 and Table 6.13). This indicates that more pronounced moves towards extending the use of a cost-effective treatment—or restricting the use of non-cost-effective treatments—are anticipated to result in greater numbers of benefits.

Table 6.12: Deterministic sensitivity analysis assuming higher current prescription (60 percent) for Gem+Carb (NSCLC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
Gem+Carb current prescription share	60%		
Gem+Carb future prescription share	75%	60%	25%
INMBs optimistic combination	£1,886,474		
INMBs neutral combination	£2,094,564		
INMBs pessimistic combination	£2,958,080		

Table 6.13: Deterministic sensitivity analysis assuming lower current prescription (40 percent) for Gem+Carb (NSCLC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
Gem+Carb current prescription share	40%		
Gem+Carb future prescription share	75%	40%	25%
INMBs optimistic combination	£2,078,410		
INMBs neutral combination	£1,672,323		
INMBs pessimistic combination	£1,806,504		

In the base case analysis, the 'unfavourable' outcome was set so that Gem+Carb is more expensive and less effective than Gem+Cisp, that is, it is dominated by the latter. However, an alternative scenario for the 'unfavourable' outcome may be that Gem+Carb is neither cost-effective at £30,000 per QALY, nor is it dominated (i.e. it is more costly but also slightly more effective than Gem+Cisp). This outcome would transpire if the BTOG-2 trial showed a lower probability of progression at one-year follow-up for Gem+Carb than that in the base case, for instance 0.75 as opposed to 0.87 in the base case. In such a case, Gem+Carb would be associated with an additional cost of £870 and 0.02 additional QALYs compared to Gem+Cisp.

Given the alternative 'unfavourable' scenario of Gem+Carb not being dominated, a situation with research would result in cost savings (due to restricting the use of the more costly Gem+Carb) and a loss in QALYs (due to limiting the prescription of the more effective Gem+Carb) as compared to the situation without research. Here, cost savings would offset the loss of QALYs to give additional non-weighted INMBs of £422,800. The non-weighted INMBs for the 'unfavourable' outcome under this scenario are lower than those for the same outcome under the base case scenario, indicating that there is more value in restricting the use of a dominated treatment, than there is in limiting the prescription of a treatment that is effective, but not cost-effective. This is also reflected on the weighted results shown in Table 6.14 which are lower than the equivalent results in the base case analysis.

Table 6.14: Deterministic sensitivity analysis for alternative 'unfavourable' outcome (NSCLC)

	Optimistic combination	Neutral combination	Pessimistic combination
Alternative scenario for unfavourable outcome (Gem+Carb not dominated)	£1,132,705	£761,791	£682,818

Last, the base case time horizon—the period over which information from a trial is expected to be useful—was replaced by alternative values. Time horizon has a significant bearing on results, as it determines the number of future patients that are expected to benefit from decision-making in the light of the generated information. Assuming that information produced by the BTOG-2 trial would be useful for ten years almost doubled the expected gains from the trial, making the case for funding and carrying out the BTOG-2 study even stronger (Table 6.15).

Table 6.15: Deterministic sensitivity analysis assuming different time horizons (NSCLC)

Time horizon	Eligible patients	Optimistic combination	Neutral combination	Pessimistic combination
One year	2,856	£143,304	£125,484	£226,066
Five years	13,797	£1,982,442	£1,883,444	£2,382,292
Ten years	25,270	£3,910,826	£3,726,711	£4,643,152

6.2.3.B. Gradual change in uptake analysis

The analyses above are based on the assumption that, once research results are disseminated, changes in treatments' prescription share take place almost instantly. However, in practice, a treatment's uptake is likely to change progressively, over time. With this in mind, further sensitivity analyses were carried out to explore the effect of gradual

change in treatments' uptake. In this illustrative analysis it was assumed that uptake changes in a linear fashion over the five-year time horizon. On this premise, prescription shares would assume the values given in Table 6.16 and would result in the non-weighed results given in Table 6.17 below.

Table 6.16: Assumed prescription share for gradual change in uptake (NSCLC)

Year	Favourable outcome		Inconclusive outcome		Unfavourable outcome	
	Gem+Carb	Gem+Cisp	Gem+Carb	Gem+Cisp	Gem+Carb	Gem+Cisp
Current	50%	50%	50%	50%	50%	50%
Year 1	55%	45%	50%	50%	45%	55%
Year 2	60%	40%	50%	50%	40%	60%
Year 3	65%	35%	50%	50%	35%	65%
Year 4	70%	30%	50%	50%	30%	70%
Year 5	75%	25%	50%	50%	25%	75%

According to this analysis, for a 'favourable' outcome, a hypothetical situation where the BTOG-2 trial has taken place would be associated with additional NMBs of £1.16 million. Compared to the 'favourable' outcome of the base case analysis, these benefits are lower by about £1.07 million. Under an 'inconclusive' outcome, there is a loss of £336,700 in NMBs, representing the cost of the trial. Last, an 'unfavourable' outcome would result in additional NMBs of £2.09 million, which is approximately £1.73 million lower than the NMBs of the equivalent outcome in the base case analysis.

As expected, the trend shows that delays in moving towards a treatment that appears cost-effective has a negative impact on the benefits expected to accrue to the eligible population.

Table 6.17: Non-weighted costs and benefits of research for gradual change in uptake (NSCLC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
With research			
Cost	£88,744,040	£86,511,060	£84,008,195
Trial cost	£336,721	£336,721	£336,721
QALYs	8537	8241	7963
Without research			
Cost	£86,868,565	£86,511,060	£85,477,767
QALYs	8425	8241	7931
Net implications			
Net cost	£2,212,195	£336,721	-£1,132,851
Net QALYs	112	0	32
Cost per QALY	**£19,698 per additional QALY**	**Costs for no additional QALYs**	**Cost savings for additional QALYs**
$NMB_{with\ research}$	£167,038,324	£160,370,345	£154,536,752
$NMB_{without\ research}$	£165,881,317	£160,707,066	£152,446,181
INMBs (£30,000 per QALY)	**£1,157,007**	**-£336,721**	**£2,090,571**

This trend can be also seen in the weighted results of this analysis (Table 6.18, Table 6.19 and Table 6.20), which appear consistently lower than the corresponding results in the base case analysis. It must be noted that results in this analysis are highly dependent on assumptions around the pattern of change in clinical practice over the five-year time period.

Table 6.18: Weighted costs and benefits of research for gradual change in uptake ('optimistic' combination)(NSCLC)

Optimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.50	£1,106,098	56	£578,503
Inconclusive	0.25	£84,180	0	-£84,180
Unfavourable	0.25	-£283,213	8	£522,643
Sum of weighted net costs and QALYs	-	£907,065	64	-
Weighted ICER and INMBs	£14,143 per additional QALY			£1,016,966

Table 6.19: Weighted costs and benefits of research for gradual change in uptake ('neutral' combination)(NSCLC)

Neutral combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.33	£730,024	37	£381,812
Inconclusive	0.33	£111,118	0	-£111,118
Unfavourable	0.33	-£373,841	11	£689,888
Sum of weighted net costs and QALYs	-	£467,301	48	-
Weighted ICER and INMBs	£9818 per additional QALY			£960,583

Table 6.20: Weighted costs and benefits of research for gradual change in uptake ('pessimistic' combination)(NSCLC)

Pessimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.25	£553,049	28	£289,252
Inconclusive	0.25	£84,180	0	-£84,180
Unfavourable	0.50	-£566,426	16	£1,045,285
Sum of weighted net costs and QALYs	-	£70,803	44	-
Weighted ICER and INMBs	£1608 additional QALY			£1,250,357

6.2.3.C. Probabilistic sensitivity analysis

Further work was undertaken to illustrate the potential use of probabilistic sensitivity analysis in 'payback'. Such analysis aimed to assess the effect of uncertainty around the hypothesised change in clinical practice and the likelihood weights associated with possible outcomes. For change in clinical practice, 1000 values of the possible levels of uptake were drawn from a uniform distribution in a way that drawn values were limited to the intervals between 0.5 (current prescription share) and 1 (maximum possible prescription share), and 0.5 (current prescription share), and 0 (minimum possible prescription share) for a 'favourable' and 'unfavourable' outcome, respectively. As prescription under an inconclusive outcome is assumed not to change, this was fixed at its current value (0.5). Results were calculated for each of the 1000 drawn prescription values. The average costs, QALYs and INMBs together with 95 percent confidence intervals are given in Table 6.21.

Table 6.21: Non-weighted costs and benefits of research in probabilistic analysis (NSCLC)

	Favourable outcome (Lower and upper 95% confidence intervals)	**Inconclusive outcome** (Lower and upper 95% confidence intervals)	**Unfavourable outcome** (Lower and upper 95% confidence intervals)
With research			
Cost	**£90,143,177** (£90,029,124 to £90,257,230)	**£86,511,060**	**£83,008,830** (82,918,557 to £83,099,105)
Trial cost	**£336,721**	**£336,721**	**£336,721**
QALYs	**8621** (8614 to 8627)	**8241**	**7984** (7982 to 7986)
Without research			
Cost	**£86,868,565**	**£86,511,060**	**£85,477,767**
QALYs	**8425**	**8241**	**7931**
Net implications			
Net cost	**£3,611,332** (£3,497,280 to £3,725,385)	**£336,721**	**-£2,132,215** (-£2,222,490 to -£2,041,941)
Net QALYs	**196** (189 to 203)	**0**	**54** (52 to 56)
Cost per QALY	**£18,417 per additional QALY**	**Cost for no additional QALYs**	**Cost savings for additional QALYs**
$NMB_{with\ research}$	**£168,152,672** (£168,061,834 to £168,243,509)	**£160,370,345**	**£156,187,402** (£156,038,296 to £156,366,507)
$NMB_{without\ research}$	**£165,881,317**	**£160,707,066**	**£152,446,181**
INMBs (£30,000 per QALY)	**£2,271,355** (£2,180,517 to £2,362,192)	**-£336,721**	**£3,741,221** (£3,592,115 to £3,890,327)

The results of this analysis appeared considerably similar to those in the base case analysis above; if further research revealed a 'favourable' or 'unfavourable' outcome for Gem+Carb and clinical practice changed accordingly, there would be additional NMBs, while in the case of an inconclusive outcome there would be a cost equal to the cost of the trial. It must be acknowledged that, the extent to which these results would differ to those in the base case analysis depends largely on the distribution assigned to prescription share values. 'Steep'

distributions that concentrate around the mean (base case) prescription share will naturally give values closer to the base case results.

Further probabilistic sensitivity analysis was carried out to propagate the uncertainty around likelihood weights used in forming combinations. Likelihood weights were considered to be random values, drawn from distributions with means equal to those in the base case and standard errors set, for illustration purposes, at one-quarter of their respective means. To account for the multinomial nature of the likelihood weights, where three independently drawn values need to sum up to one, likelihood weights were drawn from Dirichlet distributions using the normalized sum of independent gamma draws method[51;261]. Each of the drawn likelihood weights was combined with the probabilistic results obtained from varying the prescription rates to give 1000 estimates of the weighted net costs and QALYs for each combination.

Mean values of these estimates are given in Table 6.22, Table 6.23 and Table 6.24. The results of this analysis showed all combinations to be associated with positive INMBs ranging from about £1.88 million for the 'neutral' combination to £2.31 million for the 'pessimistic' combination.

Table 6.22: Weighted costs and benefits of research in probabilistic analysis ('optimistic' combination)(NSCLC)

Optimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.50	£1,783,070	97	£1,123,673
Inconclusive	0.25	£87,339	0	-£87,339
Unfavourable	0.25	-£529,777	13	£930,009
Sum of weighted net costs and QALYs	-	£1,340,632	110	-
Weighted ICER and INMBs	£12,162 per additional QALY			£1,966,343

Table 6.23: Weighted costs and benefits of research in probabilistic analysis ('neutral' combination)(NSCLC)

Neutral combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.33	£1,183,358	64	£743,478
Inconclusive	0.33	£112,327	0	-£112,327
Unfavourable	0.33	-£710,116	18	£1,246,941
Sum of weighted net costs and QALYs	-	£585,568	82	-
Weighted ICER and INMBs	£7130 per additional QALY			£1,878,092

Table 6.24: Weighted costs and benefits of research in probabilistic analysis ('pessimistic' combination)(NSCLC)

Pessimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.25	£906,743	49	£570,500
Inconclusive	0.25	£86,754	0	-£86,754
Unfavourable	0.50	-£1,041,717	26	£1,828,483
Sum of weighted net costs and QALYs	-	-£48,219	75	-
Weighted ICER and INMBs	Cost savings of £48,219 and 75 additional QALYs			£2,312,228

In summary, results from the base case and additional analyses suggest that, given the specific assumptions about possible results and change in clinical practice, funding and conducting the proposed BTOG-2 trial would results in additional benefits to the population.

6.3. 'Payback' analysis for HRPC

This section gives the methods and results of applying the PATHS methodology to the HRPC case study. In this case study, the TRAPEZE phase III randomised controlled trial was proposed to assess the effectiveness of four chemotherapy combinations (DP, DP+ZA, DP+Sr89, DP+ZA+Sr89).

Applications of PATHS to case studies involving comparisons between more than two treatments are rare and require stronger assumptions than those for two-treatment comparisons. Given this, two separate analyses were carried out and are reported below. The first analysis is concerned with a two-treatment analysis, focusing on the treatments which were non-dominated in the cost-effectiveness analysis reported in Chapter 5 (DP and DP+Sr89). The second analysis looks into the application of PATHS to the four-arm TRAPEZE trial. Both the analyses are primarily intended as an illustration of the employed methods.

6.3.1. Base case analysis of the two-treatment HRPC case study

The section reports the methods and results of the base case analysis applied to the two-treatment version of the HRPC case study.

6.3.1.A. Required information

The annual number of new cases of prostate cancer has been estimated at approximately 40,800[234], nearly half of which cases are expected to eventually develop hormone-refractory disease[239]. About 50 percent of these cases are expected to be eligible for first-line chemotherapy [personal communication with Professor N. James, Oncologist, University of Birmingham, 16-02-2011] and, of those patients, approximately 33 percent will be candidates for docetaxel-based treatment[188]. This results in about 3330 eligible patients per year. Given the emergence of new chemotherapy agents, such as cabazitaxel, DP-based treatments may be soon superseded. Thus, based on expert opinion, the time horizon for which the information generated by the proposed trial is expected to be relevant was chosen to be two years [personal communication with Professor N. James, Oncologist, University of Birmingham, 16-02-2011]. Results from the TRAPEZE phase III trial were expected to be disseminated in 2013. Given this, the discounted number of patients affected by the adoption decision over the hypothesised time horizon after research results are disseminated was estimated at 5101 patients.

No estimates of the possible prescription shares of DP and DP+Sr89 in 2013 were identified. Given the fact that DP+Sr89 is not provided routinely to patients and on the assumption that in the absence of further information practice will remain as in 2011, experts suggested that in 2013 the treatment may be given to as little as 5 percent of the eligible population [personal communication with Ms K. Metcalf and Mr A. Stanley, NHS Pan Birmingham

Cancer Network, 11-11-2011]. This value represents the 'current' prescription share for DP+Sr89.

The fixed cost of the trial (i.e. expenditure for researchers' salaries and facilities) amounted at £627,156. For the analysis involving only the DP and DP+Sr89 arms, the excess cost due to trial patients receiving treatments other than usual care provided by the NHS (DP) was estimated at £57,720, resulting in a total trial cost of £684,876. As before, the value of the ceiling ratio used is £30,000 per additional QALY.

6.3.1.B. Specification of possible outcomes

Similarly to the NSCLC case study, three broad possible outcomes were specified:

a. 'Favourable' outcome: Trial shows results (disease progression rates) such that, when these results are entered in the HRPC decision model, DP+Sr89 is cost-effective compared to DP (i.e. $ICER_{DP+Sr89}$ well below £30,000 per additional QALY);

b. 'Inconclusive' outcome: Trial shows results (disease progression rates) such that, when these results are entered in the HRPC decision model, DP+Sr89 appears non-conclusively more (or less) cost-effective than DP (i.e. $ICER_{DP+Sr89}$ near £30,000 per additional QALY), and

c. 'Unfavourable' outcome: Trial shows results (disease progression rates) such that, when these results are entered in the HRPC decision model, DP+Sr89 appears not cost-effective compared to DP (i.e. $ICER_{DP+Sr89}$ well above £30,000 per additional QALY).

Values representing the probability of a patient on DP+Sr89 moving from a progression-free to a progressive health state needed to be specified in order to match the 'favourable', 'inconclusive' and 'unfavourable' outcomes. In the HRPC model, the probability of transition between these states under DP+Sr89 was derived from observed counts of patient transitions taken from the TRAPEZE phase II trial and was calculated to be 0.05. To match the specified outcomes, new values of this transition were specified to be 0.04, 0.12 and 0.14 for the 'favourable', 'inconclusive' and 'unfavourable' outcomes, respectively. Given these transition probabilities, the new costs and QALYs associated with DP+Sr89 appear in Table 6.25 in bold.

Table 6.25: Cost-effectiveness results based on 'favourable', inconclusive and 'unfavourable' outcomes (two-treatment HRPC)

Outcome	DP		DP+Sr89		ICER (DP+Sr89 vs. DP)
	Total Cost	Total QALYs	Total Cost	Total QALYs	
Favourable to DP+Sr89	£8949	0.81	**£10,164**	**0.99**	£6602
Inconclusive	£8949	0.81	**£10,190**	**0.85**	£30,395
Unfavourable to DP+Sr89	£8949	0.81	**£10,193**	**0.83**	£58,270

6.3.1.C. Change in clinical practice

As before, possible change in clinical practice, in terms of treatments' prescription shares, was determined with the help of experts from the NHS Pan Birmingham Cancer Network (Mrs K. Metcalf, Dr D. Ford and Mr A. Stanley). Under 'current' (2013) practice, only a limited number of the eligible patients, about 5 percent, are expected to receive DP+Sr89. This is because of uncertainty about the treatment's clinical and cost-effectiveness, as well as due to limitations in the infrastructure for administering strontium-89 radioisotope fractions. For the purposes of this two-treatment analysis, it was assumed that other chemotherapy treatments (DP+ZA, DP+ZA+Sr89) were given to a negligible proportion of the population.

In the absence of a further trial, experts agreed that clinical practice would be unlikely to change and the great majority of patients would still receive DP. In the light of a 'favourable' outcome for DP+Sr89, there was a general agreement that DP+Sr89 uptake would increase, although the treatment's prescription share would be unlikely to reach a high level within the specified time horizon, mostly due to practical obstacles around the wider introduction and use of the treatment. Under the inconclusive outcome, the expectation was that clinical practice will remain as it was before the results became available, with the great majority of patients receiving the well-established and easily accessible DP. Last, under an 'unfavourable' outcome, it was expected that even fewer patients would be treated with DP+Sr89. The assumed changes in treatment uptake associated with the 'with research' and 'without research' states are given in Table 6.26.

Table 6.26: Treatment uptake following the specified research outcomes (two-treatment HRPC)

Outcome	Prescription in 'with research' state	Prescription in 'without research' state
Favourable to DP+Sr89 (i.e. DP+Sr89 cost-effective)	DP: 60% DP+Sr89: 40%	DP: 95% DP+Sr89: 5%
Inconclusive (i.e. DP+Sr89 not clearly cost-effective)	DP: 95% DP+Sr89: 5%	
Unfavourable to DP+Sr89 (i.e. DP+Sr89 not cost-effective)	DP: 99% DP+Sr89: 1%	

6.3.2. Results of the base case analysis for the two-treatment HRPC case study

As in the PATHS application to NSCLC, the cost and QALYs associated with the 'with research' and the 'without research' states in HRPC can be estimated according to the cost-effectiveness of each treatment and the proportion of the eligible population expected to receive each treatment in the view of the specified possible outcomes. The calculated costs, QALYs and INMBs for each possible outcome are given in Table 6.27.

The results show that, given an outcome 'favourable' to DP+Sr89, carrying out research would be associated with a gain of 329 QALYs at an extra cost of about £2.85 million, suggesting positive INMBs of about £7 million. On the other hand, observing an 'unfavourable' outcome for DP+Sr89 (that is, DP+Sr89 being slightly more effective and more costly than DP, but overall not being cost-effective at £30,000 per QALY) and switching clinical practice away from this treatment would result in additional costs (largely due to the cost of the trial) and a few QALYs forgone (due to the fact that less people would be taking the slightly more effective DP+Sr89). The latter case is associated with negative INMBs of about £561,800. Under the 'inconclusive' scenario, as always, there are costs equal to the

expenditure for the trial for no additional QALYs, which translates to negative INMBs of £684,900.

Table 6.27: Non-weighted costs and benefits of research in base case analysis (two-treatment HRPC)

	Favourable outcome	**Inconclusive outcome**	**Unfavourable outcome**
With research			
Cost	£48,129,173	£45,966,706	£45,713,552
Trial cost	£684,876	£684,876	£684,876
QALYs	4489	4124	4114
Without research			
Cost	£45,960,005	£45,966,706	£45,967,268
QALYs	4160	4124	4119
Net implications			
Net costs	£2,854,044	£684,876	£431,161
Net QALYs	329	0	-4
Cost per QALY	**£8686 per additional QALY**	**Costs for no additional QALYs**	**Additional costs for QALYs forgone**
$NMB_{\text{with research}}$	£85,845,233	£77,054,816	£77,028,156
$NMB_{\text{without research}}$	£78,842,093	£77,739,692	£77,589,940
INMBs (£30,000 per QALY)	**£7,003,141**	**-£684,876**	**-£561,784**

As before, an 'optimistic', a 'neutral' and a 'pessimistic' combination were formed by attaching weights to possible outcomes. The weighted results of this analysis appear in Table 6.28, Table 6.29 and Table 6.30 below. Under the 'optimistic' combination, conducting the TRAPEZE phase III trial appears to be beneficial, resulting in INMBs of £3.19 million. The

trial appears less beneficial under the 'neutral' and 'pessimistic' combinations, resulting in INMBs of £1.90 million and £1.30 million, respectively.

Table 6.28: Weighted costs and benefits of research in base case analysis ('optimistic' combination)(two-treatment HRPC)

Optimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted $INMB_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.50	£1,427,022	164	£3,501,570
Inconclusive	0.25	£171,219	0	-£171,219
Unfavourable	0.25	£107,790	-1	-£140,446
Sum of weighted net costs and QALYs	-	£1,706,031	163	-
Weighted ICER and INMBs	£10,454 per additional QALY			£3,189,905

Table 6.29: Weighted costs and benefits of research in base case analysis ('neutral' combination)(two-treatment HRPC)

Neutral combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted $INMB_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.33	£941,835	108	£2,311,036.41
Inconclusive	0.33	£226,009	0	-£226,009.08
Unfavourable	0.33	£142,283	-1	-£185,388.71
Sum of weighted net costs and QALYs	-	£1,310,127	107	-
Weighted ICER and INMBs	£12,245 per additional QALY			£1,899,639

Table 6.30: Weighted costs and benefits of research in base case analysis ('pessimistic' combination)(two-treatment HRPC)

Pessimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted $INMB_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.25	£713,511	82	£1,750,785.16
Inconclusive	0.25	£171,219	0	-£171,219.00
Unfavourable	0.50	£215,580	-2	-£280,891.99
Sum of weighted net costs and QALYs	-	£1,100,310	80	-
Weighted ICER and INMBs	£13,760 per additional QALY			£1,298,674

As noted earlier, the difference between the NSCLC and the HRPC case studies is that in the latter patient-level evidence around the effectiveness of the compared treatments was available from the TRAPEZE phase II trial. This information can be used as an indication of the likelihood of observing different outcomes in subsequent trials. In the phase II trial DP+Sr89 was shown to be more effective than DP. Therefore, it appears reasonable to assume that a further study—the TRAPEZE phase III trial—is likely to show similar results for DP+Sr89. On this premise, the most relevant outcome would be a 'favourable' one with the most relevant combination being the 'optimistic'. Given this, conducting the TRAPEZE phase III trial would be expected to result in NMBs of about £3.19 million.

6.3.3. Additional analyses for the two-treatment HRPC case study

As in the NSCLC case study, additional analyses were carried out to assess the sensitivity of the results to different assumptions. Additional, probabilistic sensitivity analysis was also undertaken by attaching probability distributions to key parameters.

6.3.3.A. Deterministic sensitivity analyses

Deterministic sensitivity analysis was undertaken to look into the impact of alternative assumptions on the results. First, it was hypothesised that research results may have a more pronounced impact on clinical practice than that assumed in the base case analysis. On this premise, a 'favourable' outcome would lead to a more marked uptake of DP+Sr89 (i.e. 50 percent or 60 percent of eligible patients taking DP+Sr89 as opposed to 40 percent in base case analysis), while an 'unfavourable' outcome would result in a sharper decrease in prescription shares (0.5 percent or 0 percent taking DP+Sr89 as compared to 1 percent in the base case analysis). For these values, results revealed increased INMBs for all combinations, suggesting that, if this was the prevailing implementation pattern, the option of conducting research would be more desirable than that associated with the base case pattern (Table 6.31 and Table 6.32).

Table 6.31: Deterministic sensitivity analysis assuming higher uptake (50 percent) following 'favourable' results (two-treatment HRPC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
DP+Sr89 current prescription	5%		
DP+Sr89 prescription following a possible outcome	50%	5%	0.5%
INMBs optimistic combination	£4,292,040		
INMBs neutral combination	£2,629,586		
INMBs pessimistic combination	£1,855,511		

Table 6.32: Deterministic sensitivity analysis assuming higher uptake (60 percent) following 'favourable' results (two-treatment HRPC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
DP+Sr89 current prescription	5%		
DP+Sr89 prescription following a possible outcome	60%	5%	0%
INMBs optimistic combination	£5,394,175		
INMBs neutral combination	£3,359,534		
INMBs pessimistic combination	£2,412,349		

In contrast, it is possible that change in clinical practice may be less responsive to research results. This could be, for instance, due to obstacles in implementation. In such a case it would be expected that, following a 'favourable' outcome, there would be a weaker uptake of DP+Sr89 than in base case (30 percent and 20 percent as opposed to 40 percent in the base case analysis) and following 'unfavourable' outcomes there would be a smaller decrease in the number of patients using DP+Sr89 (2 percent and 3 percent of all patients would be receiving DP+Sr89, as opposed to 1 percent in the base case). The analysis showed this uptake pattern to result in lower INMBs for all combinations, suggesting that carrying out research in this case would still be the preferable option, but, as this pattern is associated with uptake rates further away from the 'optimal implementation' (i.e. all patients receiving the most cost-effective treatment), conducting research would be less beneficial than in the base case (Table 6.33 and Table 6.34, below).

Table 6.33: Deterministic sensitivity analysis assuming lower uptake (30 percent) following 'favourable' results (two-treatment HRPC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
DP+Sr89 current prescription	5%		
DP+Sr89 prescription following a possible outcome	30%	5%	2%
INMBs optimistic combination	£2,083,924		
INMBs neutral combination	£1,164,613		
INMBs pessimistic combination	£734,144		

Table 6.34: Deterministic sensitivity analysis assuming lower uptake (20 percent) following 'favourable' results (two-treatment HRPC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
DP+Sr89 current prescription	5%		
DP+Sr89 prescription following a possible outcome	20%	5%	3%
INMBs optimistic combination	£977,943		
INMBs neutral combination	£429,588		
INMBs pessimistic combination	£169,613		

Last, assuming that the current prescription share of DP+Sr89 is higher than the base case estimate (10 percent as opposed to 5 percent in the base case) while clinical practice change is as in base case resulted in lower INMBs for all combinations (Table 6.35). This is driven by the fact that, here, the proportion of people who would benefit from a switch in clinical practice under the 'favourable' outcome (from the assumed level of 10 percent to 40 percent) is smaller than that in the base case (from 5 percent to 40 percent).

By analogy, assuming that DP+Sr89 is currently not prescribed at all (0 percent as opposed to 5 percent in the base case) and this would remain the case after 'inconclusive' or

'unfavourable' outcomes, led to slightly higher INMBs than in the base case for all combination. This is not surprising given that, under a 'favourable' outcome, more patients (40 percent, from 0 percent to assumed uptake of 40 percent) would switch to a beneficial treatment than in the base case analysis (35 percent, from 5 percent to 40 percent) (Table 6.36).

Table 6.35: Deterministic sensitivity analysis assuming higher current prescription (10 percent) for DP+Sr89 (two-treatment HRPC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
DP+Sr89 current prescription	10%		
DP+Sr89 prescription following a possible outcome	40%	10%	1%
INMBs optimistic combination	£2,679,228		
INMBs neutral combination	£1,587,979		
INMBs pessimistic combination	£1,101,035		

Table 6.36: Deterministic sensitivity analysis assuming no current prescription for DP+Sr89 (two-treatment HRPC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
DP+Sr89 current prescription	0%		
DP+Sr89 prescription following a possible outcome	40%	0%	0%
INMBs optimistic combination	£3,708,276		
INMBs neutral combination	£2,221,453		
INMBs pessimistic combination	£1,511,700		

In the base case analysis, the 'unfavourable' outcome was specified so that DP+Sr89 is not cost-effective compared to DP (i.e. DP+Sr89 is much more costly and slightly more effective

than DP), neither is it dominated by this treatment. An alternative scenario for the 'unfavourable' outcome could be that DP+Sr89 is more costly and less effective than DP (i.e. DP+Sr89 is dominated). This scenario would be observed if the probability of a patient on DP+Sr89 moving from the 'progression-free' state to progressive disease was set at a higher value than in base case (i.e. at 0.185 compared to 0.140 for the 'unfavourable' outcome in the base case). This made the treatment more costly by about £1246 and less effective by 0.004 QALYs compared to DP. For the specific 'unfavourable' outcome, this scenario resulted in additional non-weighted NMBs in the population, due to restricting the use of a more costly and less effective treatment. When these results were combined with likelihood weights in weighted analysis, there were positive INMBs for the comparison between the 'with research' and 'without research' situations, and additional NMBs as compared to the base case analysis. This suggests that gains in NMBs are more pronounced when practice switches away from relatively 'worse' (i.e. dominated) treatments, than from treatments which may be effective but not cost-effective (Table 6.37).

Table 6.37: Deterministic sensitivity analysis for alternative 'unfavourable' outcome (two-treatment HRPC)

	Optimistic combination	Neutral combination	Pessimistic combination
Alternative scenario for 'unfavourable' outcome (DP+Sr89 dominated)	£3,228,068	£1,950,013	£1,374,999

Last, the effect of different time horizons can be seen in Table 6.38. The results confirm that as the number of patients to be affected by a beneficial change in practice increases, so do the expected NMBs from conducting research and changing clinical practice.

Table 6.38: Deterministic sensitivity analysis assuming different time horizons (two-treatment HRPC)

Time horizon	Eligible patients	Optimistic combination	Neutral combination	Pessimistic combination
One year	2668	£1,341,357	£669,907	£352,378
Five years	12,057	£8,473,396	£5,414,437	£4,003,361
Ten years	22,314	£16,265,010	£10,597,743	£7,991,988

6.3.3.B. Gradual change in uptake analysis

While the above analyses presume that research results have an immediate effect on treatments' uptake, it is likely that change in clinical practice occurs gradually. Thus, additional analysis was based on the assumption that uptake may change in a linear fashion over the two-year period. Under this assumption, prescription would be expected to develop as shown in Table 6.39 below.

Table 6.39: Assumed prescription rates in gradual uptake analysis (two-treatment HRPC)

Year	Favourable outcome		Inconclusive outcome		Unfavourable outcome	
	DP+Sr89	DP	DP+Sr89	DP	DP+Sr89	DP
Current	5.0%	95.0%	5.0%	95.0%	5.0%	95.0%
Year 1	22.5%	77.5%	5.0%	95.0%	3.0%	97.0%
Year 2	40.0%	60.0%	5.0%	95.0%	1.0%	99.0%

Based on these uptake patterns, non-weighted results were re-calculated and are given in Table 6.40.

Table 6.40: Non-weighted costs and benefits of research in gradual uptake analysis (two-treatment HRPC)

	Favourable outcome	Inconclusive outcome	Unfavourable outcome
With research			
Cost	£48,246,889	£46,651,582	£46,464,766
Trial cost	£684,876	£684,876	£684,876
QALYs	4403	4124	4115
Without research			
Cost	£45,960,005	£45,966,706	£45,967,268
QALYs	4160	4124	4119
Net implications			
Net cost	£2,286,884	£684,876	£497,498
Net QALYs	243	0	-3
Cost per QALY	**£9424 per additional QALY**	**Costs for no additional QALYs**	**Additional costs for QALYs forgone**
$NMB_{\text{with research}}$	£83,835,093	£77,054,816	£76,995,972
$NMB_{\text{without research}}$	£78,842,093	£77,739,692	£77,589,940
INMBs (£30,000 per QALY)	**£4,993,000**	**-£684,876**	**-£593,968**

In general, non-weighted results predicted lower INMBs than those found in the base case analysis. This is unsurprising given that, in this case, the move to a more cost-effective treatment is introduced gradually. Accordingly, weighted results, given in Table 6.41, Table 6.42 and Table 6.43 below, are also lower than in the base case analysis, at £2.18 million, £1.23 million and £780,000 NMBs for an 'optimistic', 'neutral' and 'pessimistic' combination (as opposed to £3.19 million, £1.9 million and £1.3 million for the respective combinations in base case). As in the gradual uptake analysis for NSCLC, it must be noted that the results dependent largely on the hypothesised prescription rates and the assumed pattern of change in uptake.

Table 6.41: Weighted costs and benefits of research in gradual uptake analysis ('optimistic' combination)(two-treatment HRPC)

Optimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.50	£1,143,442	121	£2,496,500
Inconclusive	0.25	£171,219	0	-£171,219
Unfavourable	0.25	£124,375	-1	-£148,492
Sum of weighted net costs and QALYs	-	£1,439,036	121	-
Weighted ICER and INMBs	£11,939 per additional QALY			£2,176,789

Table 6.42: Weighted costs and benefits of research in gradual uptake analysis ('neutral' combination)(two-treatment HRPC)

Neutral combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.33	£754,672	80	£1,647,690
Inconclusive	0.33	£226,009	0	-£226,009
Unfavourable	0.33	£164,174	-1	-£196,009
Sum of weighted net costs and QALYs		£1,144,855	79	-
Weighted ICER and INMBs	£14,489 per additional QALY			£1,225,671

Table 6.43: Weighted costs and benefits of research in gradual uptake analysis ('pessimistic' combination)(two-treatment HRPC)

Pessimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.25	£571,721	61	£1,248,250
Inconclusive	0.25	£171,219	0	-£171,219
Unfavourable	0.50	£248,749	-2	-£296,984
Sum of weighted net costs and QALYs	-	£991,689	59	-
Weighted ICER and INMBs	£16,792 per additional QALY			£780,047

6.3.3.C. Probabilistic sensitivity analysis

The impact of change in clinical practice on weighted and non-weighted 'payback' results was assessed in probabilistic sensitivity analyses. In doing so, 1000 estimates of the possible levels of uptake were drawn from distributions attached to the 'favourable' and 'unfavourable' outcomes. As in the equivalent NSCLC analysis, the distributions had means equal to the base case uptake rates and standard errors hypothesised to be one-quarter of the mean values. As DP+Sr89 prescription in the light of an 'inconclusive' outcome is assumed to remain unchanged, this parameter was held at its current prescription value (5 percent). The mean values and confidence intervals for costs, QALYs and INMBs resulting from the probabilistic analysis are given in Table 6.44.

The results of this analysis are almost identical to those in the base case analysis, suggesting that further research would be beneficial only if it showed a 'favourable' outcome for DP+Sr89. In the case of 'inconclusive' or 'unfavourable' outcomes, conducting research would result in negative INMBs of £684,900 and £562,100, respectively. It must be stressed that, similarly to the NSCLC study, the results are highly dependent on assumptions about the distributions representing the uptake levels.

Table 6.44: Non-weighted costs and benefits of research in probabilistic analysis (two-treatment HRPC)

	Favourable outcome (Lower and upper 95% confidence intervals)	**Inconclusive outcome** (Lower and upper 95% confidence intervals)	**Unfavourable outcome** (Lower and upper 95% confidence intervals)
With research			
Cost	**£48,122,483** (£48,084,132 to £48,160,834)	**£45,966,706**	**£45,714,133** (£45,713,155 to £45,715,110)
Trial cost	**£684,876**	**£684,876**	**£684,876**
QALYs	**4488** (4482 to 4493)	**4124**	**4114** (4113 to 4414)
Without research			
Cost	**£45,960,005**	**£45,966,706**	**£45,967,268**
QALYs	**4160**	**4124**	**4119**
Net implications			
Net cost	**£2,847,355** (£2,809,004 to £2,885,705)	**£684,876**	**£431,741** (£430,764 to £432,718)
Net QALYs	**328** (322 to 333)	**0**	**-4** (-4 to -4)
Cost per QALY	**£8516 per additional QALY**	**Costs for no additional QALYs**	**Additional costs for QALYs forgone**
$NMB_{with\ research}$	**£85,821,525** (£85,685,602 to £85,957,448)	£77,054,816	**£77,027,874** (£77,027,400 to £77,028,348)
$NMB_{without\ research}$	**£78,842,093**	**£77,739,692**	**£77,589,940**
INMB (£30,000 per QALY)	**£6,979,432** (£6,843,509 to £7,115,355)	**-£684,876**	**-£562,065** (-£562,539 to -£561,591)

Additional analysis was carried out to assess the uncertainty around likelihood weights used in forming combinations. As in the NSCLC application, likelihood weights were drawn from a Dirichlet distribution using the normalized sum of independent gamma draws method[51;261]. The drawn likelihood weights were combined with the probabilistic results obtained from varying the prescription rates to give 1000 estimates of the weighted net costs and QALYs

for each combination. Results, in the form of average values across the 1000 cost and QALY estimates are given in Table 6.45, Table 6.46 and Table 6.47. These showed values from £1.31 million to £3.13 million for the 'pessimistic' and 'neutral' combinations, respectively, supporting the conclusion that the two-arm trial would be a 'cost-effective' investment.

Table 6.45: Weighted costs and benefits of research from probabilistic analysis ('optimistic' combination)(two-treatment HRPC)

Optimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with research vs. without research)}$**
Favourable	0.50	£1,406,860	162	£3,445,261
Inconclusive	0.25	£172,382	0	-£172,382
Unfavourable	0.25	£109,283	-1	-£142,288
Sum of weighted net costs and QALYs	-	£1,688,526	161	-
Weighted ICER and INMBs	£10,511 per additional QALY			£3,130,591

Table 6.46: Weighted costs and benefits of research from probabilistic analysis ('neutral' combination)(two-treatment HRPC)

Neutral combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted INMB$_{(with research vs. without research)}$**
Favourable	0.33	£949,019	109	£2,324,863
Inconclusive	0.33	£225,899	0	-£225,899
Unfavourable	0.33	£145,184	-1	-£189,060
Sum of weighted net costs and QALYs	-	£1,320,102	108	-
Weighted ICER and INMBs	£12,261 per additional QALY			£1,909,903

Table 6.47: Weighted costs and benefits of research from probabilistic analysis ('pessimistic' combination)(two-treatment HRPC)

Pessimistic combination				
Outcome	**Likelihood weight**	**Weighted net cost**	**Weighted net QALYs**	**Weighted $INMB_{(with\ research\ vs.\ without\ research)}$**
Favourable	0.25	£718,251	83	£1,761,200
Inconclusive	0.25	£172,569	0	-£172,569
Unfavourable	0.50	£214,134	-2	-£278,772
Sum of weighted net costs and QALYs		£1,104,953	80	-
Weighted ICER and INMBs	£13,727 per additional QALY			£1,309,859

6.3.4. Base case analysis of the four-treatment TRAPEZE trial

While the steps in carrying out PATHS analysis for two-treatment comparisons are well documented[67], applications involving more than two treatments are scant and their methodology is relatively unclear. Particular ambiguity exists around the most appropriate way of specifying possible outcomes. Given this, additional analyses were carried out to illustrate a possible way of applying PATHS to multi-treatment assessments.

6.3.4.A. Specification of outcomes

Only one application of PATHS on a case study involving multiple treatments was found in the literature[116]. In this study, Fleurence[116] used PATHS to assess the value of funding and conducting a particular study, the Record Trial, which compared three active treatments (vitamin D and calcium (VDC), vitamin D (VD), and calcium (C)) against placebo for the

prevention of fractures in patients with osteoporosis. In this application, the active treatments were grouped together, so that the study was, in essence, reduced to a two-treatment comparison. For instance, under the 'positive' outcome, the author assumed that the trial would show all of the active treatments (VDC, VD and C) to be effective and cost-effective, with different degrees of effectiveness attached to them. While in a trial more than one treatment may be shown effective against placebo, in an economic evaluation, where the cost-effectiveness of a treatment is determined in relation to the rest of the compared treatments, only one of VDC, VD and C can be cost-effective. Therefore, additional sub-scenarios need to be specified in order to account for the fact that each of the compared treatments may be proven cost-effective.

6.3.4.B. Specification of possible outcomes

With this in mind, possible outcomes were specified as follows:

a. The TRAPEZE phase III trial gives results such that, when these results are entered into the HRPC model, the most cost-effective treatment is shown to be DP+Sr89 (i.e. DP+Sr89 shows the highest NMBs amongst the compared treatments);

b. The TRAPEZE phase III trial gives results such that, when these results are entered into the HRPC model, all treatments are shown to be of similar cost-effectiveness ($NMB_{DP} = NMB_{DP+ZA} = NMB_{DP+Sr89} = NMB_{DP+ZA+Sr89}$),

c. The TRAPEZE phase III trial gives results such that, when these results are entered into the HRPC model, the most cost-effective treatment is shown to be DP (i.e. DP shows the highest NMBs amongst the compared treatments),

d. The TRAPEZE phase III trial gives results such that, when these results are entered into the HRPC model, the most cost-effective treatment is shown to be DP+ZA (i.e. DP+ZA shows the highest NMBs amongst the compared treatments), and

e. The TRAPEZE phase III trial gives results such that, when these results are entered into the HRPC model, the most cost-effective treatment is shown to be DP+ZA+Sr89 (i.e. DP+ZA+Sr89 shows the highest NMBs amongst the compared treatments).

The cost of the four-treatment TRAPEZE trial, as stated in the trial proposal submitted for funding, was £2.54 million.

The first task involved working out the trial results that would be expected from the proposed trial, if the specified outcomes (i.e. 'favourable', 'inconclusive' and 'unfavourable') were to transpire. These results—expressed in terms of probabilities of transition from a progression-free state to progressive disease—are given in the last column of Table 6.48, alongside the existing probabilities used in the HRPC model (second column in the table).

Table 6.48: Transition probabilities used in four-treatment HRPC case study

Outcome	Base case transition probability from 'Progression-free, not on treatment' to 'Progression'	Adjusted transition probability to match each possible outcome
DP+Sr89 cost-effective	0.05	0.04
DP cost-effective	0.06	0.01
DP+ZA cost-effective	0.09	0.01
DP+ZA+Sr89 cost-effective	0.04	0.01

On the basis of the specified transition probabilities, the cost and QALY estimates would assume the values in Table 6.49. In scenarios where a particular treatment was not cost-effective, it was assumed that its cost and benefits remained at their 'current' values, as calculated in the HRPC decision model (non-bold values).

For example, under the specified outcome of DP being the cost-effective treatment, the cost and QALYs associated with this treatment was £8901 and 0.92, respectively, so that DP is the most cost-effective of all treatments (i.e. it results in the highest NMBs), while the values for the rest of the treatments remained as in base case. In the case of an 'inconclusive' scenario, transition probabilities were adjusted so that all treatments are of comparable cost-effectiveness (i.e. all treatments showed very similar NMBs).

Table 6.49: Cost-effectiveness results based on 'favourable', 'inconclusive' and 'unfavourable' outcomes (four-treatment HRPC)

Outcome	DP		DP+ZA		DP+Sr89		DP+ZA+Sr89	
	Total cost	Total QALYs	Total cost	Total QALYs	Total cost	Total QALYs	Total cost	Total QALYs
Base case	£8949	0.81	£9855	0.83	£10,172	0.96	£11,436	0.87
DP+Sr89 cost-effective	£8949	0.81	£9855	0.83	**£10,164**	**0.99**	£11,436	0.87
Inconclusive	**£8928**	**0.87**	**£9842**	**0.90**	**£10,181**	**0.91**	**£11,405**	**0.96**
DP cost-effective	**£8901**	**0.92**	£9855	0.83	£10,172	0.96	£11,436	0.87
DP+ZA cost-effective	£8949	0.81	**£9822**	**0.95**	£10,172	0.96	£11,436	0.87
DP+ZA+Sr89 cost-effective	£8949	0.81	£9855	0.83	£10,172	0.96	**£11,381**	**1.00**

6.3.4.C. Change in clinical practice

As before, estimates of possible prescription rates were obtained through discussion with experts from the NHS Pan Birmingham Cancer Network. Such estimates were based on deliberations about the direction and strength of change in practice in the light of different eventualities and are given in Table 6.50. In general, should a treatment appear cost-effective, its uptake would be expected to increase, although the magnitude of increase would depend on possible obstacles to implementation. For example, the available infrastructure for strontium-89 administration is limited and thus, if treatments involving radioisotope fractions (i.e. DP+Sr89 or DP+ZA+Sr89) were found to be cost-effective, their uptake would be expected to be lower than that of treatments not requiring strontium-89 administration (i.e. DP, DP+ZA). This is reflected in the assumed prescription shares, where finding in favour of DP+Sr89 or DP+ZA+Sr89 is presumed to lead to 40 percent of the eligible patients taking those treatments, as opposed to 50 percent for DP+ZA.

In the case that DP (the usual practice) appeared cost-effective, the uptake of this treatment would be expected to increase (from 85 percent to 90 percent), while the uptake of the rest of the treatments would decrease (from 5 percent to 3.3 percent). This reflects the presumption that, if the currently standard treatment (DP) is confirmed to be cost-effective, there will be a pressure for restrictions in the use of non-standard, experimental treatments (DP+ZA, DP+Sr89, DP+ZA+Sr89). On the other hand, if any of the non-standard treatments appears cost-effective (either of DP+ZA, DP+Sr89, DP+ZA+Sr89), the pressure for restrictions would be expected to ease off and the non-cost-effective experimental treatments would be likely to remain at current implementation levels (5 percent).

Table 6.50: Treatment uptake following the specified research outcomes (four-treatment HRPC)

Prescription	Possible outcome	Percentage use of DP	Percentage use of DP+ZA	Percentage of DP+Sr89	Percentage of DP+ZA+Sr89
Current prescription		85.0%	5.0%	5.0%	5.0%
Prescription in the light of each possible outcome	**DP cost-effective**	90.0%	3.3%	3.3%	3.3%
	DP+Sr89 cost-effective	50.0%	5.0%	40.0%	5.0%
	DP+ZA cost-effective	40.0%	50.0%	5.0%	5.0%
	DP+ZA+Sr89 cost-effective	50.0%	5.0%	5.0%	40.0%

6.3.5. Results of the four-treatment HRPC case study

The non-weighted results of the base case analysis are given in Table 6.51. Under a 'favourable' scenario for DP+Sr89, there would be costs due to conducting the trial and moving towards a more costly interventions but also gains in QALYs, resulting in overall gains of £5.13 million in NMBs.

Table 6.51: Non-weighted costs and benefits of research in base case analysis (four-treatment HRPC)

Outcome	DP+Sr89 cost-effective	Inconclusive	DP cost-effective	DP+ZA cost-effective	DP+ZA+Sr89 cost-effective
With trial					
Cost	£48,994,528	£46,727,957	£46,213,291	£48,822,245	£51,156,266
Trial cost	£2,537,116	£2,537,116	£2,537,116	£2,537,116	£2,537,116
QALYs	4511	4487	4655	4525	4549
Without trial					
Cost	£46,825,405	£46,727,957	£48,838,426	£46,818,994	£46,813,533
QALYs	4183	4487	4648	4204	4206
Net implications					
Net costs	£4,706,240	£2,537,116	-£88,019	£4,540,368	£6,879,849
Net QALYs	328	0	7	320	342
Cost per QALY	**£14,351 per additional QALY**	**Costs for no additional QALYs**	**Cost savings for additional QALYs**	**£14,175 per additional QALY**	**£20,101 per additional QALY**
$NMB_{with\ research}$	£83,798,160	£85,358,679	£90,893,840	£84,383,144	£82,763,340
$NMB_{without\ research}$	£78,666,300	£87,895,795	£90,586,669	£79,313,999	£79,375,245
INMBs (£30,000 per QALY)	**£5,131,860**	**-£2,537,116**	**£307,171**	**£5,069,145**	**£3,388,095**

Under the 'inconclusive' scenario, there would be a loss of £2.54 million due to the expenditure for the 'inconclusive' trial. If DP was found cost-effective, there would be cost savings from restricting the use of more expensive non-standard treatments and additional QALYs, resulting in overall positive NMBs of £307,200. In the case of DP+ZA or DP+ZA+Sr89 being cost-effective, there would be gains in QALYs, which would offset the increased costs to result in positive NMBs of £5.07 million and £3.39 million, respectively.

Each outcome was attached a likelihood of being observed, to form combinations. In each of a series of alternative combinations, a weight of 0.5 was given to observing 'favourable' results for a specific treatment, with the likelihood weight for the rest of the results being 0.125. For example, in alternative combination B where DP is specified as the most cost-effective treatment the weight for DP being associated with a 'favourable' outcome was 0.5, while a weight of 0.125 was given for the rest of the outcomes.

According to the results, funding and conducting the more expensive four-arm TRAPEZE phase III trial would be beneficial, and it is expected to lead to NMBs between £468,500 and £3.34 million (Table 6.52).

Table 6.52: Weighted costs and benefits of research (four-treatment HRPC)

Combination	Weighted INMB(with research vs. without research)
Alternative combination A (i.e. greater weight on 'favourable' results for DP+Sr89)	£3,344,342
Neutral combination	£468,476
Alternative combination B (i.e. greater weight on 'favourable' results for DP)	£1,535,084
Alternative combination C (i.e. greater weight on 'favourable' results for DP+ZA)	£3,320,824
Alternative combination D (i.e. greater weight on 'favourable' results for DP+ZA+Sr89)	£2,690,430

On the premise that the phase III trial will show results similar to the phase II study, the most likely combination would be Combination A, which is at the top end of the estimated range of 'payback' results.

6.3.6. Additional analyses for the four-treatment HRPC case study

As noted earlier, the four-treatment HRPC analysis was conducted to illustrate a possible way to using PATHS in multi-treatment comparisons, rather than to obtain results and draw conclusions on possible patterns. However, additional deterministic sensitivity analyses were carried out to assess the sensitivity of the obtained results to different assumptions. These are given in Table 6.53.

Table 6.53: Deterministic sensitivity analysis results (four-treatment HRPC)

Combination	Base case results (Weighted $INMB_{\text{with research vs.without research}}$)	Sensitivity analysis: greater change in clinical practice (Weighted $INMB_{\text{with research vs.without research}}$)	Sensitivity analysis: less marked change in clinical practice (Weighted $INMB_{\text{with research vs.without research}}$)	Sensitivity analysis: perfect implementation (Weighted $INMB_{\text{with research vs.without research}}$)
Alternative combination A (i.e. greater weight on 'favourable' results for DP+Sr89)	£3,344,342	£6,310,375	£1,536,599	£12,297,626
Neutral combination	£468,476	£1,791,157	-£517,591	£4,426,801
Alternative combination B (i.e. greater weight on 'favourable' results for DP)	£1,535,084	£3,278,912	£315,045	£5,961,351
Alternative combination C (i.e. greater weight on 'favourable' results for DP+ZA)	£3,320,824	£5,277,360	£1,067,046	£10,176,529
Alternative combination D (i.e. greater weight on 'favourable' results for DP+ZA+Sr89)	£2,690,430	£5,282,799	£1,069,519	£10,163,452

As before, a first analysis looked into the possibility that research results would have a greater impact on clinical practice than that hypothesised in the base case analysis (i.e. a

'favourable' outcome for a treatment would result in greater uptake of this treatment in practice). As expected, a more pronounced move towards the cost-effective treatment would be associated with greater expected NMBs than those predicted in the base case analysis. Alternatively, on the assumption that change in clinical practice would be less marked than that assumed in the base case, the expected NMBs from research are expected to be lower, and, in the case of a 'neutral' combination, they become negative suggesting that, in this situation, the benefits from the trial would not exceed its cost. Last, if change in clinical practice was 'perfect', namely, all patients were given the treatment shown to be the most cost-effective, the gains from conducting research and informing clinical practice would be substantial, from £4.43 million to £12.30 million.

6.4. Discussion

This chapter reports the application of the PATHS model to the NSCLC and HRPC case studies. The application aimed to give results on the potential value of carrying out the BTOG-2 and TRAPEZE phase III trials, as well as to give an insight into the employed methods, their strengths and their limitations. Overall, the analysis suggested that funding and conducting the proposed trials would be beneficial. In the base case analysis, carrying out the BTOG-2 trial and changing clinical practice according to its findings was estimated to result in additional NMBs of between £1.88 million and £2.38 million. Similarly, in the case of HRPC, the base case analysis predicted gains from conducting the TRAPEZE phase III trial between £1.30 million and £3.19 million for the two-treatment comparison, and between £468,500 and £3.34 million for the four-treatment comparisons. On the premise that the

TRAPEZE phase III trial may show results similar to those observed in the phase II study (i.e. 'favourable' to DP+Sr89), conducting the proposed phase III trial would be expected to result in NMBs of about £3.19 million and 3.34 million for the two-treatment and four-treatment comparisons, respectively.

The application followed the core methods of the PATHS model, as laid out by Townsend and colleagues[67], with two notable diversions. First, the present analysis accounts for the fact that evidence from clinical trials typically comes in the form of measures of clinical effectiveness (e.g. disease progression rates associated with a treatment), and, for this evidence to be useful in decision-making, it need to be translated into estimates of the treatments' per-patient costs and benefits. In the present study, this was done by using the NSCLC and HRPC decision analytic models. The second modification relates to specifying how different results are expected to affect clinical practice, which is a key requirement in 'payback' models[67;90;98]. In previous applications, this was determined by asking local experts and decision-makers to indicate the level of effectiveness and cost-effectiveness above which they would expect an assessed technology to be adopted in clinical practice[67]. Although local decision-makers are likely to give 'richer' answers that, in addition to cost-effectiveness, may reflect other relevant considerations (e.g. a local authority's budget, need for other services, local population's epidemiological profile), such answers appear to be restricted to the specific locality and may not be appropriate for assessing the costs and benefits of research the results of which would be disseminated nationally (or even internationally) and inform decision-making at central level. In the present application,

evidence on a treatments' cost-effectiveness was linked to changes in clinical practice—that is, adoption or restriction of treatments—on the basis of the rationale that underpins adoption decisions made by NICE: a health care intervention is likely to be recommended if it offers and additional unit of benefit at a cost lower than a willingness-to-pay threshold[114;246]. Here, this threshold was set at £30,000 per QALY.

Additional analyses were carried out to look into factors that may have a bearing on the results. Deterministic analyses showed that the magnitude of the obtained results depends largely on the treatments' effectiveness and cost-effectiveness as specified in possible outcomes, as well as on how clinical practice (prescription shares) is expected to develop in view of these outcomes. Scenarios where changes in practice occur gradually, over time, were also explored; these suggested that the 'sharper' the move towards adopting a cost-effective treatment (or restricting the use of a non-cost-effective one), the greater the expected benefits to the population.

A further, probabilistic component was added to the analyses, where uncertain parameters, such as treatments' prescription shares and likelihoods of observing particular outcomes, were given probability distributions, rather than single values. These analyses gave results together with ranges within which the results are expected to fall, although, as noted earlier, results in this case are highly dependent on the characteristics of the distributions assigned to the parameters. A way of addressing multiple-treatment comparisons which involved specifying several possible outcomes was also outlined and applied to the four-treatment HRPC case study.

In order for the model to work, certain assumptions were required. First, in line with existing 'payback' models[67;90;92;100], possible outcomes need to be specified under the assumption that a discrete number of well-defined results may appear, although, in practice, results can assume values anywhere between the specified effectiveness estimates. Further, the application assumed that a trial would give evidence on a single parameter of one treatment (here, rates of disease progression for Gem+Carb and DP+Sr89) though, in practice, a trial would look into more parameters (e.g. adverse events and quality of life) and different treatments. Accounting for different parameters and treatments would make the specification of possible outcomes particularly complex, as there would be an infinite number of combinations of different parameter values that would make a treatment appear cost-effective, inconclusive or non-cost-effective.

The analysis made extensive use of assumptions. Information on future uptake of treatments and time horizons does not usually exist in empirical form and thus in line with existing studies[67;92] it was obtained through expert opinion. Similarly, the employed likelihood weights for combinations were based on previous applications of PATHS in the literature[67;116]. In this illustrative analysis, assumptions were also needed for the gradual implementation analysis (i.e. rate of treatments' uptake increase or decrease over the specified time horizons) and the probabilistic sensitivity analysis. In the latter, no values on the variance of the distributions for uptake rates and likelihood weights were available and, thus, in order to illustrate the methods, estimates were chosen arbitrarily.

Last, the views of experts on possible changes in clinical practice were obtained through discussion in the context of semi-structured interviews. The use of formal elicitation techniques might have added rigour to this exercise, however, such an undertaking was considered to be beyond the pragmatic constraints of this doctoral project.

6.5. Chapter overview

The chapter presents and discusses the application of the latest 'payback' model (PATHS) to the NSCLC and HRPC case studies. For each case study, the application involved a base case analysis, as well as additional deterministic and probabilistic sensitivity analyses. An approach to carrying out PATHS analysis for four-treatment comparisons was also illustrated and applied to the HRPC study.

In summary, results suggested that, given specific assumptions about possible research outcomes, change in clinical practice and likelihoods of the outcomes to transpire, undertaking the BTOG-2 and TRAPEZE trials would result in additional benefits. In addition to giving results, the analysis provided useful information around the methodological and practical challenges associated with the framework. These are discussed in more detail in Chapters 8 and 9.

CHAPTER 7. Practical application of 'value of information' to case studies

This chapter focuses on the application of 'value of information' (VoI) to the case studies of NSCLC and HRPC. As in the 'payback' analysis, the application aims to give an insight into the frameworks' strengths and challenges, rather than to inform actual funding decisions. The first part of the chapter gives a description of the methods involved in VoI analysis, while the second part reports and interprets the obtained results. The chapter concludes with a discussion of points which arose in the course of this analysis.

7.1. Description of methods in 'value of information'

A description of the rationale underpinning the VoI framework has been given earlier in this thesis, in Chapter 3. In essence, the framework seeks to establish the gains from—and thus the value of—obtaining further information through evaluative research by looking into the additional benefits that would be expected to accrue from decision-making in the light of this information[97;179;180]. Given this, the framework has been often advocated as a means of establishing the value of conducting further evaluative research[55;80;121].

Different concepts exist in 'value of information' analysis, such as the expected value of perfect information for all or a subset of uncertain parameters, expected value of sample information, as well as concepts related to the value of implementation. These concepts are described below with a focus on the methods involved in calculating relevant results.

7.1.1. Expected value of perfect information

A first, core step in 'value of information' analysis relates to establishing the expected value of perfect information (EVPI). EVPI represents the difference between the net monetary benefits (NMBs) to be gained if a treatment adoption decisions (e.g. the decision on which NSCLC treatment to adopt) was made under perfect information (effectively, with no uncertainty) and the NMBs expected from the same decision made in the light of current, imperfect information. EVPI can be estimated parametrically, using analytic formulae[97;262]; however, such calculations require assumptions about normally distributed NMB results to hold true, which is in practice rare, especially when cost-effectiveness results and estimates of uncertainty are obtained from non-linear decision analytic models[51;190]. In such cases, EVPI can be calculated non-parametrically, using estimates of uncertainty in the form of simulated results obtained from probabilistic sensitivity analysis. Adopting the notation in Briggs *et al.*[51], for *j* alternative interventions and a set of all uncertain parameters θ, EVPI is given by the formula:

$$EVPI = E_\theta \, max_j \, NMB(j, \theta) - max_j \; E_\theta NMB(j, \theta)$$

and can be estimated following the steps given in Box 7.1[51;263]. EVPI calculated in this way represents the expected additional benefits due to making a decision about treating a single patient under perfect information, as opposed to making the same decision under current information. Once information is generated, it can be used to inform all decisions for all

current and future eligible patients over a specific time horizon representing the period over which the information is expected to be useful[183].

1. Carry out probabilistic sensitivity analysis to obtain a large number of cost and effect (e.g. QALYs) pairs for each treatment j under assessment. Each of these pairs represents the results for a specific resolution of the uncertain parameters. Convert cost and effects into NMBs.

2. For each treatment j, calculate the average across all the simulated NMBs estimates, to obtain the expected NMBs associated with each treatment. Compare the expected NMBs across treatments to establish the treatment associated with the highest expected NMBs ($max_j \ E_\theta NMB(j,\theta)$). This represents the optimal treatment under current information.

3. If perfect information about the correct values of the uncertain parameters (i.e. the correct resolution of parameters) affecting the decision was available, one would always choose the treatment resulting in the maximum NMBs ($max_j NMB(j,\theta)$). As it is not known which resolution of parameters represents the true NMBs, calculate the average over all the maximum expected NMBs to get the expectation of the maximum NMBs ($E_\theta \ max_j \ NMB(j,\theta)$). This represents the expected benefits from a decision with perfect information.

4. Find the difference between the expectation of the maximum NMBs ($E_\theta \ max_j \ NMB(j,\theta)$) and the maximum of the expected NMBs ($max_j \ E_\theta NMB(j,\theta)$), in other words the difference between the expected gains from a decision with perfect and current information. This value represents the EVPI associated with the adoption decision (i.e. choice between treatments).

Box 7.1: Steps in non-parametric calculation of EVPI

The EVPI for the population of eligible patients is the sum of the individual EVPI multiplied by the discounted number of patients affected by the decision in each period over the time horizon for which the generated information is expected to be useful (I_t), starting from the point when information is disseminated. The formula for calculating the population EVPI is:

$$EVPI_{population} = EVPI_{individual} \times \sum_{t=D+1,D+2,..,D+T} \frac{I_t}{(1+r)^t}$$

Here, D is the time lag between a funding decision and dissemination of results, measured in relevant time periods, and r is the discount rate. As EVPI shows the maximum NMBs from pursuing perfect information around a treatment adoption decision through conducting research, EVPI results represent the maximum amount of resources that a rational decision-maker should devote to research around this adoption decision. If EVPI exceeds the cost of research, conducting research is potentially—although not necessarily—worthwhile[110;123].

7.1.2. Expected value of perfect parameter information

EVPI analysis can be extended to establish the value of undertaking research to eliminate uncertainty around one or a subset of all the uncertain parameters affecting an adoption decision. Formally, the expected value of perfect parameter information (EVPPI) for a group of uncertain parameters φ of all uncertain parameters θ associated with a decision between j alternative interventions is given by the formula below:

$$EVPPI_{\varphi} = E_{\varphi}\, max_j\, E_{\psi|\varphi} NMB(j, \varphi, \psi) - max_j E_{\theta} NMB(j, \theta)$$

This represents the difference between the NMBs expected to accrue from a decision made with perfect information about the parameters of interest φ (and imperfect information about the remaining parameters ψ) and a decision made with current, imperfect information about all parameters θ. Individual EVPPI can be extrapolated to the population of patient that stand to benefit from further information in a way identical to that of EVPI.

Different ways of calculating EVPPI have been proposed, discussed and evaluated in the literature[55;180;264]. The method which is commonly suggested as the most appropriate makes use of nested two-level Monte Carlo simulation loops, in recognition of the fact that calculation of the expected NMBs under perfect information involve two expectations[190;263;264]. The process of calculating the EVPPI for one (or a group of) uncertain parameter(s) φ through the nested two-level Monte Carlo approach is given in

Box 7.2.

As a result of using two loops, EVPPI calculations typically require a large number of simulations[265], although under certain assumptions (i.e. linear relationship between input parameters and calculated NMBs, and no correlation between parameter of interest and remaining parameters) shortcuts such as single-level Monte Carlo simulations may also be appropriate[190;264]. The minimum number of inner and outer loop runs required to obtain accurate EVPPI has been subject to debate, although it is generally agreed that this number depends on factors related to the structure of the decision model, the number of uncertain

parameters of interest and the impact of these parameters on the final cost-effectiveness results[263;264]. Brennan and colleagues[263] carried out an empirical investigation on the impact of the number of inner and outer loop runs on the EVPPI result and concluded that a combination of 100 outer and 500 inner loops would, in most situations, lead to convergence and sufficiently accurate EVPPI results. In general, higher numbers of inner and outer loop simulations result in lower levels of sampling error and improved accuracy[263;264;266].

EVPPI shows the difference between the NMBs expected to arise from a decision made under perfect information around a group of parameters, over the NMBs expected under current information around all parameters affecting the decision. Given this, EVPPI analysis has been suggested as a means of indicating those parameters for which pursuing further information would be beneficial[184;185;263;267]. Similarly to EVPI, if the expected benefits of eliminating uncertainty around a group or a single parameter through carrying out a research programme (e.g. a clinical trial) exceed the cost of the programme, research to obtain better information on the parameter(s) would be potentially cost-effective[51;185].

1. Draw a set of values (i.e. one value for each parameter) from the probability distributions of parameters (outer loop).

2. Holding the drawn set of values for parameters fixed, carry out a large number (e.g. *m*= 1000) of Monte Carlo simulations, where values are drawn randomly for the remaining uncertain parameters ψ (inner loop). For each simulation and each treatment, calculate costs and effects, and convert these into NMBs. Record the 1000 NMBs for each treatment.

3. For each treatment, calculate the average across the obtained 1000 NMB estimates. This gives a set of expected NMB values (one for each treatment), given a fixed (considered certain) set of values of and non-fixed (uncertain) values of the remaining parameters ψ ($E_{\psi|\varphi}NMB(j,\varphi,\psi)$). Repeat steps 1 to 3 a large number of times (e.g. k =1000) to obtain 1000 sets of expected NMBs (1000 NMBs for each treatment)

4. Calculate the average across all the expected NMB estimates generated in step 3. Choose the treatment with the highest average NMBs ($max_j E_\theta NMB(j,\theta)$). This is the optimal treatment under current information (continued in the following page).

5. For each set of expected NMB estimates obtained in step 3, calculate and record the maximum expected NMBs. Taking the average of the maximum expected NMBs will give the expected value of the maximum expected NMBs ($E_\varphi\ max_j\ E_{\psi|\varphi}NMB(j,\varphi,\psi)$).This value represents the average gains from making a decision with perfect information about the parameters of interest .

6. Last, subtract the average gains with current information ($max_j E_\theta NMB(j,\theta)$) from the gains expected from a decision with perfect information about the parameters of interest ($E_\varphi\ max_j\ E_{\psi|\varphi}NMB(j,\varphi,\psi)$). This gives the EVPPI for the specific group of parameters .

Box 7.2: Steps in non-parametric calculation of EVPPI

7.1.3. Expected value of perfect information and implementation

It has often been argued that benefits expected from treatment adoption decisions are rarely realised in full, as, in many cases, the uptake (or implementation) of cost-effective treatments does not reach an optimal level[186;268]. Implementation programmes can be undertaken to improve the uptake of beneficial treatments, but such programmes come at a cost[188]. In view of this, Fenwick and colleagues[187] have introduced a conceptual framework which considers the value of devoting resources to implementation strategies.

Value of implementation analysis focuses on the benefits expected to accrue in different 'states of the world', where each state represents a unique combination of different levels of information available for making a decision and implementation of the decision in clinical practice. Implementation is expressed as the prescription share, or uptake, of a treatment in clinical practice; this can be 'current', where only a proportion of the eligible patients receive the treatment that appears cost-effective, or 'optimal', where all eligible patients receive the cost-effective treatment. Different 'states of the world', as summarised by Fenwick and colleagues[187] are given in Table 7.1.

Table 7.1: Possible eventualities regarding information and implementation

Possible 'states of the world'		**Information**	
		Current	**Perfect**
Implementation	**Current**	state A	state B
	Optimal	state C	state D

The NMBs associated with each of the above states can be estimated from the results of the probabilistic model, using the following formulae[187]:

- $State\ A = \ Pop \times \sum_{j=1}(P_j^i \times E_\theta NMB(j,\theta))$
- $State\ B = Pop \times E_\theta \sum_{j=1}(P_j^i \times NMB(j,\theta))$
- $State\ C = Pop \times max_j E_\theta NMB(j,\theta)$
- $State\ D = Pop \times E_\theta max_j NMB(j,\theta)$

Here, Pop represents the eligible population of patients, P_j^i represents the proportion of patients taking treatment j given implementation i (i=current or perfect) and θ represents the uncertain parameters affecting a decisions. Comparisons between the NMBs expected to accrue in different states are thought to give useful measures of the value of pursuing better implementation[124]. An important assumption underpinning such comparisons is that acquiring additional information will not, in itself, have an impact on implementation, which will always remain at current level unless active implementation strategies are put in place to alter it. Fenwick *et al.*[187] acknowledged that this is a somewhat simplistic view and it is more likely that acquisition of information does affect subsequent implementation.

7.1.4. Expected value of sample information and expected net benefit of sampling

As explained above, EVPI and EVPPI results show the NMBs expected to accrue from eliminating uncertainty around all or a subset of parameters affecting a decision. However, eliminating uncertainty is in practice unfeasible, as this would require carrying out 'perfect research', in the form of a study involving an infinitely large number of participants. In

reality, research studies, such as clinical trials, employ finite, and often small, sample sizes and are expected to reduce, but not eliminate, uncertainty[51;107].

The expected reduction in uncertainty due to conducting research of a particular sample size and, thus, the expected value of sample research, can be quantified by the expected value of sample information (EVSI) measure. The measure gives the actual expected NMBs from a particular study of a particular sample size. EVSI is calculated as the difference between the NMBs expected to accrue from a decision based on sample information obtained from a study of sample size n and the NMBs of a decision made with current information[51;97;115;178].

EVSI calculations are based on the concept of Bayesian inferrence[107;189]. In simple terms, before a study (e.g. a clinical trial) is conducted, adoption decisions are made in view of cost-effectiveness results calculated on the basis of existing (prior) evidence about parameters (e.g. existing evidence on a treatment's effectiveness). Conducting a study generates new evidence (e.g. evidence on disease progression), which can be combined with the prior evidence to provide improved, 'posterior' information. Decisions made on the basis of posterior information are expected to be less uncertain and, as a result, they will be associated with additional NMBs as compared to decisions made under existing evidence.

Given the above, the main aim in EVSI is to compare the benefits from a decision made with current information against the benefits expected to accrue from a decision made under 'improved' information obtained from conducting a proposed clinical trial. Improved

(posterior) information is a combination of existing information and sample (simulated) evidence from the trial.

Early applications of EVSI analysis have been carried out parametrically using simple mathematical formulae, under the assumption that results of economic evaluations—expressed, for example, as net monetary benefits—are normally distributed[97;115]. However, the recognition that results from decision analytic models are characterised by complex, non-linear structures and are very unlikely to be normally distributed[175] has turned the attention towards non-parametric simulation methods[190;269].

Combining prior and sample (trial) information is fairly straightforward when prior and sample evidence is expressed in terms of distributions which are conjugate, that is, they can be combined by using simple analytic solutions (e.g. beta-binomial, normal, gamma-Poisson distribution). Often, prior and sample data are expressed as distributions which are disconnected (e.g. prior information parameterised as Weibull distribution and sample information expressed as Normal distribution). When this is the case, as Brennan and Kharrubi[270[p.1206]] explain, combining existing evidence with simulated sample results is *"a substantial computational expense itself and must be repeated for each simulated data-set, which can result in very substantial computation times"*. In such cases, possible ways of combining prior and data involve Bayesian approximation methods and Bayesian updating in specialist software, typically WinBUGS[271].

EVSI can be calculated from the output of probabilistic analytic model using non-parametric methods, which usually involve running Monte Carlo simulations and taking advantage of

special relationships between distribution of parameters (i.e. conjugacy)[190;263]. Owing to the complexity of EVSI calculations, the method is typically restricted to assessing the value of a clinical trial in informing one or a group of similar parameters (e.g. disease progression), rather than all parameters investigated in the trial[51].

Assuming that a study of sample n is considered to provide evidence on a single parameter or a group of parameters of interest φ of all uncertain parameters θ, the per-patient EVSI can be calculated following the sequence of steps described in Box 7.3 and is given by the formula:

$$EVSI_n = E_D max_j \, E_{\varphi|D} \, NMB(j, \varphi) - \, max_j \, E_\theta \, NMB(j, \theta)$$

The EVSI for the population that is expected to benefit from a decision in the light of improved information is calculated in the same way as the population EVPI and EVPPI.

1. Draw a set of values of the uncertain parameters φ from their existing (prior) distributions.

2. On the basis of the drawn values, simulate possible sample results D on φ.

3. Combine the prior distribution and sample results obtained from step 2 to get a posterior distribution. Repeat steps 1 to 3 a large number of times (e.g. k =1000) for each treatment j.

4. Draw a large number of values (e.g. n =1000) from each of the posterior distributions obtained in step 3, and calculate the resulting NMBs for each treatment j.

5. Average across the NMBs obtained in step 4, to get the expected NMBs ($E_{\varphi|D}\ NMB(j,\varphi)$) for each posterior distribution and for each treatment j.

6. Find the maximum expected NMBs across treatments for each posterior distribution ($max_j\ E_{\varphi|D}\ NMB\ (j,\varphi)$).

7. As it is not known which posterior distribution (i.e. trial results) will transpire, average across the maximum expected NMBs to obtain the expected maximum NMBs ($E_D max_j\ E_{\varphi|D}\ NMB\ (j,\varphi)$). This represents the expected NMBs from making a decision with sample information.

8. Subtract the NMBs associated with a decision with sample information ($E_D max_j\ E_{\varphi|D}\ NMB\ (j,\varphi)$) from those based on a decision made under current information ($max_j\ E_\theta\ NMB\ (j,\theta)$) to get the EVSI.

Box 7.3: Steps in non-parametric calculation of EVSI

7.2. 'Value of information' analysis for NSCLC

This section reports the results of the application of EVPI, EVPPI and value of implementation analyses to the NSCLC case study.

7.2.1. Expected value of perfect information analysis for NSCLC

EVPI analysis was carried out to establish whether conducting further research to inform the choice between Gem+Cisp and Gem+Carb—that is, the relevant treatment adoption decision in NSCLC—would be potentially worthwhile or should be ruled out. EVPI results for the individual patient were obtained by using the output of the probabilistic NSCLC model. Individual (per-patient) and population EVPI results are given in Table 7.2.

Table 7.2: Individual and population EVPI (NSCLC)

Ceiling ratio	Individual	Population (five-year time horizon)
£0 per QALY	£66	£908,419
£30,000 per QALY	£948	£13,077,504
£80,000 per QALY	£2287	£39,827,506

Population EVPI was calculated by projecting the individual EVPI to the population of current and future patients which is expected to be affected by the NSCLC adoption decision. As noted in the previous chapter, assuming a five-year time horizon and an estimated number of 3830 eligible patients per year, the total discounted number of eligible patients was estimated to be 13,797.

At a ceiling ratio value of £0 per QALY (that is, when a decision-maker is not willing to pay any amount of money for additional QALYs), there is little uncertainty around Gem+Cisp being the optimal treatment. Thus, the value of eliminating this uncertainty, that is, the EVPI, is relatively low, at about £66 for the individual patient and £908,400 for the population. As the ceiling ratio increases, so does the uncertainty around Gem+Cisp being the optimal treatment, which, in turn, leads to increases in EVPI. This trend can be seen in the plots of the population EVPI shown in Figure 7.1. At £30,000 per QALY, the EVPI is £950 and £13.08 million for the individual and the population, respectively, while at a high ratio of £80,000 per QALY, individual and population EVPI values are £2290 and £39.83 million.

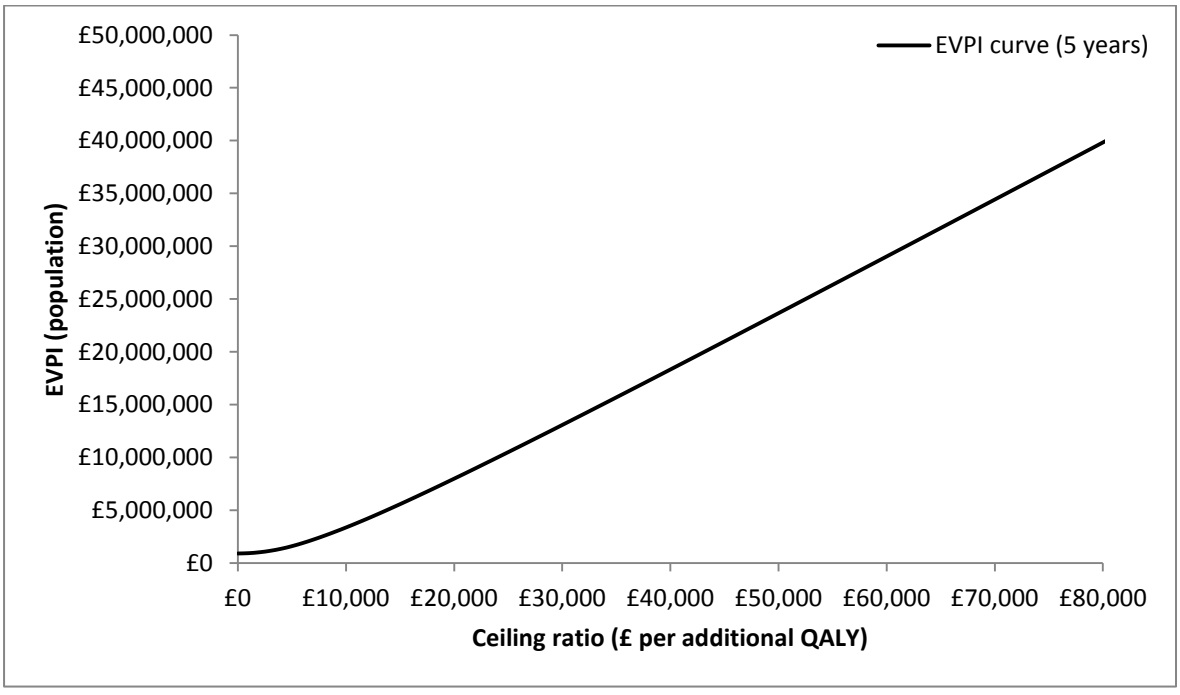

Figure 7.1: Population EVPI for NSCLC

The obtained EVPI results suggest that, at £30,000 per QALY, conducting further research (such as the proposed BTOG-2 trial) to provide further evidence around the NSCLC treatment adoption decision would be potentially worthwhile if the research programme costs less than £13.08 million. On this basis, funding and carrying out the proposed BTOG-2 trial, which costs £336,700, would be a potentially worthwhile investment.

7.2.2. Expected value of perfect parameter information analysis for NSCLC

Additional analysis was carried out to investigate the expected gains from acquiring perfect information about specific parameters. Five groups of parameters were formed:

a. utility values for the 'Progression-free' and 'Progression' health states in the NSCLC model;

b. transition probabilities (rates of progression and survival) associated with Gem+Cisp;

c. transition probabilities (rates of progression and survival) associated with Gem+Carb;

d. costs associated with Gem+Cisp, and

e. costs associated with Gem+Carb.

EVPPI calculations were carried out through nested two-level (inner and outer loop) Monte Carlo computations[51;263]. The analysis involved a relatively high number of simulations—

1000 for the inner loop and 1000 for the outer loop—in order to improve accuracy in the results and was carried out in Microsoft Excel 2007® using commands (macros) written in the VBA® programming language.

Individual and population EVPPI for different parameters assuming a ceiling ratio of £30,000 per QALY are given in Figure 7.2 and Figure 7.3. Parameters with the highest individual EVPPI were transition probabilities for Gem+Cisp and Gem+Carb, at £400 and £657, respectively, while the equivalent values for all eligible patients over five years were £5.52 million and £9.07 million. On the other hand, the value of research associated with Gem+Cisp and Gem+Carb costs appeared low, at £4 (£51,250) and £3 (£35,000) for the individual patient (entire population), respectively. The EVPPI associated with utility scores at £30,000 per QALY was zero for both an individual patient and the population.

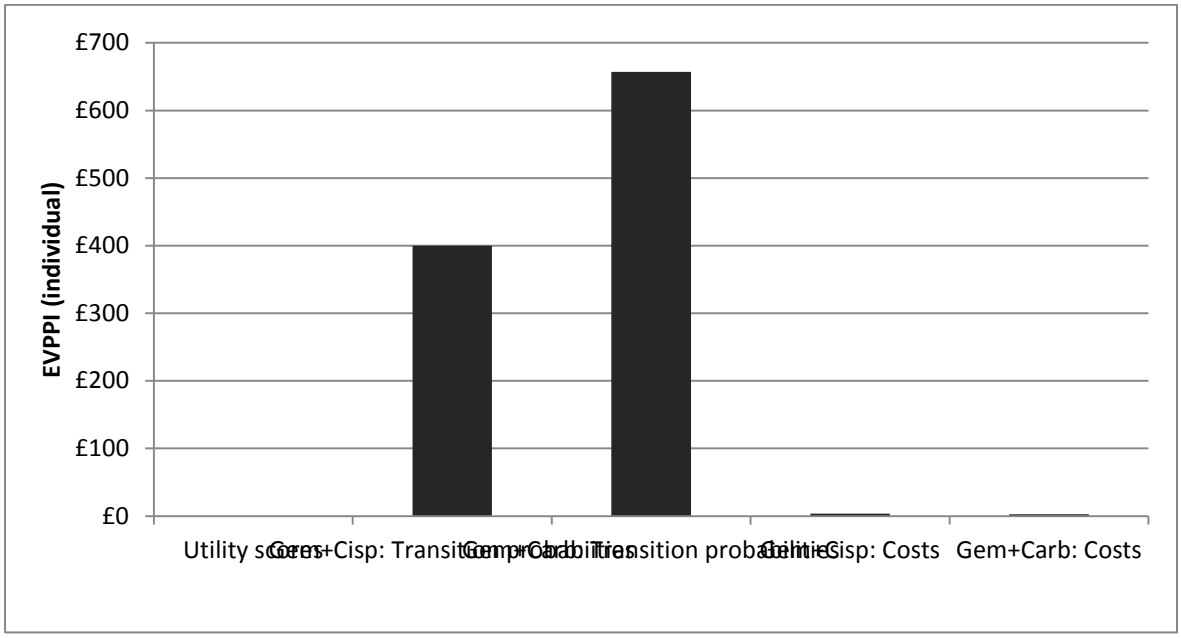

Figure 7.2: Individual patient EVPPI for NSCLC at £30,000 per QALY

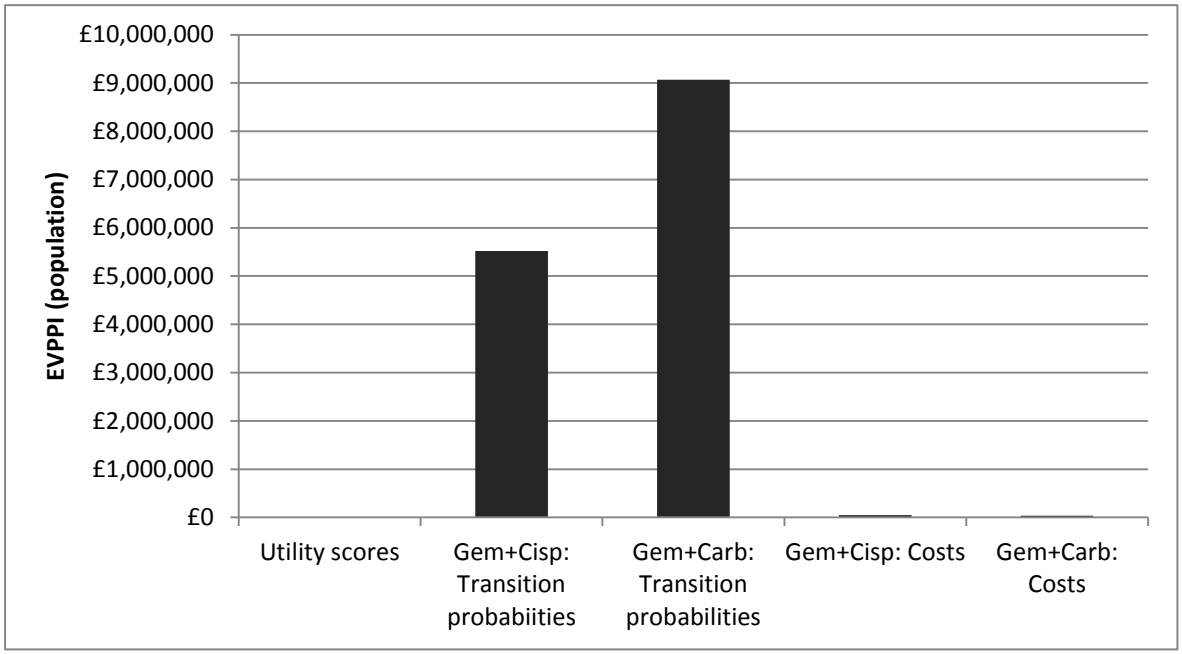

Figure 7.3: Population EVPPI for NSCLC at £30,000 per QALY

Curves showing the population EVPPI for different groups of parameters are given in Figure 7.4. With regards to transition probabilities, higher values of the ceiling ratio led to higher EVPPI. At a high ceiling ratio of £80,000 per QALY, EVPPI values for transition probabilities for Gem+Cisp and Gem+Carb reached £20.41 million and £29.18 million, respectively. The EVPPI associated with utility scores started to increase at ceiling ratios over £86,000 per QALY and, at the very extreme value of £250,000 per QALY (not shown here) it only reaches about £97,700. On the other hand, the population EVPPI values for cost parameters declined as the ceiling ratio rose and, at £80,000 per QALY they were low, at £215 and £7080 for Gem+Cisp and Gem+Carb, respectively. The decline in these values reflects the fact that uncertainty around costs has an impact on the choice between Gem+Cisp and

Gem+Carb when the ceiling ratio is low (i.e. when the cost of treatments is important) but, as the ceiling ratio increases, the importance of costs in overall decision uncertainty diminishes.

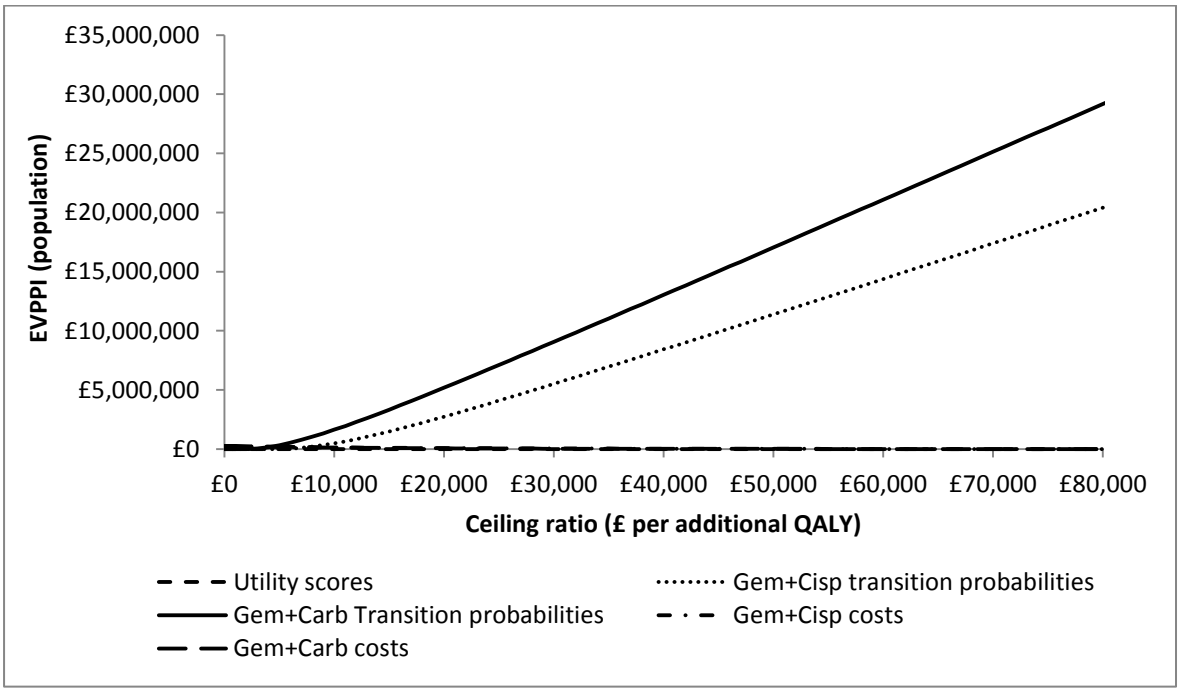

Figure 7.4: Population EVPPI at different ceiling ratios for NSCLC

In summary, the results of the EVPPI analysis suggest high returns from research looking into disease progression rates for Gem+Cisp and Gem+Carb. At £30,000 per QALY, further research on transition probabilities would be potentially cost-effective if the research study costs less than £5.52 million. Given the fact that the proposed BTOG-2 trial has a cost of £336,700, the analysis suggests that the trial would be potentially cost-effective. On the other hand, the EVPPI associated with research on preference-based quality of life (utility) scores was particularly low (£0 at £30,000 per QALY) suggesting that further research on

utility scores should be ruled out. Similarly, EVPPI for costs was low and it is unlikely that research on this parameter only would be potentially cost-effective; nonetheless, there may be scope for including cost and utility parameters as additional endpoints in a trial with a primary focus on disease progression rates.

7.2.3. Impact of parameter uncertainty on results

As explained above, VoI measures translate the uncertainty around a decision problem into possible gains expected from obtaining better information. Because of this, EVPI and EVPPI results are highly dependent on the degree of the uncertainty around a decision, with higher uncertainty resulting in higher EVPI and EVPPI. Often, estimates of the uncertainty around parameters affecting a decision are not available and, in such situations, they are replaced by assumptions. With this in mind, further analysis was carried out to explore how different assumptions about the extent of uncertainty around key parameters in the NSCLC model may affect the generated VoI results. In doing so, it was assumed that uncertainty around key parameters, represented by standard errors, may take on values double or half as high as the base case values. Results of this analysis are given in Table 7.3.

Introducing greater uncertainty around cost and preference-based quality of life parameters had a small effect on EVPI results, which became less than five percent higher than the base case value. However, increasing the standard errors around transition probabilities for Gem+Cisp and Gem+Carb had a sizeable impact on EVPI, which rose by about 40 percent and 80 percent, respectively. This is explained by the fact that transition probabilities are a

key determinant in the choice between Gem+Cisp and Gem+Carb (i.e. in the adoption decision) and thus uncertainty around such parameters has an impact on the uncertainty around the adoption decision and, in turn, on the EVPI. As expected, an effect of the opposite direction was observed when the standard errors around parameters were set at lower values. In this case, the greatest reduction in EVPI (about 43 percent) resulted by decreasing the uncertainty around transition probabilities for Gem+Carb.

Table 7.3: Results of illustratory sensitivity analysis on impact of uncertainty on EVPI

Varied parameter	**Population EVPI** (£30,000 per QALY; five-year time horizon)
Base case	£13,077,504
Gem+Cisp transition probabilities (standard error twice the base case value)	£17,941,963
Gem+Carb transition probabilities (standard error twice the base case value)	£29,264,503
Gem+Cisp costs (standard error twice the base case value)	£13,280,117
Gem+Carb costs (standard error twice the base case value)	£13,288,161
Preference-based quality of life (utility) scores (standard error twice the base case value)	£13,538,582
Gem+Cisp transition probabilities (standard error half the base case value)	£11,376,279
Gem+Carb transition probabilities (standard error half the base case value)	£7,422,768
Gem+Cisp cost (standard error half the base case value)	£12,697,379
Gem+Carb cost (standard error half the base case value)	£12,895,010
Preference-based quality of life (utility) scores (standard error half the base case value)	£12,928,053

As noted in Chapter 4, estimates of preference-based quality of life (utility scores) for different states of advanced NSCLC were not available in the literature. Thus, mean utility

scores and estimates of the uncertainty around them (standard errors) were obtained from expert opinion. On the basis of these scores, the EVPPI analysis reported above showed a very low value of research on utility values. Additional EVPPI analyses were carried out to explore whether assigning greater uncertainty around the utility scores (i.e. higher standard errors than those specified by experts) would suggest this parameter as a good candidate for further research. However, even when the uncertainty around utility scores was doubled, EVPPI at £30,000 per QALY was zero, suggesting that, at conventional ceiling ratio values, the impact of utility scores on the adoption decision and thus the value of obtaining better estimates around such scores, is negligible.

7.2.4. Value of information and implementation analysis for NSCLC

VoI analysis was extended to take into account different situations regarding availability of information and treatments' uptake (implementation) using the conceptual framework introduced by Fenwick and colleagues[187]. The main analysis was based on the assumption that there is no direct relationship between acquisition of information and treatment uptake (i.e. implementation), that is, uptake changes only in response to implementation strategies. Further analysis was carried out to illustrate a possible extension to the framework by looking into the more realistic scenario where availability of improved information has an impact on implementation. In these analyses, results reflect NMBs expected to accrue from decisions made under different 'states of the world' related to information and implementation, using a ceiling ratio of £30,000 per QALY.

7.2.4.A. Assuming acquisition of information has no impact on implementation

In this analysis, information is needed on 'current implementation' (that is, treatments' prescription shares in the absence of any implementation strategies) and 'optimal' implementation. Information on current implementation, that is treatments' prescription shares in the year evidence from research is expected to become available (i.e. 2011), suggested that each of Gem+Cisp and Gem+Carb is provided to 50 percent of the eligible patients. On the other hand, optimal implementation means that the treatment that appears cost-effective would be implemented perfectly, (i.e. it would be provided to the entire population of eligible patients) and, on the other hand, the treatment which is not cost-effective will not be provided to any patients.

The expected NMBs for different 'states of the world', assuming that improved information has no impact on the level of implementation, are given in Table 7.4. A treatment adoption decision with current information implemented at current levels (state A) is expected to bring about benefits of £152.70 million, the same as in the case of making a decision with perfect information and current implementation (state B). A decision with current information followed by optimal implementation (state C) is expected to result in NMBs of approximately £160.76 million, while a decision with perfect information that will be implemented optimally (i.e. the 'ideal' situation), is expected to result in NMBs of about £173.84 million (state D).

Table 7.4: Expected NMBs for different 'states of the world' regarding information and implementation (NSCLC)

States of the world		Information	
		Current	Perfect
Implementation	Current	£152,696,292 (state A)	£152,696,292 (state B)
	Optimal	£160,760,214 (state C)	£173,837,480 (state D)

On the basis of these results, one can calculate different measures of the value of acquiring better information and improving implementation. These measures are given in Table 7.5. The EVPI (i.e. difference between states D and C) was estimated at £13.08 million. This figure represents the expected NMBs from a decision made with perfect as opposed to current information given that in either case the decision will be implemented optimally, and indicates the maximum returns to pursuing better information through research.

Table 7.5: Expected NMBs for different measures of information and implementation (NSCLC)

Measure*	NMBs (at £30,000 per QALY)
EVPI (state D – state C)	£13,077,266
rEVPI (state B – state A)	£0
EVPIM$_{\text{perfect information}}$ (state D – state B)	£21,141,187
EVPIM$_{\text{current information}}$ (state C – state A)	£8,063,922
EVP (state D – state A)	£21,141,187
EVIIT (state B – state C)	-£8,063,922
*Featured measures are explained in the following text	

The comparison between current and optimal implementation gives the expected value of implementation (EVPIM). Under current information, that is, with Gem+Cisp being superior

to Gem+Carb, the difference between providing the superior treatment to the whole population (i.e. optimal implementation) and continuing with current prescription shares (i.e. current implementation) is £8.06 million. This value is the $EVPIM_{current\ information}$ and represents the maximum expected gains from investing resources in the pursuit of better implementation of the treatment which is believed to be cost-effective under current evidence.

On the other hand, the $EVPIM_{perfect\ information}$ is the difference between the expected benefits obtained with current and optimal implementation, assuming that perfect information is available. The $EVPIM_{perfect\ information}$ was estimated at £21.14 million and represents the maximum gains from ensuring optimal implementation of the treatment that appears cost-effective under perfect information. This is seen as the maximum value that decision-makers should commit to implementation strategies under the premise that perfect information is available[187].

This comparison between the value of a decision made with perfect information and optimal implementation, and one made with current information and current implementation gives the expected value of perfection (EVP). In this analysis, EVP is £21.14 million, the same as the $EVPIM_{perfect\ information}$, owing to the assumption that acquiring perfect information does not improve implementation and it is only implementation strategies that can affect treatment uptake in clinical practice. In other words, as long as the practice remains at current levels, it is irrelevant whether there is perfect or current information. As

a result, EVP provides the maximum gains from, and thus the maximum value of, acquiring perfect information and pursuing perfect implementation.

The difference between the expected NMBs in a situation with perfect information and current implementation (state B) and one with current information and current implementation (state A) gave the realisable EVPI (rEVPI). As expected, under the assumption that acquiring perfect information does not improve implementation, there is no 'realisable' benefits from acquiring perfect information, that is, rEVPI is zero.

The final comparison is between perfect information and current implementation (state B), and current information and optimal implementation (state C). Observing a positive difference from this comparison would mean that acquiring further information should take priority over investing in better implementation, assuming that the cost of further research and the cost of implementation programmes are comparable. If the difference is negative, the gains from a decision with current information and optimal implementation are expected to exceed those from a decision with perfect information which will only be implemented at current, sub-optimal levels. Obviously, a decision between investing in research or implementation will also depend on the cost of information and implementation programmes. To the author's best knowledge, no published work has explored this comparison before, which, for the purposes of this project, has been termed 'expected value of information-implementation trade-off' (EVIIT).

EVIIT is useful in that it shows whether priority should be given to further research, or resources should be committed to programmes that would improve adherence to the treatment that appears cost-effective under current information. The results gave a value of -£8.06 million, suggesting that a decision under perfect information which will be implemented only at current rates will be expected to result in lower NMBs than a perfectly implemented decision made under current information. Assuming comparable costs for research and implementation programmes, EVIIT suggested that pursuing better implementation is potentially more beneficial than carrying out a study such as the BTOG-2 trial to obtain better information.

7.2.4.B. Assuming acquisition of information affects implementation

Fenwick *et al*.[124] recognise that the assumption of no interaction between acquisition of information and implementation (i.e. only implementation strategies can improve adherence to cost-effective treatments) may be simplistic. Given this, additional analysis was carried out to look into a situation where acquisition of information is expected to affect implementation. This illustrative analysis makes use of estimates of future prescription shares obtained for the purposes of the 'payback' analysis through discussion with experts at the Pan-Birmingham Cancer Network (Table 7.6).

In general, having better information is expected to lead to an increase in the uptake of the treatments shown to be cost-effective and a decline in the use of non-cost-effective treatments. However, in the absence of active implementation strategies, uptake would be

expected to only reach an 'improved' level of implementation, which would be higher than the equivalent level under the assumption of no interaction between information and implementation, but lower than the 'optimal' implementation level.

Table 7.6: Prescription shares in the light of different eventualities about treatments' cost-effectiveness (NSCLC)

Implementation	Eventuality	Percentage use of Gem+Cisp	Percentage use of Gem+Carb
Current implementation		50%	50%
Improved implementation	**Gem+Cisp cost-effective**	75%	25%
	Gem+Carb cost-effective	25%	75%

For example, if further information showed Gem+Cisp to be the most cost-effective treatment, this was hypothesised to bring about an increase in the treatment's prescription share, from the current level of 50 percent to 75 percent and, equivalently, trigger a decrease in the use of the non-cost-effective treatment (Gem+Carb), from 50 percent to 25 percent. Given this, the expected NMBs given perfect information and 'improved' implementation will be approximately £163.27 million (Table 7.7).

Table 7.7: Revised expected NMBs for different 'states of the world' regarding information and implementation (NSCLC)

State of the world		NMBs (at £30,000 per QALY)
Current information	**Current implementation** (state A)	£152,696,292
	Optimal implementation (state C)	£160,760,214
Perfect information	**Improved implementation** (state B)	£163,266,886
	Optimal implementation (state D)	£173,837,480

The assumption that perfect information induces a beneficial change in implementation has an impact on three measures: the rEVPI, the $EVPIM_{perfect\ information}$ and the EVIIT. Revised results for these measures are shown in Table 7.8.

Table 7.8: Revised expected NMBs for different measures of information and implementation (NSCLC)

Measure	NMBs (at £30,000 per QALY)
rEVPI (state B- state A)	£10,570,594
$EVPIM_{perfect\ information}$ (state D- state B)	£10,570,594
EVIIT (state B- state C)	£2,506,672

The revised rEVPI was estimated to be £10.57 million. This value shows the difference in the expected NMBs between a decision with perfect information and improved implementation, and one with current information and current implementation. Similarly, the $EVPIM_{perfect\ information}$ is revised and it now shows the difference between perfect information followed by optimal implementation and perfect information followed by improved implementation. This value was calculated to be £10.57 million, which is half of the £21.14 million estimated under the assumption of no connection between information and implementation. This reflects the fact that the revised $EVPIM_{perfect\ information}$ is calculated given that there has already been a change of 25 percent in prescription shares after further information became available (i.e. from 'current' implementation to 'improved' implementation) therefore the scope for further improvements in uptake through implementation—and, thus, the scope for further NMBs—is more limited.

Finally, when the assumption of no interaction between improved information and implementation is relaxed, the revised EVIIT is positive. This is in contrast to the negative EVIIT calculated with the above assumption holding and it reflects the fact that, in the current case, research is expected to bring about additional benefits due to inducing greater adherence to the recommended treatment. Under the assumption that further research will provide information as well as prompt a beneficial change in practice, carrying out a trial such as the BTOG-2 would be potentially more beneficial than funding implementation strategies.

7.2.5. Expected value of sample information and expected net benefits of sampling for NSCLC

As explained at the beginning of this chapter (section 7.1.4) a central task in undertaking EVSI analysis involves combining existing (prior) information (in NSCLC, taken from Zatloukal *et al*.[208]) with new (sample) information which is hypothesised to arise from a proposed trial. This task typically represents the main difficulty in undertaking EVSI, as combining prior and sample information represented by probability distributions requires these distributions to be connected by a special relationship, termed 'conjugacy', where one quantity can be combined with the other using a simple analytic solution. Examples of conjugate distributions are the beta/Dirichlet distributions, gamma distribution and the normal distribution[261]. When parameters of interest (e.g. progression to a worse health state) are not expressed as conjugate distributions (as was the case in HRPC, where progression was derived from counts of events and was represented as beta and Dirichlet distributions),

there is no simple analytic solution, and other ways of combining prior and sample information need to be explored.

As explained in Chapter 4, evidence on the effectiveness of Gem+Cisp and Gem+Carb was obtained from a randomised controlled trial following 87 and 89 patients on Gem+Cisp and Gem+Carb, respectively. The outcomes of the trial were given in the form of overall survival (OS) and time-to-progression (progression-free survival (PFS)) curves, showing the probability of a NSCLC patient being alive and progression-free, respectively, at different points in time (months) after randomisation. This information was used as the basis for fitting Weibull overall survival and progression-free survival curves to the available data by using ordinary least squares (OLS) regression (see Chapter 4, section 4.4.2):

$$ln[-lnS(t)] = intercept + coefficient \times ln(t) + \varepsilon$$

where $S(t)$ can be either of the OS and PFS curve, t is time in months after the onset of the observation and ε is the random error term of the regression model. The model intercept and the regression coefficient of $ln(t)$ (i.e. the slope of the fitted regression model) were used to derive the shape parameter alpha (α) and the scale parameter beta (β) of the Weibull functions for OS and PFS. The specified Weibull functions, which can be seen in Appendix 3.C, gave the probability of a patient being alive and progression-free at different points in time, which was used to populate the three health states of the NSCLC model. Estimates of uncertainty around these probabilities were obtained directly from the regression analysis, in terms of the standard errors around the intercept and regression coefficient parameters.

7.2.5.A. Process for obtaining posterior distributions of effectiveness (progression-free survival and overall survival)

A way to circumvent the problem of combining non-conjugate distributions such as the Weibull distributions used in NSCLC is by representing prior and sample information in terms of the same type of outcomes. In this application, an appropriate and convenient representation is in terms of numbers of patients who are alive and progression-free at different points in time.

Translating prior information from Zatloukal *et al.*[208] to numbers of patients alive and progression-free at different points was relatively straightforward. Under the assumption that patients had not been lost to follow-up, the PFS and OS curves can be obtained by multiplying the probability of a patient being at a particular state (alive or progression-free) by the number of patients in a trial arm at the beginning of the trial

$$Number\ of\ patients\ alive\ at\ t_i = n \times P(alive\ at\ t_i)$$

$$Number\ of\ progression\ free\ patients\ at\ t_i = n \times P(progression\ free\ at\ t_i)$$

Here, t_i represents a point in time after randomisation (e.g. month 3) and n represents the number of participants in the trial arm at the onset of the follow-up period. For example, for patients on Gem+Cisp, given a probability of being progression-free at 3 months of 0.75, 65 out of the total 87 patients on treatment would be expected to be alive.

In terms of sample information, hypothetical results of a proposed trial—expressed in terms of numbers of progression-free and alive patients at different points in time—were

generated through individual patient sampling. The process of individual sampling aimed to simulate the results that a future trial of a specific sample size may produce, based on current indications of the effectiveness of the treatments. The process involved the following steps.

First, a set of parameters (intercept and regression coefficient, and resulting α and β parameters of the Weibull model) was drawn from the existing distribution of the intercept and regression coefficient obtained from the OLS regression. In essence, these parameter gave possible PFS and OS drawn from the distribution of the prior OS and PFS. For example, if the drawn intercept and regression coefficient are -2.88 and 1.34 for PFS and -3.01 and 1.09 for OS, the resulting Weibull curves will give a probability of a patient being progression-free at 3 months of 0.8 and a probability of death at the same point in time of 0.05.

An individual patient who starts the trial is expected to progress according to the drawn parameters and the resulting PFS and OS curves. To replicate the stochastic process of patient progression, at each point in time a random number between 0 and 1 is drawn. Using the probabilities in the previous paragraph, if the number is lower or equal to 0.8, the patient remains progression-free, if the number is between 0.8 and 0.95 (i.e. 1-0.05) the person presents progressive disease, and if the number is between 0.95 and 1, the patient dies. For later time points, appropriate conditional probabilities were used.

Repeating this process for patients n equal to the number of the proposed trial gives numbers of patients who are alive, progression-free and dead at specific points in time.

These represent the simulated sample information produced by a hypothetical trial of n patients. The sample size of the proposed BTOG2 trial was 450 patients per arm, and thus the size n of the simulated trials is 450 patients. As these results represent one possible outcome of the trial (based on one draw from the prior distribution of the parameters of interest), the process is repeated a large number of times (k = 1000) to give 1000 simulated trial results, as seen in Table 7.9.

Table 7.9: Illustrative table showing data (number of patients alive on different follow-up points) generated from 1000 simulated trials for Gem+Cisp

Time point (month)	**0**	**3**	**6**	**…**	**27**	**30**
Prior information (derived from Zatloukal et al.[208])	87	72	59		9	9
Sample information$_{\text{trial 1}}$	450	382	312	…	37	31
Sample information$_{\text{trial 2}}$	450	355	258	…	14	10
Sample information$_{\text{trial 3}}$	450	383	307	…	74	54
…	…	…	…	…	…	…
Sample information$_{\text{trial 1000}}$	450	397	327	…	63	55

The process was repeated to replicate the sample OS and PFS curves for both Gem+Cisp and Gem+Carb. Having obtained 1000 simulated data, each of them was combined with prior information as

$$Posterior_i = Prior + Sample\ information_{trial\ i}$$

For example, for OS in Gem+Cisp at three months, the results for Posterior$_1$—expressed in terms of number of patients alive—are equal to the prior information (72 patients alive at month 3) and simulated sample information$_{\text{trial 1}}$ (382 patients alive) resulting in a total of

454 patients alive at month 3 out of the 537 (i.e. 87+450) patients under observation, giving a probability of being alive at 3 months of approximately 0.85.

Having obtained a representation of the posterior distribution, this information needs to be translated into Weibull curves, which can then be used in the NSCLC model. For each of the posterior values for OS and PFS, Weibull curves were fitted by the same process that was used to fit Weibull curves to prior information from Zatloukal *et al*.[208]. In brief, fitting Weibull to posterior data involved translating number of patients (fourth row in Table 7.10) into probabilities of a patient staying alive at different points (fifth row in Table Table 7.10), manipulating the Weibull survival function to obtain a linear relationship between time (logarithm of time) and $S(t)$ (logarithm of the negative logarithm of the survival function) and regressing $S(t)$ against time.

Table 7.10: Table illustrating the steps in translating posterior information to quantities (ln(t) and ln(-ln(S(t))) to be used in OLS regression.

Month	0	3	6	9	12	15	18	21	24	27	30
Prior information (number of patients alive at different points)	87	72	59	43	29	23	15	15	15	9	9
Sample information$_{\text{trial 1}}$ (number of patients alive at different points)	450	382	312	240	187	142	106	79	55	37	31
Post$_1$ (no of patients alive at different points)	537	454	371	283	216	165	121	94	70	46	40
S(t|Post$_1$) (i.e. probability of patient being alive on basis of Post$_1$)	1	0.85	0.69	0.53	0.4	0.31	0.23	0.18	0.13	0.09	0.07
Ln(t) (i.e. natural logarithm of time)	-	1.1	1.79	2.2	2.48	2.71	2.89	3.04	3.18	3.3	3.4
Ln(-ln(S(t|Post$_1$)))	-	-1.78	-0.99	-0.44	-0.09	0.16	0.4	0.56	0.71	0.9	0.95

For each $Post_i$ (i=1,2,…1000), OLS regression gave a set of parameters (intercept, regression coefficient of time, standard errors of the intercept and regression coefficient) and the resulting α and β parameters of the Weibull models for PFS and OS (see Table 7.11 and Table 7.12 for an example of generated data for Gem+Cisp and Gem+Carb).

Each of the obtained 1000 sets (i.e. $Post_1$, $Post_2$,…, $Post_{1000}$) was treated as a representation of the true progression-free survival and overall survival associated with Gem+Cisp and Gem+Carb. These sets were subsequently entered in the NSCLC model one at a time and, for each set, 1000 Monte Carlo simulations were carried out to give 1000 estimates of each treatment's NMBs given the specific posterior.

Table 7.11: Example of set of parameters representing the posterior distributions of PFS and OS for Gem+Cisp given a trial of 450 patients.

Posteriors (translated into Weibull curves)	**Progression-free survival (PFS)**						**Overall survival (OS)**					
	Intercept	SE intercept	Slope	SE slope	alpha	beta	Intercept	SE intercept	Slope	SE slope	alpha	beta
$Post_1$	-3.298	0.038	1.530	0.016	1.53	12.50	-3.117	0.033	1.209	0.012	1.21	19.08
$Post_2$	-2.735	0.043	1.227	0.019	1.23	13.46	-2.772	0.034	1.181	0.013	1.18	15.13
…												
$Post_{1000}$	-3.132	0.031	1.485	0.014	1.48	11.93	-3.272	0.053	1.206	0.020	1.21	21.84

Table 7.12: Example of set of parameters representing the posterior distributions of PFS and OS for Gem+Carb given a trial of 450 patients.

Posteriors (translated into Weibull curves)	**Progression-free survival (PFS)**						**Overall survival (OS)**					
	Intercept	SE intercept	Slope	SE slope	alpha	beta	Intercept	SE intercept	Slope	SE slope	alpha	beta
$Post_1$	-2.336	0.030	1.248	0.013	1.248	9.417	-3.176	0.053	1.249	0.020	1.249	18.40
$Post_2$	-2.502	0.026	1.169	0.011	1.169	12.304	-3.645	0.046	1.271	0.017	1.271	25.49
…												
$Post_{1000}$	-2.180	0.033	1.205	0.014	1.205	8.835	-3.022	0.074	1.207	0.028	1.207	17.705

7.2.5.B. EVSI and ENBS results

Generated NMBs estimates were averaged across their distribution to give an estimate of the expected NMBs ($E_{\varphi|Post_i}NMB(j,\varphi)$) given the specific posterior value. The process was repeated for all the 1000 posteriors, to give 1000 estimates of $E_{\varphi|Post_i}NMB(j,\varphi)$ for each treatment (second and third columns in Table 7.13), of which estimates the maximum expected NMBs ($max_j\ E_{\varphi|Post_i}NMB(j,\varphi)$) were selected (fourth column in Table 7.13).

Table 7.13: Illustrative example of EVSI calculations for a trial in NSCLC of sample size of 450 patients per arm

	Gem+Cisp	**Gem+Carb**	$\boldsymbol{max_j E_{\varphi\|Post_i} NMB\ (j,\varphi)}$
$\mathbf{E_{\phi\|Post_1} NMB(j,\varphi)}$	£11,728	£9,790	£11,728
$\mathbf{E_{\phi\|Post_2} NMB(j,\varphi)}$	£9,674	£14,849	£14,849
$\mathbf{E_{\phi\|Post_3} NMB(j,\varphi)}$	£13,047	£10,032	£13,047
…	…	…	…
$\mathbf{E_{\varphi\|Post_{999}} NMB(j,\varphi)}$	£9,825	£10,983	£10,983
$\mathbf{E_{\varphi\|Post_{1000}} NMB(j,\varphi)}$	£13,254	£9,350	£13,254
$\boldsymbol{E_{Post_i}\ max_j\ E_{\varphi\|Post_i} NMB(j,\varphi)}$			£12,328

As it is not known which of the $Sample\ information_{trial\ i}$ will be observed (i.e. which posterior distribution represents the true underlying distribution), the $max\ E_{\varphi|Post_i}NMB(j,\varphi)$ need to be averaged over their posterior distribution, to give the $E_{Post_i}\ max_j\ E_{\varphi|Post_i}NMB(j,\varphi)$ (bottom right cell in Table 7.13).

This value represents an estimate of the expected NMBs from making a decision under sample information and, when compared to the expected NMBs of making a decision under

current information ($max_j E_\varphi NMB(j, \varphi)$), it gives the EVSI$_n$ for the particular trial of size n. In this analysis, for the proposed BTOG2 trial of 450 patients per treatment arm and given a ceiling ratio of £30,000 per QALY, the expected maximum NMBs with sample information were estimated at £12,328. Given the estimated expected NMBs from a decision under current information of £11,652 at £30,000 per QALY, EVSI for an individual patient was found to be

$$EVSI_{n=450} = NMB_{sample\ info\ (n=450)} - NMB_{current\ info} = £12{,}328 - £11{,}652 = £677$$

EVSI for the population of eligible patients over five years following dissemination was calculated at £9.33 million (at £30,000 per QALY).

Similarly to other measures of value of information, EVSI results vary according to the assumed value of the ceiling ratio. The population EVSI (dotted line) for ceiling ratios ranging from £0 to £80,000 can be seen in Figure 7.5 alongside the population EVPI (solid line). As anticipated, at each ceiling ratio the benefits from a decision with sample information (EVSI) are lower than those of a decision made under perfect information (EVPI). The EVSI curve follows a pattern analogous to the EVPI curve; in this case of NSCLC, EVSI increases for higher values of the ceiling ratio.

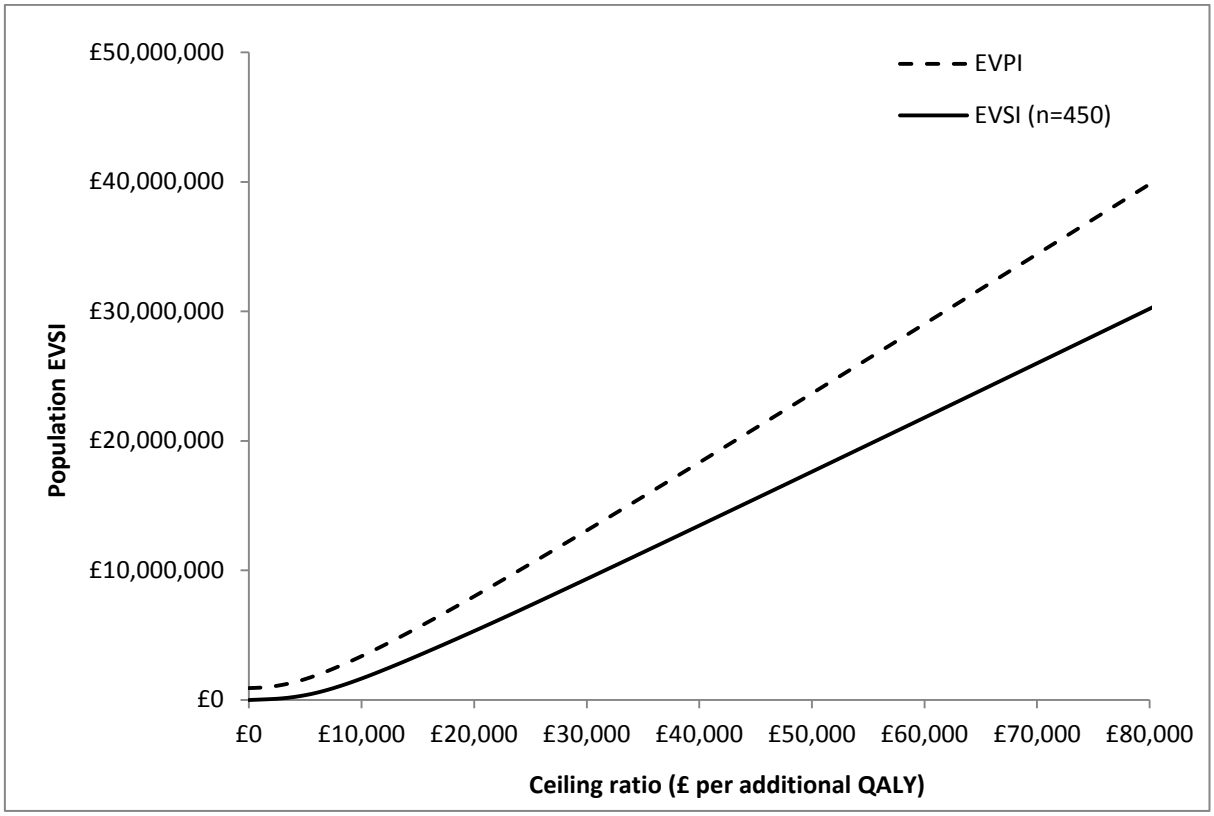

Figure 7.5: Population EVSI (for sample size of 450 patients per arm) and EVPI at different ceiling ratio values

The obtained EVSI estimates can be compared with the cost of the trial to give the expected net benefits of sampling (ENBS) for a trial such as BTOG2. In the context of the NSCLC case study, the cost of the BTOG2 phase III trial consisted of a fixed component, which is assumed not vary with the size of the trial and includes researchers' salaries and expenditure for trial co-ordination (£134,220), as well as a variable part, which depends on the size of the trial. This includes a 'support service' cost (£202,500) covering expenditures for patient recruitment, collection of QoL information through questionnaires and additional clinic time. As patients in the trial were treated with chemotherapies which would have been also used in usual practice outside the trial, no excess treatment cost due to receiving care different than that routinely provided was included. On the basis of the above, the marginal cost of

recruiting an additional patient in the trial was estimated at a relatively low value of £225. The total cost of the trial of 450 patients per treatment arm, including fixed and variable costs, was stated as £336,720. Given this, the ENBS for a trial of 450 participants per treatment arm was estimated at:

$$ENBS_{n=450} = EVSI_{n=450} - trial\ cost_{n=450} = £8{,}997{,}161$$

According to the obtained ENBS figure, the proposed trial of 450 patients per arm aiming to obtain better information on disease progression for NSCLC patients treated with Gem+Cisp and Gem+Carb would result in ENBS of about £9 million.

Additional analyses were carried out for different possible sample sizes (400 and 500 participants per arm). As expected, increases in the number of trial participants led to decreases in the uncertainty around the treatments' effectiveness and increases in the EVSI (second and third columns in Table 7.14). At the same time, recruiting more participants raises the cost of the trial due to increase in variable costs (fourth column in Table 7.14).

For the particular sample sizes investigated here (n=400, 450 and 500 patients per arm), the expected benefits exceeded the increase in costs (i.e. there are positive marginal benefits), thus, a study involving any of these sample sizes would be beneficial. This will be the case up to the point where the additional benefits from recruiting an additional participant would be offset by the additional cost from involving this extra person. Assuming no additional benefits from allocating participants unequally between arms, the 'optimal' sample size is found at the level where the ENBS is maximum, that is, when the difference between trial cost and NMBs is the greatest possible. For a ceiling ratio of £30,000 per QALY, the resulting

ENBS suggest that the preferable sample size for BTOG2 among the sizes considered here would be 500 patients in each arm. As in this proposal the cost of recruiting further patients is very low due to the fact that no excess treatment costs are involved, the 'value' of recruiting more patients comfortably exceeds the marginal cost of recruitment, and, it is anticipated that this will also be the case for sample sizes over 500 patients per arm.

Table 7.14: Individual EVSI for studies of different sample sizes

Trial sample size	**Individual EVSI** (£30,000 per QALY)	**Population EVSI** (£30,000 per QALY)	**Total cost of trial** (£30,000 per QALY)	**ENBS** (£30,000 per QALY)
400 participants per arm	£654	£9,023,757	£314,221	£8,709,536
450 participants per arm	£677	£9,333,882	£336,721	£8,997,161
500 participants per arm	£758	£10,453,550	£359,221	£10,094,329

In summary, the EVSI analysis suggested that a proposed trial to provide better information on disease progression employing 450 patients per arm and of a cost of about £340,000, such as the BTOG2 phase III trial, would result in benefits in excess of its cost. On this basis, carrying out such a trial appears to be a 'cost-effective' use of research funds.

7.3. 'Value of information' analysis for HRPC

This section presents the results of 'value of information' analyses (EVPI, EVPPI, value of implementation and EVSI) applied to the HRPC case study.

7.3.1. Expected value of perfect information for HRPC

EVPI for an individual patient was estimated by using the output (i.e. simulated costs and QALYs) of the probabilistic HRPC model, which compared docetaxel plus prednisolone (DP), against DP plus zoledronic acid (DP+ZA), DP plus strontium-89 (DP+Sr89) and DP plus zolendronic acid and strontium-89 (DP+ZA+Sr89) as a first-line treatment option for HRPC. Individual patient EVPI was subsequently weighted by the discounted number of eligible patients over a two-year time horizon (5101 patients; see Chapter 6) to give the population EVPI. Individual and population EVPI results are given in Table 7.15. At a ceiling ratio of £0 per QALY, the per-patient EVPI was estimated at approximately £190 (£947,100 for the population), while at a ceiling ratio of £30,000, the value rose to £1680 (£8.55 million for the population). At a very high ceiling ratio of £80,000, individual and population EVPI were approximately £4330 and £22.09 million, respectively.

Table 7.15: Individual and population EVPI for HRPC

Ceiling ratio	Individual EVPI	Population EVPI (two year time horizon)
£0 per QALY	£186	£947,120
£30,000 per QALY	£1676	£8,550,438
£80,000 per QALY	£4331	£22,093,720

The relationship between EVPI and the ceiling ratio can be seen in the EVPI curve in Figure 7.6. Starting from £0 per QALY, as the ceiling ratio increases, the EVPI rises. When the ceiling ratio becomes equal to the ICER for DP+Sr89—the point where the NMBs of DP are equal to

the NMBs of DP+Sr89 and uncertainty is high—EVPI reaches a local maximum of about £4.50 million.

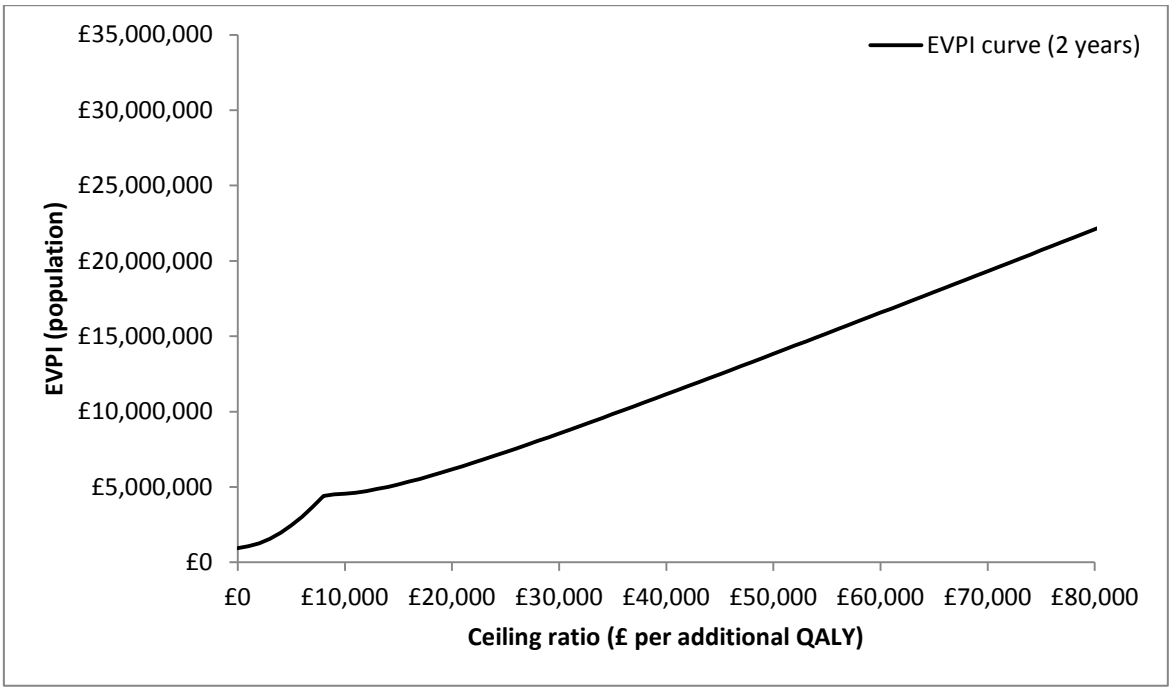

Figure 7.6: Population EVPI for HRPC

In summary, the EVPI at a ceiling ratio of £30,000 per QALY appears high, at £8.55 million. Further research that costs less than this amount would be potentially cost-effective and, given this, funding and conducting the proposed TRAPEZE phase III trial, which is expected to cost approximately £2.54 million, would be a potentially cost-effective investment.

7.3.2. Expected value of perfect parameter information analysis for HRPC

EVPPI analysis was carried out to establish the benefits of eliminating the uncertainty around specific parameters affecting the HRPC adoption decision. For each treatment, three subsets of uncertain parameters were investigated:

a. transition probabilities (disease progression rates);

b. utility scores for different health states, and

c. costs.

This gave a total of 12 different parameter subsets to be investigated. EVPPI calculations were carried out following the two-level (inner and outer loop) Monte Carlo simulations method detailed in Briggs *et al.*[51] and Brennan *et al*.[263], and calculations involved 1000 runs in each loop.

EVPPI for a single patient and the population of eligible patients at a ceiling ratio of £30,000 per QALY are presented in Figure 7.7 and Figure 7.8.

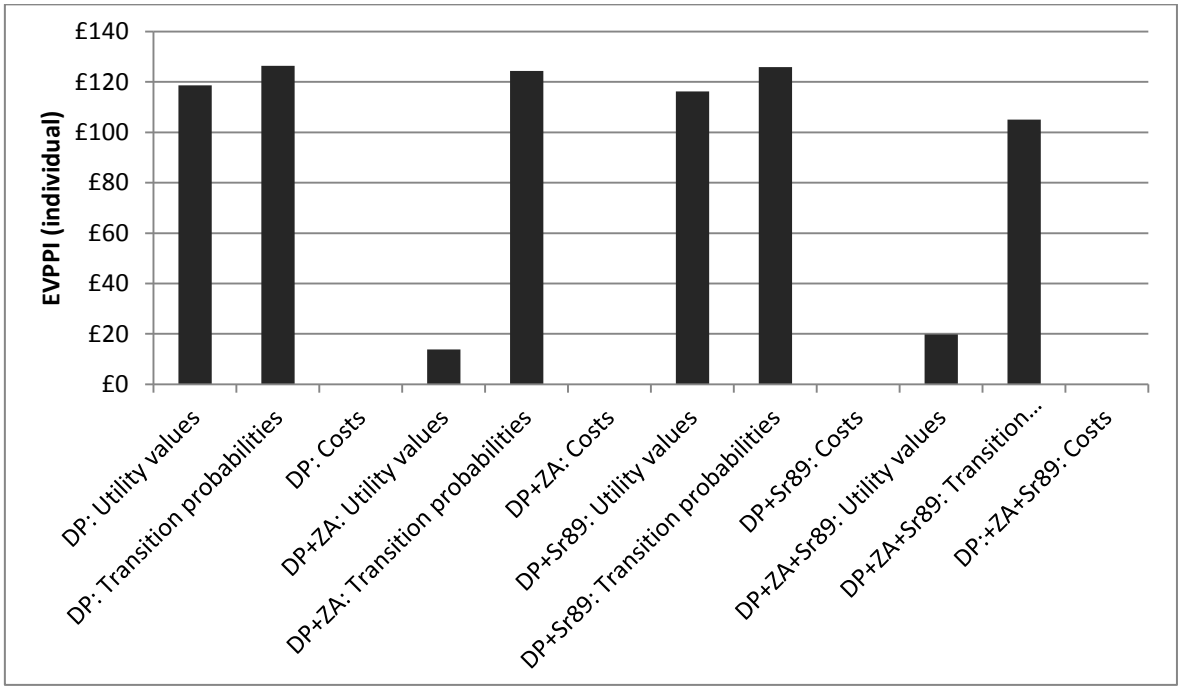

Figure 7.7: Individual EVPPI for HRPC at £30,000 per QALY

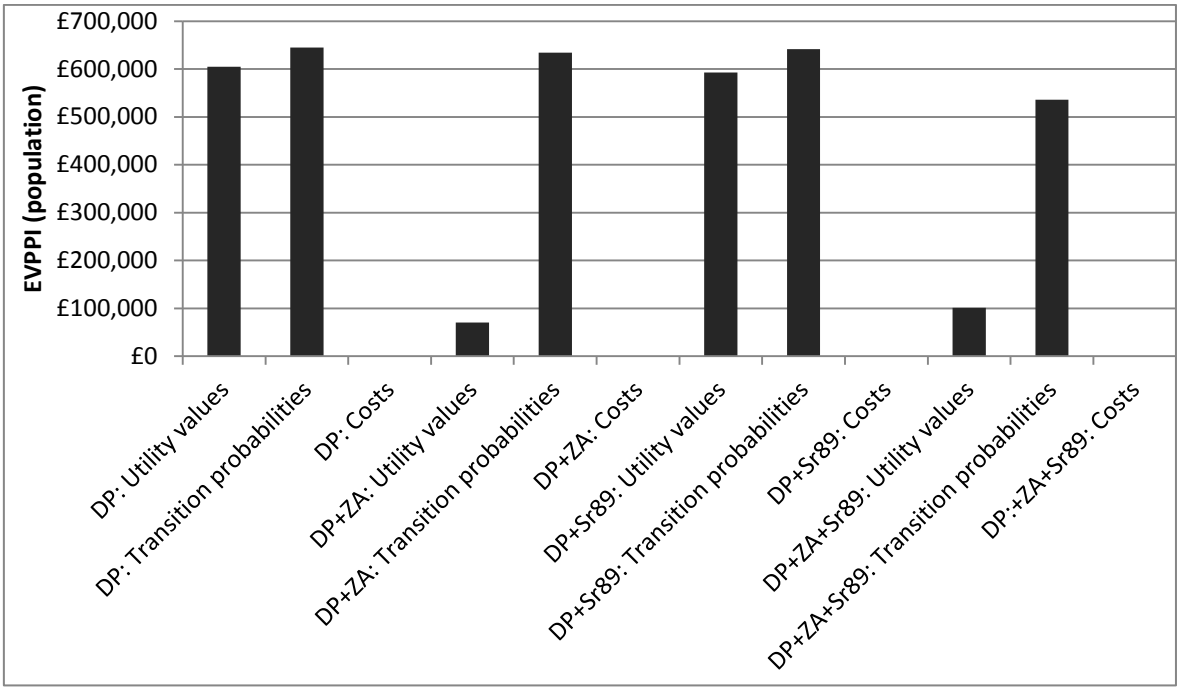

Figure 7.8: Population EVPPI for HRPC at £30,000 per QALY

The analysis indicated a relatively low value of EVPPI associated with transition probabilities; at £30,000 per QALY, population EVPPI ranged from £536,000 for DP+ZA+Sr89 to approximately £645,000 for DP. The results also suggest a relatively low population EVPPI for utility scores associated with DP and DP+Sr89 (£605,000 and £593,000 respectively). Cost parameters had a negligible impact on the decision uncertainty and, at £30,000 per QALY, population EVPPI associated with these parameters was zero for all treatments.

Population EVPPI curves for each of the 12 subsets of parameters are given in Figure 7.9. For low values of the ceiling ratio, where DP is more likely to be the optimal treatment, acquiring better estimates of specific subsets is unlikely to change the decision and further research on any of the specific parameters has low value. EVPPI for each subset of parameters increases as the ceiling ratio approaches the value where decision uncertainty is greatest

(i.e. the ICER) and peaks near the £8000 mark, above which the decision switches from DP to DP+Sr89.

In summary, the EVPPI analysis suggests that the maximum benefits from carrying out studies to eliminate the uncertainty around transition probabilities, costs or quality of life are relatively low and unlikely to exceed the cost of separate trials for each of these parameters. However, a trial which would investigate transition probabilities, resource use and utility scores simultaneously may be potentially worthwhile. In the light of the EVPI and EVPPI results, the TRAPEZE phase III trial which investigates transition probabilities, quality of life and resource use, appears potentially worthwhile.

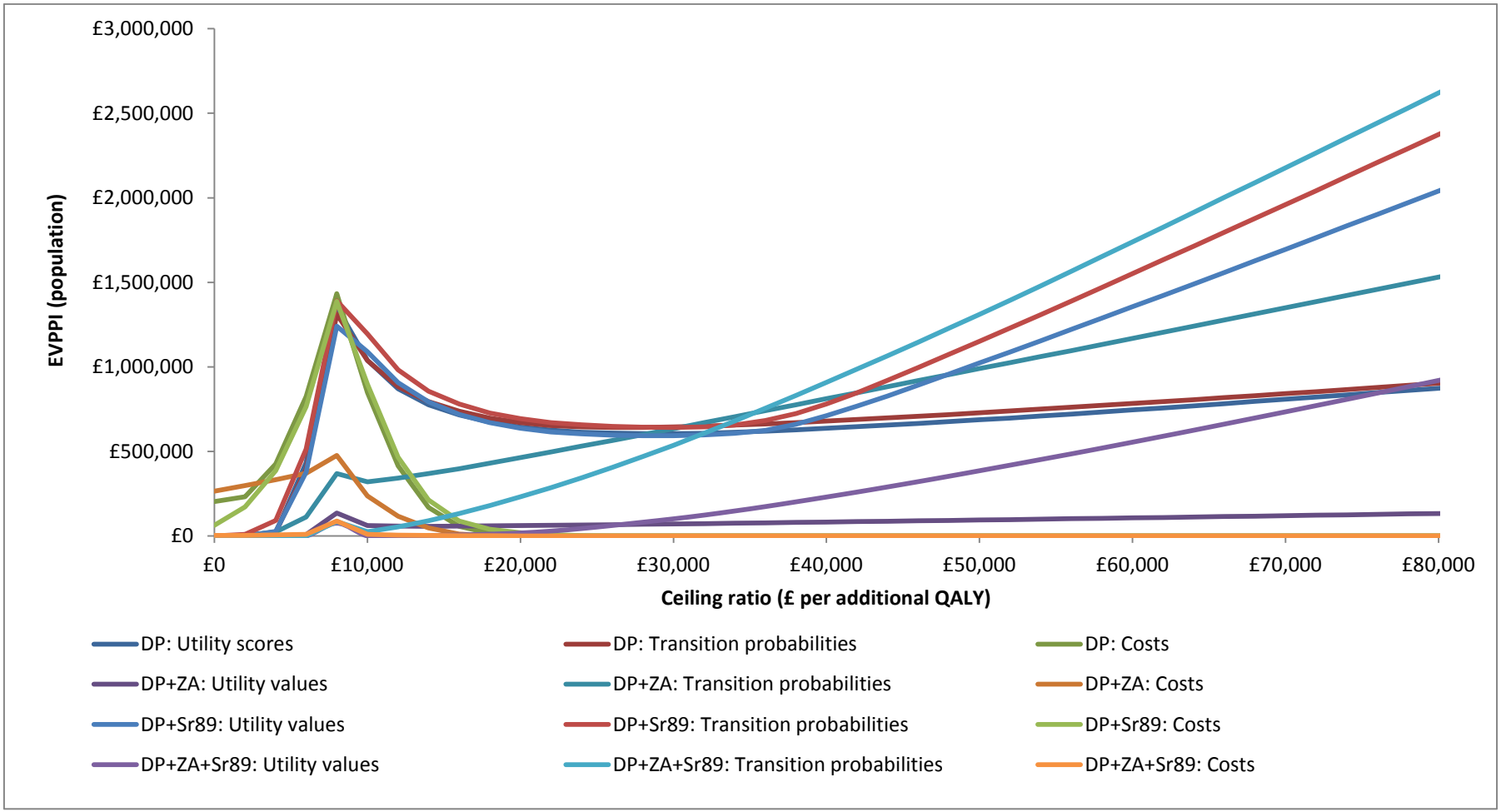

Figure 7.9: Population EVPPI at different ceiling ratios for HRPC

7.3.3. Value of information and implementation analysis for HRPC

Additional analysis was conducted to explore the value of implementation associated with HRPC under the assumptions that acquisition of information does, and does not, lead to a beneficial change in implementation.

7.3.3.A. Assuming acquisition of information has no impact on implementation

At a ceiling ratio of £30,000 per QALY and under current information, the treatment that appears cost-effective is DP+Sr89. Expert opinion suggested that, at the time information from further research is expected to become available, that is, in 2013, the majority of the HRPC patients treated with chemotherapy are anticipated to receive DP (85 percent) with the remaining treatments provided to a low proportion of patients (5 percent each for DP+ZA, DP+Sr89 and DP+ZA+Sr89). These values represent the prescription share under current implementation. Under current information and optimal implementation, DP+Sr89 would be provided to all eligible patients (100 percent implementation).

The results of this analysis are given in Table 7.16. As expected, the greatest gains—approximately £103.10 million in NMBs—are anticipated from a decision made under perfect information which will be implemented perfectly (state D). The benefits from a decision with current information but optimal implementation (state C) are somewhat lower, at about £94.56 million. Under the assumption that acquisition of further information does not affect implementation, the benefits expected to arise from a decision with perfect information and current implementation (state B) are the same as those from a decision

with current information and current implementation (state A), at approximately £78.40 million.

Table 7.16: Expected NMBs for different 'states of the world' regarding information and implementation (HRPC)

States of the world		Information	
		Current	Perfect
Implementation	Current	£78,396,604 (state A)	£78,396,604 (state B)
	Optimal	£94,555,968 (state C)	£103,104,695 (state D)

The estimated values of different measures of implementation are given in Table 7.17. EVPI, that is, the difference between a decision made under perfect information and optimal implementation and current information and optimal implementation, was estimated at £8.55 million. This estimate gives the upper bound of the value of conducting further research. The values of $EVIMP_{perfect\ information}$ and $EVIMP_{current\ information}$ were calculated at £24.71 million and £16.16 million, respectively. These values provide a measure of the upper ceiling of resources to be devoted to implementation strategies, under perfect and current information, respectively. The EVP, a measure of the benefits due to moving from current information and current implementation to perfect information and perfect implementation was found to be at £24.71 million. This is an estimate of the maximum amount of resources to be devoted to research and implementation programmes.

As expected, under the assumption of no interaction between information and implementation, the realisable EVPI (i.e. the difference between the expected NMBs in a situation with perfect information and current implementation (state B) and a situation with

current information and current implementation (state A)) is zero, as no change in implementation implies no additional benefits.

Last, the 'expected value of information-implementation trade-off', that is, the comparison between perfect information and current implementation (state B), and current information and optimal implementation (state C), gave a negative value of -£16.16 million. This suggests that the NMBs from a decision with current information and optimal implementation exceed those from a decision with perfect information implemented at current levels. On this basis, if the cost of research is similar to the cost of implementation strategies, investing in programmes to improve adherence to cost-effective treatments appears to be more beneficial than funding and conducting research to obtain better information.

Table 7.17: Expected NMBs for different measures of information and implementation (HRPC)

Measure	NMBs (at £30,000 per QALY)
EVPI (state D – state C)	£8,548,728
rEVPI (state B – state A)	£0
EVPIM$_{\text{perfect information}}$ (state D - state B)	£24,708,092
EVPIM$_{\text{current information}}$ (state C – state A)	£16,159,364
EVP (state D – state A)	£24,708,092
EVIIT (state B – state C)	-£16,159,364

7.3.3.B. Assuming acquisition of information affects implementation

The assumption that the level of available information has no bearing on implementation rates was relaxed using hypothesised estimates of treatments' prescription shares. These were obtained from discussion with experts, based on a series of guesses about the direction

and strength of change in practice in the light of different eventualities regarding research results, and are given in Table 7.18. In general, should a treatment appear cost-effective, its uptake would be expected to increase, although the magnitude of this increase would vary across treatments and would depend on obstacles to implementation. For example, if treatments which require radioisotope fractions (i.e. DP+Sr89 or DP+ZA+Sr89) were found cost-effective, the increase in their uptake would be expected to be lower than that of treatments not involving radioisotopes (i.e. DP, DP+ZA).

Table 7.18: Prescription shares in the light of different eventualities about treatments' cost-effectiveness (HRPC)

Implementation	Eventuality	Prescription share of DP	Prescription share of DP+ZA	Prescription share of DP+Sr89	Prescription share of DP+ZA+Sr89
Current implementation		85.0%	5.0%	5.0%	5.0%
Improved implementation	**DP cost-effective**	90.0%	3.3%	3.3%	3.3%
	DP+ZA cost-effective	40.0%	50.0%	5.0%	5.0%
	DP+Sr89 cost-effective	50.0%	5.0%	40.0%	5.0%
	DP+ZA+Sr89 cost-effective	50.0%	5.0%	5.0%	40.0%

The assumption that perfect information induces a beneficial change in implementation has an impact on three measures: the rEVPI, $EVPIM_{perfect\ information}$ and EVIIT. The new results for these measures are shown in Table 7.19. The NMBs of the revised rEVPI were estimated at £7.43 million; this value reflects the maximum expected NMBs from undertaking both research and implementation strategies. The revised $EVPIM_{perfect\ information}$ was calculated at £17.27 million; this value shows the difference between the NMBs expected to arise from a state where perfect information is followed by optimal implementation and a state where perfect information is followed by improved implementation. Last, the revised EVIIT was

negative, suggesting that a decision in a state with perfect information and improved implementation would result in lower NMBs than the same decision taken under current information and perfect implementation. The revised EVIIT results are in the same direction as those under the assumption of no interactions between information and implementation, suggesting that, despite the hypothesised improvements in implementation following research, resources would still be more prudently invested in implementation strategies.

Table 7.19: Revised expected NMBs for different 'states of the world' regarding information and implementation (HRPC)

Measure	NMBs (at £30,000 per QALY)
rEVPI (state B- state A)	£7,433,518
EVPIM$_{\text{perfect information}}$ (state D- state B)	£17,274,574
EVIIT (state B – state C)	-£8,725,846

7.3.4. Expected value of sample information and expected net benefits of sampling for HRPC

Expected value of sample information (EVSI) analysis was carried out to explore the applicability of the method, as well as to develop a basis for a subsequent discussion of its practicality, usefulness and limitations. The application was carried out using the HRPC Markov model described in Chapter 5. Non-parametric applications of EVSI on the basis of the output of Markov models are rare; no other such EVSI analyses carried out in Microsoft Excel®—a spreadsheet application commonly used for model building—are known to the author and the supervisory team of this thesis. The analysis was assisted by the fact that parameters of interest (transition probabilities) were represented by distributions with

specific properties (conjugate distributions), as well as by the use of an 'individual sampling model' exercise.

EVSI seeks to establish the additional benefits expected to arise from making an adoption decision under 'improved' (posterior) information about a parameter affecting the decision—for instance, effectiveness of treatments, expressed as transition probabilities to a worse health state—as opposed to making the same decision under existing (prior) information. In such analyses, 'improved' information is a combination of existing evidence (e.g. currently known transition probabilities) and possible evidence from further research, that is, 'sample' information from a proposed trial. As the trial has not taken place, sample information has not been observed and needs to be predicted.

Combining prior and sample information is feasible when the distributions representing the prior and new information are conjugate, that is, they are of the same family and can be brought together using simple analytic formulae[51;107]. Such were the beta and Dirichlet distributions which were used to characterise transition probabilities in the HRPC model. Analyses based on non-conjugate distributions are in principle possible, but they are considerably more complex[51;190].

7.3.4.A. Process for obtaining posterior distributions of effectiveness

The first step in this analysis involved drawing a set of values from the prior distribution (Pr_i) of the parameters of interest, here, transition probabilities. The prior distribution of transition probabilities is represented by a Dirichlet distribution, with parameters of this distribution showing counts of 21-day cycles that participants spent in specific health states:

$$Draw\ a\ sample\ from\ prior\ distribution\ Pr_i \sim Dirichlet(\ ,\ ,\ ,\ )$$

The next step involved predicting the possible 'sample' information, in the form of possible results of the proposed TRAPEZE phase III trial. This was done by carrying out an 'individual patient sampling' exercise. In contrast to cohort models which follow a group of patients (such as the NSCLC and HRPC models reported in Chapters 4 and 5), individual sampling models trace the progression of hypothetical patients through health states one at a time. The aim of the individual patient sampling exercise in the particular context is to predict the possible results (in terms of accumulated transitions and, thus, transition probabilities) of a hypothetical trial of a given sample size.

The process involved in individual patient sampling is outlined below. A patient starts in a specific health state (e.g. state A). The probability of the patient transitioning to another state B is based on existing (prior) information (e.g. probability of moving from state A to state B is 0.05). A number between 0 and 1 is drawn at random; if this number is smaller or equal to 0.05, the patient moves to state B; if the number is larger than 0.05, the patient stays at the current state. This process is repeated for a sufficient period of time (i.e. for a large number of discrete model cycles) and the number of cycles the patient spent at each state is counted and can be used to give the probability of the patient transitioning across states. This process is run for each of the patients in the simulated trial. In the context of this study, the proposed TRAPEZE phase III trial was expected to recruit 300 participants per arm and thus individual patient sampling was run for 300 hypothetical patients in each arm. The obtained counts of cycles spent in specific health states can be expressed as a probability distribution, in this case a Dirichlet distribution.

$$Possible\ result\ of\ trial\ D_i\ (given\ the\ specific\ sample\ from\ prior\ Pr_i)$$

$$\sim Dirichlet\ (\ ,\ ,\ ,\ )$$

Next, the set of transition probabilities from the prior distribution were combined with the possible trial results to give the transition probabilities for the posterior distribution $Post_i$, as:

$$Posterior\ distribution\ Post_i \sim Dirichlet\ (\ +\ ,\ +\ ,\ +\ ,\ +\ )$$

This process was repeated 1000 times and gave 1000 sets of values for the posterior transition probabilities ($Post_1, Post_2, \dots, Post_{1000}$) for each treatment j. Each of these sets was subsequently entered in the HRPC model one at a time, where Monte Carlo simulations were carried out to give 1000 estimates of each treatment's NMBs, by drawing values from each set of posterior values.

7.3.4.B. EVSI and ENBS results

For each treatment (i.e. DP, DP+ZA, DP+Sr89 and DP+ZA+Sr89) and for the parameter of interest φ (transition probabilities), obtained NMBs estimates were averaged across their distribution to give an estimate of the expected NMBs ($E_{\varphi|Post_i}NMB(j,\varphi)$) given the specific posterior distribution. The process was repeated 1000 times, to give 1000 estimates of $E_{\varphi|Post_i}NMB(j,\varphi)$ for each treatment (columns 2 to 5 in Table 7.20), of which the maximum expected NMBs ($max_j\ E_{\varphi|Post_i}NMB(j,\varphi)$) were selected (column 6 in Table 7.20). As it is not known which of the sample results D_i will be observed (i.e. which posterior distribution

represents the true underlying distribution), the $max\, E_{\varphi|Post_i} NMB(j,\varphi)$ need to be averaged over their posterior distribution, to give the $E_{Post_i}\, max_j\, E_{\varphi|Post_i} NMB(j,\varphi)$ (bottom right cell in Table 7.20).

Table 7.20: Illustrative example of EVSI calculations for a sample size of 300 patients per arm

	DP	DP+ZA	DP+Sr89	DP+ZA+Sr89	$\boldsymbol{max_j E_{\varphi|Post_i} NMB\ (j,\varphi)}$
$E_{\varphi|Post_1} NMB(j,\varphi)$	£18,513	£15,973	£16,316	£13,720	£18,513
$E_{\varphi|Post_2} NMB(j,\varphi)$	£15,995	£16,408	£21,307	£16,811	£21,307
$E_{\varphi|Post_3} NMB(j,\varphi)$	£15,981	£16,704	£18,120	£12,793	£18,120
…	…	…	…	…	…
$\boldsymbol{E_{\varphi|Post_{999}} NMB(j,\varphi)}$	£14,589	£12,074	£13,911	£12,922	£14,589
$\boldsymbol{E_{\varphi|Post_{1000}} NMB(j,\varphi)}$	£15,312	£13,471	£24,018	£13,399	£24,018
$\boldsymbol{E_{Post_i}\, max_j\, E_{\varphi|Post_i} NMB(j,\varphi)}$					**£19,145**

This value represents an estimate of the expected NMBs from making a decision under sample information and, when compared to the expected NMBs of making a decision under current information ($max_j E_\varphi NMB(j,\varphi)$), it gives the EVSI$_n$ for the particular trial of size n, investigating the parameters φ.

In this analysis, for a new trial of 300 participants per treatment arm and given a ceiling ratio of £30,000 per QALY, the expected maximum NMBs with sample information were estimated at £19,145. Subtracting the estimated expected NMBs of making a decision under current information (i.e. £18,540 at £30,000 per QALY) from this value gave the EVSI for a

trial of 300 participants per arm. This was found to be £605 and £3.09 million for the individual and the population, respectively.

Similarly to EVPI and EVPPI, EVSI results vary with the employed ceiling ratios. The population EVSI (dotted line) for a range of ceiling ratios can be seen in Figure 7.10, along with the population EVPI (solid line). As expected, at each ceiling ratio the benefits from a decision with imperfect, sample information (EVSI) are lower than those of a decision made under perfect information (EVPI) (solid line). The EVSI curve follows a pattern analogous to the EVPI curve: as the ceiling ratio increases, EVSI increases and reaches a local maximum when uncertainty around a decision is greatest (i.e. near the ICER of the most cost-effective treatment).

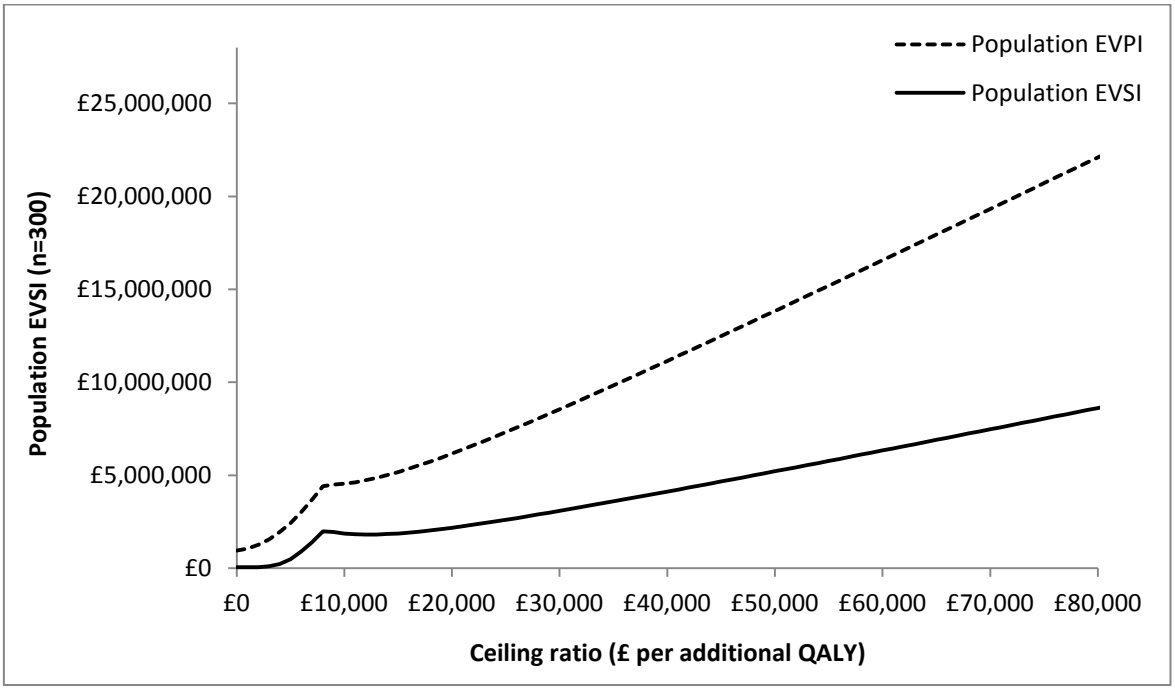

Figure 7.10: Population EVSI (for sample size of 300 patients per arm) and EVPI at different ceiling ratio values

The obtained EVSI estimates can be compared with the cost of the trial to give the expected net benefits of sampling (ENBS) for the particular trial. In the context of the HRPC case study, the cost of the TRAPEZE phase III trial consisted of a fixed component, which includes researchers' salaries and expenditure for trial co-ordination (£627,156), as well as a variable part, which depends on the size of the trial. For a trial of 300 patients per treatment arm, this includes an excess treatment cost (i.e. additional cost due to patients in experimental arms receiving care different than that routinely provided outside the study) of £1,457,560, as well as service support costs related to patient recruitment and additional clinic time of £452,400. On the basis of above, the marginal cost of recruiting an additional patient in the trial was estimated at £1590. The total cost of the trial of 300 patients per treatment arm, including fixed and variable costs, was stated as £2.54 million. Given the above, the ENBS for a trial of 300 participants per treatment arm was estimated at:

$$ENBS_{n=300} = EVSI_{n=300} - trial\ cost_{n=300} = £549{,}200$$

According to this, the proposed trial of 300 patients aiming to obtain better information on disease progression would result in ENBS of about £550,000.

Additional analyses were carried out for different possible sample sizes (n=100, 200 and 400 participants per arm). Increases in the number of trial participants led to increases in the 'value' of the trial (EVSI) (second and third columns in Table 7.21) while, at the same time, recruiting more participants raises the cost of the trial due to increase in variable costs (fourth column in Table 7.21).

For the particular sample sizes investigated here (n=100, 200, 300 and 400 patients per arm), the expected benefits exceeded the increase in costs (i.e. there are positive marginal benefits), thus, a study involving any of these sample sizes would be beneficial. This will be the case up to the point where the additional cost of recruiting an additional participant in the study will exceed the additional benefits from involving this extra person. Under the assumption that there is no additional benefits from allocating participants to one arm over another, the 'optimal' sample size can be found at the point when the difference between trial cost and NMBs is the greatest possible (that is, when ENBS is maximum). Assuming equal allocation between trial arms is optimal and a ceiling ratio of £30,000 per QALY, the resulting ENBS suggest that the preferable sample size for TRAPEZE among the ones considered here would be 200 patients in each arm.

It must be noted that significantly greater numbers of simulations may be needed to minimise the influence of sampling error and derive more accurate results. The influence of sampling error can be seen in the fluctuations in the obtained EVSI and ENBS results. For an increase in sample size from the 'optimal' size of 200 to 300 patients per arm, the EVSI and ENBS increase at a diminishing rate, whereas increasing the number of participants from 300 to 400 per arm shows the EVSI and ENBS to be rising at an increasing rate.

Table 7.21: Individual EVSI for studies of different sample sizes

Trial sample size	Individual EVSI (£30,000 per QALY)	Population EVSI (£30,000 per QALY)	Total cost of trial (£30,000 per QALY)	ENBS (£30,000 per QALY)
100 participants per arm	£411	£2,095,813	£1,263,809	£832,004
200 participants per arm	£569	£2,900,836	£1,900,463	£1,000,373
300 participants per arm	£605	£3,086,341	£2,537,116	£549,225
400 participants per arm	£739	£3,769,606	£3,173,769	£595,836

In summary, the EVSI analysis suggested that a trial to provide better information on disease progression, such as the TRAPEZE phase III trial, would result in benefits in excess of its cost. On this basis, funding and carrying out such a trial appears to be 'cost-effective'.

7.4. Discussion

The chapter reports the application of VoI methods to inform a decision on funding primary research in the areas of NSCLC and HRPC. At a ceiling ratio of £30,000 per QALY and assuming a five-year time horizon, results for the NSCLC case study showed a high EVPI value in excess of £13 million and relatively high EVPPI values for clinical progression. In light of these results, the maximum expected gains from research exceed the cost of the proposed BTOG-2 trial, funding and conducting the trial is potentially worthwhile. In addition, EVSI analysis was carried out to compare the benefits expected from a trial of a particular sample size to the cost of the trial. For the BTOG2 phase III trial, which was expected to cost £336,700 and involve 450 participants per treatment arm, the EVSI exceeded the cost of the trial, resulting in positive ENBS of about £9 million. This suggested that the particular study

satisfies both the necessary (i.e. EVPI in excess of trial cost) and the sufficient (EVSI in excess of trial cost) conditions, and it is worth funding and conducting.

Similar results were observed in the HRPC case study: at £30,000 per QALY and a two-year time horizon, EVPI was found to be about £8.5 million. The value far exceeds the cost of the proposed TRAPEZE phase III trial (£2.54 million), suggesting that study would be potentially beneficial and is worth considering further. EVPPI analysis suggested that the parameter where research may be most beneficial is clinical disease progression. Although investigating each parameter in separate trials is unlikely to be cost-effective, carrying out a trial which would look into disease progression, cost and quality of life for all the chemotherapy options of interest—such as the TRAPEZE phase III trial—would be potentially beneficial. Last, EVSI analysis for the TRAPEZE phase III trial (cost of £2.54 million, 300 participants per treatment arm) the EVSI outweighed the cost of the trial, resulting in positive ENBS of about £550,000. In this case, too, conducting the TRAPEZE phase III study satisfies both the necessary and the sufficient conditions, and it is worth prioritising.

The analysis followed well-established methods for calculating EVPI[51;180;190] and EVPPI[51;263-265]results and EVSI[51;190;272]. Additional analyses were conducted to explore measures of the value of implementation using the framework proposed by Fenwick and colleagues[187]. This framework was extended to illustrate the calculations of different measures of the value of implementation under the assumption that acquisition of perfect information is expected to have an impact on implementation.

EVSI analysis was also carried out to estimate the benefits from decision-making in the light of information from the proposed BTOG2 and TRAPEZE phase III trials. These analyses were novel in that they simulated the evidence that the hypothesised trials may give by using individual patient sampling simulations methods. Observations made throughout this application regarding the practicality of undertaking VoI analysis and its potential to assist research funding decisions are discussed in Chapters 8 and 9.

7.5. Chapter overview

This chapter described an application of VoI methods to case studies representing decisions to fund primary evaluative research (BTOG-2 and TRAPEZE phase III trials) in NSCLC and HRPC. According to the obtained EVPI and EVPPI results, the maximum expected gains from evaluative research in NSCLC and HRPC are high and exceed the cost of the proposed trials in these areas. Given this, both the BTOG-2 and the TRAPEZE trials are potentially worth conducting. EVSI analysis carried out for NSCLC HRPC showed that the benefits from trials such as BTOG2 and TRAPEZE outweigh their respective cost and thus funding and carrying out these trials would be beneficial. The applications allowed observations on the practical and methodological challenges of 'value of information' methods; these are discussed and summarised in the following chapters.

PART III. Assessment, discussion and conclusions

The last part of this thesis, Part III, aims to summarise and discuss the findings of the project. Chapter 8 identifies strengths, limitations, potentials and challenges associated with 'payback of research' and 'value of information' by drawing on observations made throughout the practical application of the frameworks to case studies.

The concluding chapter of this thesis, Chapter 9, summarises the project aims and methods, discusses and interprets the findings, forms conclusions and makes recommendations for further research.

CHAPTER 8. Assessment of 'payback of research' and 'value of information'

This chapter aims to assess the 'payback' and 'value of information' (VoI) frameworks by drawing on observations made throughout this study. Strengths, limitations and challenges associated with the approaches are looked at from the viewpoint of potential users of the methods. Important points arising from this assessment form the basis for the discussion and conclusions presented in Chapter 9.

8.1. Existing assessments of 'payback of research' and 'value of information'

Two studies which looked at the strengths and limitations of 'payback' and 'value of information' in a comparative way were identified in the literature. In the first study, Chilcott *et al*.[55] undertook a review to assess the role of decision modelling in designing and prioritising clinical trials. The authors discussed the potential of VoI and 'payback'-based models published up to 1999 to help with research funding decisions, by drawing on evidence from the existing literature, without applying the methods to case studies. Chilcott *et al*.[55] found 'payback' to be an intuitive and potentially feasible approach and suggested further research into incorporating sensitivity analysis in aspects where assumptions are typically needed (e.g. extent of change in clinical practice, specification of possible outcomes and likelihood of these outcomes to occur). With respect to VoI, the authors[55] pointed out that the framework has a sound theoretical underpinning, while, at the same time, it may be useful in indicating the type of research that may be needed and determining optimal study

characteristics. The authors recognised the complexities of carrying out decision modelling and VoI analyses, particularly as compared to more simple, deliberative approaches.

In the second study, Fleurence[116] applied 'value of information' (expected value of perfect information (EVPI)) and 'payback' (PATHS model) to case studies of proposed clinical trials in the areas of osteoporosis and pressure ulcers. The author pointed out that, although the use of 'payback' and VoI for research priority-setting appears beneficial, both the frameworks have limitations. In the case of EVPI, it was highlighted that the method fails to translate the expected benefits from research to benefits in clinical practice, while, on the other hand, 'payback' was criticised on the basis of the assumption that a clinical trial is valuable only if it can lead to changes in clinical practice, which, it is argued, should not be the primary aim of research[116]. Both Chilcott *et al*.[55] and Fleurence[116] agreed that the use of analytic approaches may bring about improvements in decision-making as compared to currently used deliberative methods.

Although these studies offer a useful insight into the strengths and limitations of 'payback' and VoI, their scope and conclusions appear limited. The study by Chilcott *et al.*[55] was intended as a review and did not involve a 'hands-on' application of VoI or recent 'payback' models (e.g. PATHS[67]) to a case study; therefore conclusions about the practicality and methodological challenges associated with the frameworks were not based on actual observations. On the other hand, the study by Fleurence[116] did involve a practical application of PATHS and VoI to case studies, but VoI analysis was limited to EVPI analysis only.

Work undertaken in the present study aimed to address limitations and extent the methods used in the existing literature, with a view to providing a more comprehensive assessment of 'payback' and VoI. To this end, VoI analysis reported in Chapter 6 involved not only EVPI, but also expected value of partial information (EVPPI) and expected value of sample information (EVSI) analyses. EVPPI and EVSI are central concepts and, without looking into these concepts, a discussion on the potentials and challenges associated with VoI analysis is bound to be incomplete. In addition, the analysis accounted for the value of implementation, a recently introduced conceptual framework aimed to strengthen VoI. The concept is important as it aspires to address the widespread criticism that VoI results do not present the real, tangible benefits that would accrue to the patients, as they are based on the unlike assumption that adoption decisions will always be implemented perfectly. Work reported in Chapter 6 incorporated this conceptual framework into the VoI analysis, and attempted to extend it by acknowledging that additional information is likely to result in improved, rather than perfect, implementation.

With regards to the application of 'payback', additional empirical work involved sensitivity analysis to identify and pinpoint the main parameters affecting 'payback' results, illustrated the addition of probabilistic sensitivity analysis and looked into the methodology of conducting 'payback' analysis for research involving multiple treatment comparisons.

This additional analysis gave the opportunity for a more comprehensive assessment of the frameworks. In particular, undertaking this additional analysis helped to identify methodological weaknesses associated with the frameworks, gave a more complete idea of the type and nature of assumptions required by the methods, and allowed an insight into

the feasibility and difficulties of conducting especially complex and time-intensive concepts such as EVPPI and EVSI. Overall, the additional analyses offered a more complete picture of 'payback' and VoI, which allowed a closer insight into the potentials, strengths and weaknesses of these analytic frameworks.

8.2. Interpretation and validity of results

Both 'payback' and VoI generated estimates of the expected benefits from research in NSCLC and HRPC. In general, 'payback' and VoI results appeared to agree in direction, suggesting that both the BTOG-2 and TRAPEZE phase III trials are potentially cost-effective investments. In particular, VoI results showed that, given certain assumptions about the eligible population and the time horizon over which the produced information is expected to be relevant, the additional benefits expected from research in NSCLC and HRPC exceeded the cost of the proposed trials in these areas, suggesting that funding and carrying out the trials would be beneficial. Similarly, the PATHS analysis revealed additional NMBs from undertaking the trials and changing clinical practice according to their results.

Despite this broad agreement, there are important differences in the way in which results are derived and interpreted. 'Payback' seeks to predict the NMBs that would be realised should research take place and generate results that would trigger a hypothetical change in clinical practice. Three factors have a major effect on the magnitude of 'payback' results: a) the extent to which the cost-effectiveness of the treatment of interest as revealed by research will differ to the currently perceived cost-effectiveness, b) the magnitude of the

beneficial change in clinical practice in the light of new evidence from a trial and c) the cost of the trial.

In the case of VoI, the expected value of perfect information and the expected value of perfect parameter information (EVPPI) show the maximum NMBs that would be expected from making a decision under perfect as opposed to current information about all or a specific parameter, while the expected value of sample information (EVSI) gives the additional expected benefits from a decision made in the light of improved, as opposed to current, information. In this context, results are driven by the extent of existing uncertainty (i.e. the probability that the treatment which appears inferior under current information is actually more cost-effective than the currently preferred option) and the expected loss of benefits if the latter turns out to be the case. Given this, further research appears more desirable when a) uncertainty is high and b) the expected (possible) loss due to uncertainty is expected to be substantial, and c) the cost of the trial is low.

A first question arising relates to whether the generated results are valid. For assessing the predictive validity of VoI and 'payback', generated results need to be compared against actual, post-research observations[55;67;273]. In order to establish that the frameworks produce valid results, the NMBs predicted in the pre-research analysis must agree with the actual results which will be realised after research has taken place.

Attempts for result validation are hindered by difficulties. First, comparisons between predicted and actual results would only be possible for research programmes which are subsequently funded and carried out, as studies that are not carried out cannot produce

actual, post-research evidence[55;67]. Second, a long time horizon would be needed before the actual post-research benefits were observed and care would be required to isolate the benefits which can be attributed to the specific research programme taking place, so that these could be compared against the prospective estimates.

Validation of 'payback' results has been attempted by Townsend and colleagues[67]. In their study, the authors compared prospective PATHS results obtained from case studies against post-research results, the latter being estimated by combining actual research results with predictions about change in clinical practice. The authors[67] reported that the value of the predicted (*ex ante*) results agreed with those of the presumed actual (*ex post*) results in two of the three case studies where such a comparison was feasible. One of the difficulties with this approach is that pre-research 'payback' methods give a number of possible results (usually three: 'favourable', 'inconclusive' and 'unfavourable') while the actual, post-research result is unique. Given this, the comparison is only meaningful when post-research results are put against the equivalent possible results, which, however, cannot be known at the time when a proposal is considered for funding.

Validation of EVPI and EVPPI poses similar difficulties. First, comparisons between pre and post-research results will only be possible for studies which have taken place. Second, EVPI and EVPPI results will show the maximum potential reduction in opportunity loss (i.e. the maximum expected benefits to be gained by eliminating uncertainty), whereas further research will only result in partial reduction in opportunity loss and thus less-than-maximum benefits. A possible way of validation may involve carrying out pre-research EVPI and EVSI analyses (for the specified number of patients in a proposed study's research protocol) and,

after research has taken place, calculating the post-research EVPI. For EVPI to be able to capture the reduction in uncertainty—and the subsequent expected benefits due to this reduction—the difference between post-research and pre-research EVPI should be comparable to the pre-research EVSI.

Overall, practical and methodological problems make 'payback' and VoI results validation difficult, although, as it is argued in the next chapter, the extent to which formal validation is needed before the results are used in practice is questionable.

8.3. Theoretical and methodological robustness

An important consideration in assessing 'payback' and VoI relates to the frameworks' theoretical and methodological soundness. Points related to these aspects are raised below.

8.3.1. Observations related to the frameworks' theoretical soundness

As it was explained earlier in this work, the 'payback of research' framework is based on the notion that research is valuable because it provides information that stimulates a beneficial change in practice. Therefore, according to this framework, the desirability of a research programme can be inferred by the additional benefits that the programme is expected to generate through informing a change in practice.

This is an intuitive idea, which seeks to account for the 'real', tangible benefits that may accrue to the population due to research. However, a direct implication of this notion is that the framework may attach greater weights to research in areas where there is a great scope for gains from a beneficial change in clinical practice, over areas where there is much

uncertainty about the appropriate use of treatments, but a lesser scope for change in practice[80]. This was confirmed in sensitivity analyses carried out in the present study, where assumptions about greater potential for change in clinical practice (i.e. greater change in prescription rates) led to greater expected NMBs from research.

In addition, an implicit assumption in this approach appears to be that new information produced by research is the sole reason for a beneficial change in practice. Research may indeed have a major effect in triggering change in clinical practice and may be responsible for a large share of the benefits due to greater use of cost-effective treatments; nonetheless other factors may be also contributing to a change in clinical practice, such as increasing familiarity with a new technology or active promotion by a treatment's manufacturers. As a result, attributing all the benefits from a change to research may overestimate the value of a proposed study.

As far as VoI is concerned, the approach has firm foundations on well-established principles of statistical decision theory. The framework is based on a 'decision-theoretic' stance, according to which research, such as a clinical trial, should be seen as a source of evidence for decision-making[115;189]. In particular, this viewpoint stipulates that a clinical trial is worth conducting as long as it is expected to add to the existing evidence base and provide input for decision-making, no matter whether the generated results will reach statistical significance, or whether the trial was powered to do so in the first place. According to this viewpoint, a trial of a few participants would be worth conducting if it produces evidence that reduces uncertainty and can be used for decision-making[115].

Although this viewpoint gains popularity[191], it is at odds with the established view that clinical trials should be carried out to test hypotheses about treatments and should be designed and powered for this purpose. As an implication, before a funding organisation decides to take VoI results into account when making funding decisions (or, more importantly, when specifying the design requirements of clinical trials), it is important that it subscribes to this 'decision-theoretic' stance.

Last, common to both approaches is the notion that research resources should be allocated with efficiency in mind, so that the greater the likely benefits associated with a research programme, the greater the desirability for conducting the programme. This notion is in agreement with currently used decision rules based on the utilitarian view that resources should be allocated to achieve 'the greatest benefits for the greatest numbers'[23;144;260]. However, just as the public may have a preference for health care resources to be used in pursuing objectives other than economic efficiency (e.g. equity[274-276]), society may also prefer to give priority to research programmes that may not necessarily result in the greatest number of benefits. It must be noted that objectives beyond maximisation of benefits are not reflected explicitly on the results of 'payback' and VoI analyses, and, in situations where they are perceived relevant, these objectives will need to be taken into account as additional considerations.

8.3.2. Observations related to the frameworks' methodological soundness

The core methods of conducting 'payback' and VoI analysis are, in general, well-established. However, a number of points and challenges related to the frameworks' methodologies were identified.

First, a central, although implicit, assumption in 'payback' models is that research results are expected to reveal the true values of parameters, on the basis of which one can calculate the stream of costs and benefits expected to accrue to the health care system. This assumption appears to overlook the fact that information from research (e.g. clinical trials) comes in the form of uncertain estimates from samples, which may have been observed by chance, or, if the study is flawed, may be incorrect or biased. As a result, the benefits that are predicted to accrue on the basis of specified trial results are also uncertain and may or may not reflect the true benefits that would be expected in the population.

A second issue pertains to the specification of possible trial results and change in practice in the light of different results. Possible trial results have a substantial impact on the stream of costs and benefits predicted to accrue from a treatment. Although these are selected according to specified outcomes (e.g. in order for a 'favourable' outcome to transpire, the BTOG-2 trial should show a probability of disease progression of 0.64 at one-year follow up), the choice of values is to some extent arbitrary. This is because other values for transition probabilities in the range of the chosen value may also result in a 'favourable' outcome.

Similarly, the pace and magnitude of change in clinical practice is difficult to predict, as it depends on a number of factors, including the magnitude and strength of the results *per se*,

the effectiveness of existing dissemination mechanisms, the availability of the infrastructure needed to make the change in practice possible, as well as the degree to which change towards cost-effective treatments (or restrictions in non-cost-effective treatments) is compulsory. Obviously, weighting up all these factors is complex and predictions about the possible change in practice are inherently prone to error. In the present study, estimates of the possible change in clinical practice were obtained through discussion with experts, who, nonetheless, highlighted that their estimates were mere guesses.

Methodological challenges also arise when applying 'payback' to case studies involving multiple comparisons. As mentioned in Chapter 6, the correct approach for dealing with such applications is unclear, while, at the same time, this task requires stronger assumptions when specifying possible outcomes; for instance, under the 'inconclusive' scenario, all four treatments are assumed to be of similar cost-effectiveness, which is an unlikely situation. In comparisons between multiple treatments, different possible outcomes need to be specified in a way that covers all the possible eventualities (i.e. each treatment to be cost-effective, as well as all treatments to be of similar cost-effectiveness). As a result, the number of weighted and non-weighted 'payback' results increases (in the analysis of the four-treatment TRAPEZE, there were five different combinations (see Chapter 6)) which may pose difficulties for selecting the combination that is more likely to transpire.

With regards to VoI, an important point relates to the conflict that appears to exist between research that is needed in order to improve scientific knowledge, and research that is useful for reducing decision uncertainty. This is clear in the case of preference-based quality of life (utility) scores for health states in NSCLC. Searches in the literature revealed a lack of

evidence on these parameters and, given this, it would be reasonable to suggest that further research should be carried out to look into patients' quality of life. However, the expected value of perfect parameter information for such scores was low, even when the uncertainty around this parameter was set at a high level (see Chapter 7) and, thus, utility values were not suggested as an area where further research should focus on.

To a large extent, this is due to the way EVPI and EVPPI results are calculated. In general, the expected value of perfect information around a parameter will be sizeable when there is a high probability that the parameter will resolve at extreme values, and, at the same time, the parameter itself has a significant impact on the difference in the cost-effectiveness between treatments and thus on the adoption decision. In this study, NSCLC utility scores affected the results for Gem+Carb and Gem+Cisp in a proportional fashion, and thus, from the perspective of decision-making, obtaining information around these parameters is not seen as a priority.

Last, it must be noted that the degree to which VoI results are correct depends largely on the validity of the decision model through which results are produced. Serious flaws in the model—for example incorrect structure, biased input estimates and, most importantly, inappropriate representation of uncertainty—are all expected to give inaccurate cost-effectiveness results, biased estimates of uncertainty and, as a consequence, incorrect VoI results. Importantly, placing confidence in the results of the cost-effectiveness analysis preceding VoI analysis is a prerequisite for taking into account the results of VoI itself.

8.4. Sensitivity to assumptions

A further important consideration relates to the extent to which 'payback' and VoI results are sensitivity to assumptions. As noted earlier, 'payback' requires a series of assumptions around possible outcomes, change in clinical practice and likelihood of observing each of the specified outcomes. Deterministic and probabilistic sensitivity analyses reported in Chapter 6 revealed that different patterns of possible change in practice, alternative research outcomes and different likelihood weights all had a sizeable effect on 'payback' results.

On the other hand, 'value of information' results were greatly affected by the degree of uncertainty surrounding parameters which affect the adoption-related decisions, with greater uncertainty being associated with greater estimates of expected benefits from research. As it was shown in Chapter 7, the effect is more pronounced when greater uncertainty surrounds parameters that are instrumental in determining the most cost-effective treatment—in the specific applications, rates of disease progression. Further, both 'payback' and VoI are sensitive to assumptions about the employed time horizon for which the produced evidence is expected to be useful; as expected, long time horizons inflated the number of patients that are affected by the availability of improved information, increased the expected benefits in the population and made further research appear more desirable.

8.5. Practicality and ease of use

Practical aspects, such as the time and expertise needed to conduct 'payback' and VoI analysis, are likely to be an important consideration for research funding organisations.

Based on observations made in the course of this study, preliminary steps needed for the application of the analytic frameworks, that is, systematic reviews and modelling, required a fair amount of time. Systematic reviews of the literature to identify evidence for the NSCLC and HRPC case studies took six and four weeks to complete, respectively. Most of this time was devoted to assessing articles for inclusion and extracting relevant information. Considerable amounts of time were required for planning the structure of the decision analytic models, for converting data (e.g. patient-level observations obtained from the TRAPEZE Phase II and survival curves taken from the published literature) into a form appropriate for use in the models, and for carrying out deterministic and probabilistic sensitivity analyses. In total, systematic reviews and modelling for the NSCLC and HRPC case studies took 24 and 28 weeks, respectively (Table 8.1).

Table 8.1: Time required for preliminary tasks of practical application (collection of evidence and decision modelling)

Task	NSCLC	HRPC
Identification of evidence	6 weeks	4 weeks
Development of model structure	6 weeks	6 weeks
Conversion of raw information into appropriate form for use in the models	3 weeks	8 weeks
Model analysis (including deterministic and probabilistic sensitivity analyses)	9 weeks	10 weeks

Carrying out 'payback' analysis using the PATHS model was relatively straightforward. Excluding the time needed for familiarising with the methods, the base case PATHS analysis took approximately two weeks, including the time needed to gather information through meetings with the involved experts. A further two weeks were required for undertaking additional deterministic and probabilistic sensitivity analyses. No particular expertise other

than knowledge of basic concepts in cost-effectiveness analysis was needed to carry out the 'payback' analysis.

EVPI and 'value of implementation' analyses were also relatively straightforward and were completed soon after the probabilistic results of the models became available. EVPPI analysis was more complex and required between six and seven weeks to undertake. A share of this time was taken up by recurring calculations of simulated results. Running a single outer loop (i.e. 1000 inner loop simulations) through Microsoft Excel® macro commands on a personal computer equipped with an Intel Core 2 Duo® processor took approximately 50 seconds. A combination of 1000 outer × 1000 inner loop simulations, which was needed to derive EVPPI results for one group of parameters in NSCLC, required approximately 14 hours of continuous computation. EVPPI analysis for the five groups of parameters in NSCLC completed in approximately 70 hours (nearly three days). EVPPI calculations for HRPC were more demanding due to the more complex structure of the HRPC model: a combination of 1000 outer × 1000 inner loop simulations for one group of parameters took approximately 35.5 hours (nearly one and a half day) to complete, while running EVPPI calculations for all the 12 separate groups of parameters in this analysis would require 423 hours (approximately 18 days). However, it was possible to run parts of these simulations in parallel on many computers at the same time, which allowed completing the computations in approximately one week using 12 separate machines. No particular expertise other than an understanding of the concept was needed for conducting EVPI; nonetheless, undertaking EVPPI analysis required acquiring familiarity with command writing in the Visual Basic® programming language.

Conducting EVSI was more complex, as it involved a two-part process. The first part involved generating a set of 1000 posterior distributions for the group of parameters of interest (effectiveness of treatments of interest) and needed to be carried out for each of the assessed treatments in NSCLC (Gem+Cisp, Gem+Carb) and HRPC (DP, DP+ZA, DP+Sr89, DP+ZA+Sr89). For one treatment, generating posterior distributions required approximately 30 minutes of continuous computational time for HRPC, and approximately 120 minutes for NSCLC. The difference in these times appears to be due to the more complex calculations needed for generating posterior distributions for NSCLC, where existing data needed to be translated into number of patients alive and progression-free at different points in time, and simulated data needed to be translated back into survival and progression-free curves.

For each of the 1000 posterior distributions, the second part involved running 1000 Monte Carlo simulations (1000 posteriors (outer loop) $\times$ 1000 MC simulations (inner) loop) for each of the assessed treatments. Generating EVSI results for one treatment and one sample size (300 patients in base case analysis of TRAPEZE, 450 patients in the base case analysis of BTOG2) took approximately 11 hours for HRPC and 12 hours for NSCLC. The process was repeated for all treatments and for different sample sizes. The time needed for payback and VoI analyses is given in Table 8.2.

Table 8.2: Time required for application of 'payback' (PATHS) and VoI methods to case studies.

Task	NSCLC case study	HRPC case study
EVPI analysis	< 1 week*	< 1 week*
EVPPI analysis	6 weeks	7 weeks
EVSI and ENBS analysis	5 weeks	10 weeks
Value of implementation analysis	1 week	1 week
Payback analysis (using the PATHS model)	4 weeks	4 weeks
*After cost-effectiveness results from Monte Carlo simulations became available.		

In summary, the average time needed for information gathering and decision modelling was six and a half months for one case study. Excluding the time needed to familiarise with the respective methods, 'payback' and VoI analyses for a single case study required on average one and three month to complete, respectively. The above observations should be seen in the light of the following considerations. First, in the present study all tasks were carried out by a single researcher, whereas, if this work was to be commissioned, systematic reviews and modelling would be undertaken by a team of researchers with expertise in different areas, such as information specialist, systematic reviewers and health economists. Clearly, if the latter was the case, these tasks would have been completed in a shorter time. Second, it is likely that researchers who have undertaken 'payback' and 'VoI' analyses before would complete the applications and obtain results in a shorter period. Last, running EVPPI and EVSI calculations in software other than Microsoft Excel®, for example WinBUGS, may have been more time-efficient, although familiarity with such software would, in itself, require considerable expertise and investment in time. Notwithstanding this, it is thought that the time for undertaking these analyses as recorded here is a fair reflection of the time that

would be needed for applying the frameworks in actual practice, for decision making purposes.

8.6. Ability to inform relevant priority-setting decisions

The ultimate purpose of analytic approaches is to inform research funding allocation decisions and, with this in mind, this section looks into how 'payback' and VoI can assist with different tasks relevant to priority-setting. The section starts by looking into the meaning of final 'payback' and VoI results, and explores whether the methods can help with ranking research proposals and establishing optimal trial design characteristics.

8.6.1. Meaning of results and decision rules

A first step in establishing the potential of the approaches to assist with research funding requires clarifying what the results show and how they can be used.

With regards to 'payback', results express estimates of the additional cost and benefits (or NMBs) expected to accrue from carrying out a given research study, under specific assumptions about the possible study results and specific hypotheses about change in clinical practice in view and in the absence of these results. The decision rules used in the framework are alike to those used in cost-effectiveness and cost-utility analyses: research is worth undertaking if a unit of incremental health benefit arising from conducting research (over not conducting research) can be obtained at a cost less than the maximum value that society (or a decision-maker) is willing to pay for this benefit (i.e. ceiling ratio):

$$EICER_{reserch\ vs.\ no\ research} < ceiling\ ratio\ \lambda$$

or, equivalently, if the NMBs associated with research at a specific ceiling ratio exceed those of no research:

$$INMB_{research\ vs\ no\ research} > £0$$

Given this, 'payback' results offer an indication of the value of carrying out a specific research study (e.g. a trial) and can be used by research funders as a simple rule for deciding if the research study should be taken forward.

On the other hand, VoI results quantify the benefits expected to arise from eliminating or reducing uncertainty around a decision by obtaining better information through research. In particular, EVPI and EVPPI show the expected additional NMBs from making a decision with no uncertainty and perfect information about all or specific parameters affecting a decision problem, as opposed to making the same decision under current, imperfect information. Because of this, EVPI and EVPPI do not show the value of a specific study *per se*, but represent the maximum expected benefits that could be realised through research and, as a result, the maximum possible value of conducting research around a specific decision problem. According to the decision rule attached to EVPI, a study is potentially—although not necessarily—worthwhile if:

$$EVPI\ (or\ EVPPI) - cost\ of\ proposed\ research > £0$$

Results from the application of EVPI to NSCLC and HRPC showed that the maximum benefits from conducting research in these areas exceeded the cost of the BTOG-2 and TRAPEZE phase III trials, thus conducting the trials would be potentially worthwhile.

Last, EVSI results show the additional benefits expected from reducing uncertainty through conducting a specific piece of research (e.g. clinical trial), of a specific sample size n[51;115]. EVSI can be compared against of cost of this study to give the expected net benefit of sampling (ENBS). Positive ENBS suggests that the specific study is worth conducting:

$$EVSI_n - cost\ of\ proposed\ research > £0$$

While EVPI and EVPPI provide only a criterion—or a hurdle— for judging whether research in the area would be potentially beneficial and should not be ruled out, EVSI provides an estimate of the benefits associated with a specific study (e.g. the TRAPEZE phase III trial) and, thus, its results can indicate whether a specific study is worth undertaking. In the application of EVSI to the HRPC case study, results showed that conducting the TRAPEZE phase III trial is worthwhile, as it would bring about expected NMBs of approximately £550,000. In the NSCLC case study, EVSI analysis suggested considerable NMBs from conducting the proposed BTOG2 trial, of the order of £9 million.

8.6.2. Priorities across research proposals

Comparisons between different research proposals would be highly useful to research funders, as this would allow ranking these proposals in order of expected returns. Such comparisons require knowing the exact—rather than the maximum—amount of benefits that a research programme may offer.

In theory, different proposals could be ranked in order of their 'payback' results. However, the recognition that 'payback' estimates are based on a number of assumptions and

uncertainties about possible research outcomes and change in clinical, make ranking challenging. An additional level of complexity arises from the fact that 'payback' generates a range of possible results (e.g. 'optimistic', 'neutral' and 'pessimistic') rather than a single estimate, which may also overlap. For example, 'payback' results for NSCLC range from £1.88 to £2.38; while the results for HRPC span from £1.30 to £3.19. In this case, it is not clear whether the BTOG-2 trial should take priority over the TRAPEZE trial.

In VoI, EVPI/EVPPI results express the maximum possible NMBs from conducting research thus these methods cannot indicate the exact (marginal) NMBs that would be expected from a specific study. Comparing research proposals on the basis of EVPI/EVPPI estimates would only be appropriate under the assumptions that: a) the marginal benefits expected from a specific research programme in a given area are proportional to the maximum expected benefits from research in this area[51], b) the cost of research is the same across the compared proposals and c) these assumption hold true across different areas. For instance, knowing that the estimated population EVPI for the NSCLC decision problem is higher than that of HRPC (£13.08 million compared to £8.55 million) does not suggest that a trial in NSCLC would necessarily result in greater marginal benefits than a trial in HRPC.

Nonetheless, comparisons across programmes are, in principle, possible by using EVSI analysis where a series of trials of different designs, in different areas, can be ranked in terms of their ENBS. Given the ENBS results for BTOG2 and TRAPEZE, it is sensible to infer that priority should be given to the BTOG2 trial (£9 million in NMBs) over the TRAPEZE trial (£550,000 in NMBs).

8.6.3. Additional decision points

Having information around different aspects of research funding may be advantageous to funding organisation. The potential of VoI to help with trial design decisions is well documented. This is based on the principle that a trial is designed optimally when the difference between the marginal benefits expected from the specific trial (i.e. EVSI) and the cost of the trial is the maximum possible (i.e. ENBS is maximum)[262]. In theory, this principle can provide a guide for identifying an array of optimal design characteristics such as sample size, choice of treatment arms as well as allocation of participants across trial arms[51]. As reported in Chapter 7, EVSI analysis applied to the NSCLC case study indicated that sample sizes greater than 450 patients per arm would result in positive ENBS, with maximum ENBS achieved when employing the largest sample size (n=500 patients) amongst those investigated in this case study. Similarly, EVSI for the HRPC case study was able to suggest that the sample size for the TRAPEZE phase III trial that maximised the ENBS at £30,000 per QALY amongst the assessed sample sizes was 200 patients per arm.

Signs of the potential of 'payback' to assist with design decisions have been seen in existing studies. One way of determining appropriate trial design characteristics is by asking decision-makers to indicate what trial outcomes they consider relevant and what magnitude of difference would persuade them to adopt or refute an assessed treatment. For example, in a case study of a trial for postnatal midwife support, Townsend *et al*.[67] found that decision-makers tasked with deciding on whether or not to commission a midwife programme would prefer trial evidence on the programme's impact on mothers' breast feeding rates or postnatal depression and, owing to this, the authors recommended changes in the trial

design to incorporating these as the primary outcomes. Additional work related to trial design considerations has looked into the relationship, or trade-off, between the choice of minimum clinical difference between treatments ( ) to be detected as significant in a trial (smaller values of require larger sample sizes and greater cost) and the fact that observing a smaller differences as significant improves the chances of accepting the trial results and changing practice in a beneficial way[91;102].

An additional design characteristic relates to selecting treatments that should be assessed in a proposed trial. Excluding irrelevant treatments from comparisons is expected to lead to cost savings, as well as to benefits from not subjecting patients to ineffective treatments. As far as 'payback' is concerned, the method may indicate treatments that are not expected to lead to changes in clinical practice and this may help decision-makers to exclude the irrelevant treatment arms from further comparisons. A more systematic criterion is suggested by VoI: at a specific ceiling ratio, a treatment is not worth considering further if, in a large number of simulations, it is never the treatment that offers the greatest NMBs (i.e. the probability of the treatment being the most cost-effective option is zero)[51]. In the EVSI applications for NSCLC and HRPC, all treatments were associated with a non-zero probability of being cost-effective, suggesting that no treatment should be excluded from comparison in the proposed BTOG2 and TRAPEZE trials.

Often, decisions are needed as to whether a phase II trial should continue to a subsequent phase III stage and, in general, whether an on-going trial should continue or terminate. Such decisions are typically made on the basis of data collected for interim analysis or at the end of phase II trials. As noted earlier, this was the case with the TRAPEZE phase III trial, where

data from the phase II stage of the trial were available and could be taken into consideration when deciding whether to continue on to a phase III study.

In essence, decisions around funding trial continuation from phase II to phase III are similar to those about starting a new phase III trial, with the difference that, the availability of interim (phase II) data offers additional advantages. Data monitoring committees deciding whether a study should continue into phase III are traditionally interested in exploring if a) there is overwhelming evidence that one or more of the assessed treatments is clearly superior to the rest (in which case it is unethical not to treat patients with the superior treatment), b) there is evidence that the trial is futile (in this case it is evident that treatments are of similar effectiveness and further experimentation will be futile), or c) there are concerns about participants' safety. No economic criteria are taken into account when deciding whether an on-going study should be terminated, although there is potential for the analytic approaches to inform such decisions[277].

In 'payback', existing phase II data can be used to give an indication of the likelihood of a proposed phase III trial to show a particular result and they can point towards the most likely 'combination' (i.e. 'optimistic', 'neutral' or 'pessimistic'). In the present study, existing results from the TRAPEZE phase II trial showed DP+Sr89 to be effective. This information pointed towards a greater likelihood of the proposed TRAPEZE phase III trial to show positive results for this treatment, indicated that a greater weight may be placed on an 'optimistic' combination (i.e. higher possibility of DP+Sr89 being cost-effective) and suggested that a phase III trial would be likely to result in positive NMBs.

In other cases, where phase II data may indicate that 'inconclusive' results are more likely to transpire—which may mean no change in practice and no additional health benefits, but additional costs due to conducting the trial—potential users of the method may decide against conducting a phase III trial. In situations where early-stage patient-level data are available, there may be scope for combining available trial information with expert opinion to obtain *a priori* indications of the likely results of a future trial, possibly by using Bayesian methods.

In VoI, available phase II data are considered as 'existing evidence', they are incorporated in the decision model and they are taken into account when assessing the extent of uncertainty and the value of conducting further research. The fact that such data come from an experimental study which is directly relevant to the population of interest gives greater confidence in the results. The availability of phase II data, especially when those are expressed in terms of 'convenient' (conjugate) distributions is also beneficial in EVSI analysis. There, phase II data can be used directly to express 'prior' evidence. All in all, trial continuation decision can be addressed by 'payback' and VoI in the same way as decisions for new phase III trials, with additional advantages due to the availability of interim data.

8.6.4. Ability to inform priority-setting for evidence synthesis studies

Apart from primary evaluative research, funding organisations such as the NIHR HTA are often called to make decisions on funding secondary studies, such as evidence syntheses and economic evaluations.

'Payback' models appear to be able to help with such tasks; on this, Townsend *et al*.[67][p.x in Executive Summary] point out that the PATHS model *"could be applied to any form of research, including secondary analysis and reviews''*. In cases where formal cost-effectiveness evidence on a decision problem is not available, 'payback' can be carried out to establish the value of undertaking a joint evidence synthesis and economic evaluation study by following the usual sequence of steps (i.e. specification of possible outcomes of the evidence synthesis, estimation of subsequent change in practice and calculation of costs and benefits). Here, instead of looking into the value of research aimed to generate new primary evidence, the question is whether it would be worth summarising the existing evidence in the first place.

On the other hand, a joint evidence synthesis and economic evaluation study that assesses all available information is a prerequisite for conducting VoI, as it is needed to establish the extent of current evidence and give estimates of the existing uncertainty around an adoption decision. Thus, on the premise that all the available information need to have been already gathered and synthesised, there will be no benefit in carrying out VoI analysis to explore the value of conducting another study to synthesise the existing information.

8.7. Fit into priority-setting

The extent to which 'payback' and VoI can be incorporated into existing priority-setting arrangements is a crucial consideration for potential users of the frameworks [personal communication with Dr P. Davidson, Director of NIHR Evaluation Trials and Studies Coordinating Centre, 13-01-2012]. As noted in Chapter 2, funding for primary evaluative

research is typically distributed via 'researcher-led' ('reactive') and 'commissioned' ('proactive') streams[65]. These work streams are commonly employed by major funders of primary research in the UK, including the NIHR Health Technology Assessment (NIHR HTA) programme, the Efficacy and Mechanisms Evaluation (EME) and medical charities such as Cancer Research UK (CRUK)[41;45;46;49]. The following section looks into the scope for, and possible ways of, incorporating 'payback' and VoI into these funding streams.

8.7.1. Scope for use of 'payback' and VoI in 'researcher-led' funding streams

The aim of 'researcher-led' streams is to prioritise and fund research proposals submitted directly by researchers on topics of their choice. 'Researcher-led' streams are the main funding route used by medical charities such as the CRUK, but they are also employed by organisations such as the NIHR HTA and EME. A variation of such streams which is employed by the NIHR HTA are 'themed calls'; these follow the same processes as in 'researcher-led' routes with the difference that, in 'themed calls', researchers submit proposals on broad predetermined disease areas. There are only subtle differences in the prioritisation processes followed by different 'researcher-led' programmes, with the core steps are similar to those in the NIHR HTA 'researcher-led' stream. These are shown in Figure 8.1 and are described below.

The first step in the process involves researchers responding to requests for applications by submitting outline proposals. Outline proposals are checked for competitiveness and eligibility and those which fall within the programme's remit are forwarded to one of six Advisory Panels for consideration. Advisory Panels assess the outline proposals, shortlist

those that are deemed as good candidates and forward the list of them to the HTA Prioritisation Group (HTA PG) for further consideration. About half of the outline proposals are rejected by the HTA PG at this stage, while the remaining proposals are sent to the HTA Clinical Evaluation and Trials Board (HTA CET Board). The HTA CET Board makes the final decision about which proposals to reject and which to invite back as full proposals. Once full proposals are submitted, these are considered further by the HTA CET Board. Proposals are scored and those which rank high are recommended to the HTA PG for final approval[78;81;278].

In this process, either 'payback' or VoI can be carried out to indicate whether conducting research on the topic or research question that the proposed study deals with would be potentially worthwhile. Clearly, proposals on topics where further research is not expected to be worthwhile can be filtered out. For example, if primary evaluative research on chemotherapies for non-small cell lung cancer is not considered worthwhile—for example due to limited uncertainty and/or little scope for gains due to change in practice—there will be no benefit in conducting a trial on NSCLC, in which case the proposal can be ruled out.

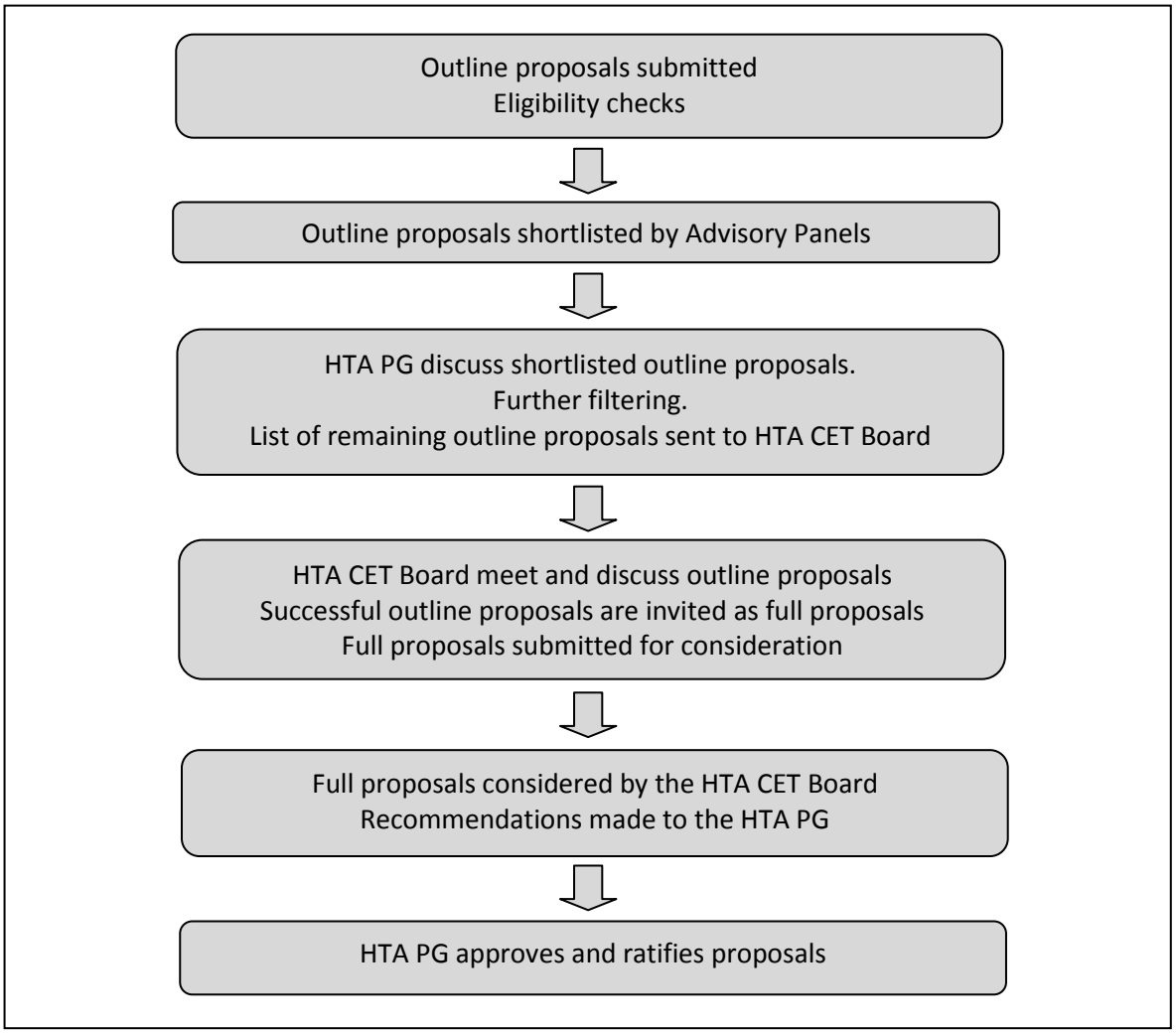

Figure 8.1: Flow chart showing prioritisation process in NIHR HTA 'researcher-led' stream

In addition, there may be scope for conducting EVSI and ENBS analysis to indicate the marginal benefits expected from a specific research proposal and to provide guidance about appropriate trial design (e.g. sample size, relevant comparators). In principle, such information can serve as a reference against which to compare the costs and the design characteristics of the proposed studies, although, as discussed in the next chapter, such comparisons will inevitably present challenges. In the context of the current arrangements, it is envisaged that 'payback' and VoI analyses could take place once full proposals are

submitted for consideration, so that their results become available when full proposals are assessed by the HTA CET Board, before final recommendations are made to the HTA PG.

8.7.2. Scope for use of 'payback' and VoI in 'commissioned' funding streams

'Commissioned' streams are central in the NIHR HTA and the EME programmes. In such streams, the aim is to identify and prioritise topics—rather than proposals—on which research is needed, as well as to commission teams of researchers to undertake this research. The process involved in the 'commissioned' stream of the NIHR HTA is shown in Figure 8.2.

A two-stage process is usually followed in commissioned streams. In the first stage, funding organisations gather suggestions for topics identified through various sources, including key stakeholders and the public. Once potential topics are gathered, they undergo an initial filtering stage to ensure that they are relevant to the NHS and do not overlap with past or on-going research. Topics deemed relevant are forwarded to a relevant Advisory Panel where they are assessed with reference to their importance. Topics that appear promising are shortlisted and are developed into 'vignettes', that is, brief documents which specify the topic question and summarise the existing evidence on the topic. Panels meet again to discuss the topics in view of the produced vignettes and decide which of the topics should be forwarded to the HTA PG for further consideration. The HTA PG meets to decide on topics to shortlist and advertise in calls for research[41;279].

Calls are typically for primary evaluative research and, less often, for secondary, evidence synthesis studies[280]. Topics are advertised together with commissioning briefs giving

clarifications about the topic, in response to which researchers submit outline proposals. Outline proposals are discussed and shortlisted by the Commissioning Board, taking into account considerations regarding the scientific quality of the proposal, justification of estimated sample size calculations and recruitment rates, ethical and social considerations related to the proposed research, as well as the proposal's cost. The most competitive outline proposals are shortlisted and invited back as full proposals. Once full proposals are received, they are discussed by the Commissioning Board in the light of external referees' comments and selected proposals are forwarded to the HTA PG for ratification[40;279;281].

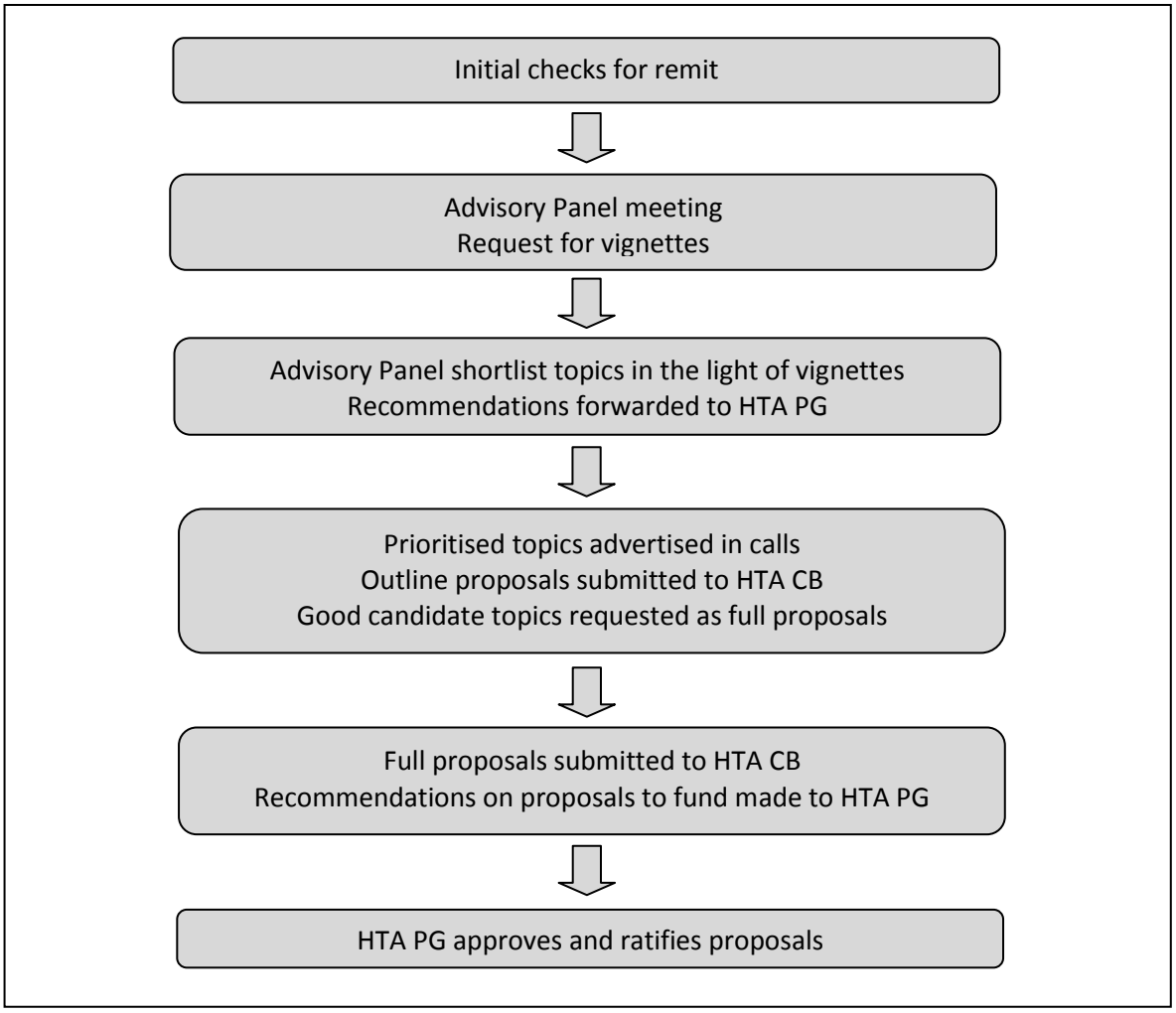

Figure 8.2: Flow chart showing prioritisation process in NIHR HTA 'commissioned' stream

In this process, 'payback' and VoI can be equally useful as an initial filter to exclude topics for which further research is not potentially worthwhile (i.e. topics for which the expected NMBs from research are unlikely to exceed the cost of commissioning and conducting research). In addition to suggestions on whether there are sufficient gains to be made from research on the topic, the approaches may be able to provide an indication of the maximum cost that a further study should not exceed. A previous study by Claxton *et al*.[110] suggested that VoI work could be carried our alongside the preparation of vignettes; however this may be problematic as, at this early stage, the research question (e.g. compared interventions) within the topic is unlikely to be well defined[110]. Thus, it is thought that 'payback' and VoI analysis would be better suited to a later stage in the process, just before the HTA PG makes recommendations for calls about topics to commission.

The ways 'payback' and VoI can help in the main funding streams of clinical evaluative research are further summarised in Table 8.3 below.

Table 8.3: Potential use of 'payback of research' and 'value of information' in 'researcher-led' and 'commissioned' funding streams

Funding stream	'Payback'	VoI
Researcher-led (such as the NIHR HTA Clinical Evaluation and Trials; EME; CRUK)	Undertake 'payback' for proposal which have been submitted as full versions. Results of 'payback' analysis can be used to inform final deliberations on whether the proposal should be funded.	Undertake VoI (EVPI) for proposals which have been invited as full submissions. EVPI results can be used to inform final deliberations as to the usefulness of research in the area. EVSI can be undertaken after the full proposal so that its results can help the HTA PG make recommendations for appropriate design, or it can be undertaken by researchers in the process of developing the proposal, to guide the design of the proposed study.
Commissioned (such as the NIHR HTA commissioned scheme)	Undertake 'payback' analysis for proposals which are considered by the HTA Prioritisation Group.	Undertake VoI (EVPI) analysis for proposals which are considered by the HTA Prioritisation Group.

It can be seen that, in principle, either of the frameworks can be suitable for either of the funding streams. In 'researcher-led' funding streams, 'payback' and VoI can be undertaken after a proposal has passed the first prioritisation state (i.e. after researchers have developed their preliminary proposal and submitted it as a full application) to provide information (i.e. expected NMBs of the study) which can be taken into account at the point where funding decisions are finalised. Similarly, in 'commissioned' programmes, VoI and 'payback' can be undertaken for topics that have passed an initial filtering topics, that is, topics which Advisory Panels judge to be important candidates for research. In this way, the filtering process is thought to provide a realistic solution to the possible concern that conducting 'payback' and VoI for the hundreds of topics and proposals submitted for funding will be, in essence, unrealistic.

8.8. Chapter overview

The chapter aimed to look into 'payback' and VoI by highlighting and discussing strengths and limitations of the methods with relation to considerations deemed relevant to potential users of the frameworks. Such considerations related to the validity and robustness of the results, practicality of undertaking the analyses, potential to inform different aspects of research funding decisions and scope for being incorporated in current priority-setting processes. Information presented in this chapter is taken forward and forms the basis for the discussion and conclusions presented in the last chapter of this thesis.

CHAPTER 9. Discussion and conclusions

This final chapter aims to discuss the findings of this research project, draw conclusions and make recommendations for further research. The first part summarises the project's aims, methods and main results. This is followed by a discussion around findings and relevant observations. The last part draws conclusions, makes policy recommendations and highlights areas for future research.

9.1. Overview of project aims and methods

Patient-level evidence generated from clinical trials is considered key input in assessing the effectiveness and cost-effectiveness of health care technologies[4;114]. Given the increasing demand for primary evidence and limited public resources for health care research, research funding organisations are routinely called to make decisions on which research proposals to fund. Although resource allocation decisions need to be informed by explicit and systematic assessments of the cost and benefit of different research programmes[55;80;90;92;100], such assessments are, at the moment, carried out implicitly, by panels of experts who infer the merits of proposals through discussion and deliberations.

A review of the literature identified nine analytic models which can assess the cost and potential benefits of proposed research in a systematic way. According to the principles underpinning them, the majority of the models were categorised into one of two overarching frameworks: 'payback of research' and 'value of information' (VoI). The 'payback' framework is based on the notion that research is worth conducting insofar as its

results can trigger a beneficial change in clinical practice. On the other hand, VoI stipulates that the value of research lies in its ability to reduce uncertainty about treatment adoption decisions and thus funding and conducting a research programme such as a clinical trial is potentially worthwhile when the expected benefits from eliminating or reducing decision uncertainty exceed its cost. Although the methodologies underpinning the frameworks are well established, their usefulness and role in informing research priority-setting remain, to a great extent, unclear. Despite the fact that this is not the first attempt to appraise 'payback' and VoI, existing studies present limitations which preclude a comprehensive assessment of the frameworks.

With this in mind, this project set out to explore the potential role and usefulness of the two most prominent analytic frameworks put forward for assisting priority-setting in primary evaluative research—'payback of research'[67] and 'value of information'[97;110]. To this end, the frameworks are assessed with regards to their ability to inform funding decisions, practicality, robustness and reliance to assumptions, theoretical and methodological soundness, and potential to fit into the existing priority-setting framework.

To obtain an insight into the frameworks' strengths and limitations, 'payback' and VoI were applied to two case studies. These represented proposals for primary evaluative research (BTOG-2 and TRAPEZE phase III trials) aimed to provide evidence for treatment adoption decisions in non-small cell lung cancer (NSCLC) and hormone-refractory prostate cancer (HRPC). At the time the present analysis took place the trials were funded and on-going thus the practical application was carried out in a retrospective manner, by looking at the points

in time when the BTOG-2 and TRAPEZE trials were considered for funding (2004 and 2006, respectively).

The analysis involved two steps. The first, preliminary step involved summarising the available pre-trial information around the treatment adoption decisions. This involved carrying out literature reviews to identify the information existing at the time that the decision to commission further research was considered, and constructing decision models to synthesise this information.

As a next step, 'payback' and VoI analyses were carried out to assess the need for, and the benefits from, obtaining additional information through funding and conducting the proposed BTOG-2 and TRAPEZE phase III trials. 'Payback' analysis was based on the PATHS[67] methodology and was applied to research projects looking into two-treatment and four-treatment comparisons (NSCLC and HRPC, respectively). VoI analysis involved all the relevant concepts, including expected value of perfect information (EVPI), expected value of perfect parameter information (EVPPI), expected value of sample information (EVSI) as well as analyses on the value of implementation.

The empirical work aimed to give a comprehensive view of the methods. In addition to estimates of the expected benefits from conducting the proposed trials, the practical application revealed strengths and limitations of the frameworks and formed the basis for the following discussion and conclusions.

9.2. Summary of practical application results

'Payback' and VoI results were calculated on the basis of a ceiling ratio of £30,000 per QALY. The assumed time horizons for the non-small cell lung cancer (NSCLC) and hormone-refractory prostate cancer (HRPC) case studies were five and two years, respectively.

'Payback' analysis for the NSCLC case study showed positive net monetary benefits (NMBs) for all the formed combinations, ranging from £1.88 million ('neutral' combination) to £2.38 million ('pessimistic' combination) in net monetary benefits (NMBs). Overall, 'payback' results suggested that, on the premise that the BTOG-2 trial would show the assumed results and would triggered a change in clinical practice of the hypothesised magnitude, carrying out the trial would result in additional benefits to the population.

For the same case study, VoI analysis showed a population EVPI in excess of £13 million and a relatively high EVPPI for clinical progression parameters. As the EVPI value exceeds the cost of the BTOG-2 trial, funding the trial would be potentially beneficial. Additional analysis around the value of implementation suggested that undertaking either research or implementation strategies is expected to be potentially worthwhile, although, in the specific case, pursuing better implementation would be preferable (i.e. it would results in greater numbers of NMBs). Last, EVSI analysis suggested that the benefits of obtaining more accurate information from the BTOG2 trial are expected to far exceed the cost of the trial, resulting in net benefits of sampling (ENBS) in excess of £9 million. Overall, both the frameworks suggested that funding and conducting the BTOG-2 trial is expected to result in additional benefits and should be funded.

In the prostate cancer case study, 'payback' results for the two-treatment comparison showed additional NMBs in the range of £1.30 million ('pessimistic' combination) to £3.19 million ('optimistic' combination), with a greater likelihood placed on the 'optimistic' combination. For the four-treatment comparison, results spanned from £468,000 to £3.34 million. In summary, 'payback' results suggested that, given the employed assumptions about possible outcomes and change in clinical practice, conducting the TRAPEZE phase III trial would result in additional NMBs.

The application of VoI to the prostate cancer case study showed a population EVPI of about £8.55 million. This value exceeds the cost of the proposed TRAPEZE phase III trial (£2.54 million), suggesting that the trial would be potentially beneficial and it is worth considering further. EVPPI was higher for clinical disease progression than for other parameters, such as costs and quality of life scores. The analysis suggested that carrying out a trial which would look into disease progression, cost and quality of life for all the chemotherapy options of interest—such as the TRAPEZE phase III trial—would be, on the basis of the obtained EVPPI values, potentially cost-effective. Further, value of implementation analysis suggested that funding either research or implementation strategies would be potentially worthwhile, although pursuing better implementation is expected to lead to greater benefits than conducting research. Last, expected value of sample information (EVSI) analysis showed that the EVSI associated with the TRAPEZE phase III study exceeded the cost of the trial, resulting in positive expected net benefits of sampling (ENBS) of about £550,000 and suggesting that the trial is a worthwhile investment.

It must be stressed that the sizeable difference between 'payback' and VoI results is, to a large extent, expected, given the differences in the methods employed to calculate them. On the one hand, EVPI results are high as they show the maximum expected benefits from conducting 'perfect' research, given that the results of research will be implemented optimally (e.g. all eligible patients will receive the most cost-effective treatment). On the other hand, 'payback' results represent the actual—not maximum—benefits that would be realised on the basis of less-than-perfect uptake of a cost-effective treatment in clinical practice and are net of the cost of the proposed trial.

9.3. Discussion

In a context of scarce public resources, normative decisions on how to allocate the available budget across different activities need to be guided by evidence on the opportunity costs and the expected benefits of competing activities[132;282]. Consistently with this, systematic, comparative assessments of treatments' costs and consequences are now routinely undertaken to inform decisions about funding and providing health care interventions and technologies[13;128;283]. On the same grounds, there is a clear justification for the use of explicit, analytic methods to assess the costs and potential benefits from funding different research programmes, given that information generated from such programmes is a public good, funded by limited public resources.

To date, none of the analytic models identified in the literature is used to guide funding decisions in the United Kingdom or, to the best of the author's knowledge, elsewhere in the world. Although as discussed in Chapter 3 there is a fair amount of literature around analytic

models—especially value of information—publications have mostly concentrated on methodological aspects, with the literature on how, and to what extent, these methods may be useful in assisting organisations with research funding decisions being limited. On this basis, it is believed that an in-depth exploration of the frameworks' potentials and limitations with regards to aspects deemed relevant to research funding organisations would help to resolve existing uncertainties.

9.3.1. Aspects related to the frameworks' theoretical and methodological validity

An important consideration for potential users relates to the validity of the produced results. As explained in the previous chapter, formal assessment of the predictive validity of 'payback' and VoI results poses significant difficulties, for both practical and methodological reasons. However, the degree to which validating 'payback' and VoI results is a strict requirement before they can be used in practice is debatable. On the one hand, confirming the validity of 'payback' and VoI results is likely to strengthen the case for using the frameworks in research priority-setting. On the other hand, unconfirmed validity does not necessarily imply that the produced results—and by extension the approaches—are invalid or flawed. It is worth noting that, despite the fact that results of widely used cost-effectiveness and cost-benefits analyses are rarely validated, these are widely accepted and used, as they are generated through methods which are thought to have robust theoretical and methodological bases[132;157].

The theoretical underpinning and methodological basis of 'payback' and VoI appear sound; however, there exist points that potential users need to be aware of. With regards to

'payback', the approach tends to give greater weight to research topics in areas where there is scope for a substantial change in clinical practice (i.e. significant increase in the use of a cost-effective treatment, or decrease in the use of a non-cost-effective treatment). In addition, the framework attributes a beneficial change in clinical practice—either greater adoption of a cost-effective treatment or restrictions on the use of a non-cost-effective treatment—solely to research. This, however, tends to overlook the impact of other factors on changes in clinical practice, such as effective marketing strategy by manufacturer or increased familiarity with a treatment. Thus, the actual contribution of research into a beneficial change towards cost-effective treatments, and therefore, the associated payback of research, is likely to be overestimated.

With reference to VoI, the framework tends to prioritise research in areas where there is substantial uncertainty around parameters affecting a treatment adoption decision. Central in this framework is the notion that research is useful insofar as it provides evidence that facilitates treatment adoption decisions. Although this notion, termed the 'decision-maker's viewpoint'[189] has supporters[115;284], it is in contrast with the currently prevalent paradigm which advocates that evaluative research should be conducted to test hypotheses and make inference about treatments' efficacy, effectiveness or cost-effectiveness[285].

In addition, in both the frameworks the estimated benefits from research—and thus, the inferred value of research—is directly proportional to the population that stands to benefit from improved information. As a result, proposals affecting larger eligible populations are more likely to result in greater NMBs. In such cases, the benefits from research are more likely to exceed the cost of research, making proposals in such areas good candidates for

funding. This, however, may be undesirable when it comes to research for rare conditions, where the fact that less people stand to benefit—and consequently less NMBs are expected from research—may lead to evaluative research in such areas appear unattractive and being overlooked.

Methodological challenges around the frameworks also exist. As discussed in Chapter 8, a prominent issue in 'payback' relates to the ambiguity in the way that 'possible outcomes' and 'likelihood weights' are specified. At the same time, the method becomes increasingly complicated for research proposals involving more than two treatments, and there is currently uncertainty around the optimal 'payback' methodology for multiple-treatment comparisons. Different methodological challenges are present in VoI. These relate to difficulties in conducting comprehensive EVSI analysis in situations where no simple analytic solution exists for combining prior evidence with predicted trial data, as well as issues related to the value of implementation in pragmatic situations where acquisition of further information is expected to have an impact on implementation.

Last, common to both the frameworks are challenges related to estimating the number of people who are expected to benefit from the research and results in the future. Such information is essential in projecting the benefits from research to the population, but it is inherently uncertain as it requires informed guesses on the number of years for which a technology of interest is expected to retain its usefulness.

9.3.2. Robustness of results and sensitivity to assumptions

A further important consideration relates to the extent to which 'payback' and VoI results are dependent on, and sensitive to, different assumptions. The empirical application of the framework showed that a series of assumptions are required in 'payback', some of which are necessary to 'operationalise' the method (e.g. assumptions about possible research results), while other assumptions are employed to compensate for lack of data (e.g. extent of change in clinical practice, likelihood of observing different results). Sensitivity analyses reported in Chapter 6 and discussed further in Chapter 8 showed that different assumptions had a notable effect on the results. On the other hand, VoI results are greatly affected by assumptions used in the decision analytic model, especially when these assumptions have an impact on the uncertainty around key parameters.

As mentioned above, both the frameworks are highly sensitive to assumptions around the number of patients who are expected to benefit from the results of research. A straightforward way of estimating this involves weighting the number of patients eligible for treatment—that is, the number of new cases per year—by the number of years before the technology is rendered obsolete. As explained later in this chapter, the latter component—commonly called the 'time horizon' of a technology—is subject to great uncertainty and it is particularly difficult to estimate with any degree of accuracy. Inevitably, this issue adds an extra layer of uncertainty to the obtained results.

Given the above, the difference between 'payback' and VoI appears to be that, in the former framework, assumptions are typically unavoidable because they are either inherent to its

methodology (e.g. specification of possible outcomes) or they represent 'guesses' for which empirical evidence is typically not available (e.g. rates of future change in clinical practice). In contrast, assumptions employed in VoI are often potentially avoidable, in that they are typically used to replace empirical data which may be unavailable in the particular instance, but they are, in general, accessible.

It must be noted that the use of assumptions in processes evaluating the costs and benefits of different activities, be it health care programmes or projects of public infrastructure, is, to a large extent, inevitable[132;286]. This would be expected to hold true for assessments of the value of future research, especially because such research is yet to take place and estimating its benefits requires 'guesses' and predictions. It is thought that, rather than discarding findings which, to some extent, are based on assumptions, the effort should be towards making sure the employed assumptions are made explicit, are plausible, and they are based on the best of the available knowledge.

9.3.3. Ease of use and practicality

It is anticipated that the extent to which 'payback' and VoI analyses are practical to undertake will be a crucial consideration for potential users. The empirical work showed that preliminary steps in the analysis—literature reviews and decision modelling—took approximately six and a half months to complete. Once decision models were constructed and their results were available, 'payback' was carried out in a relatively short period of time, within a month, while VoI took about three months, mostly due to time required for setting up the programming codes for EVPPI and EVSI, and running computations.

It must be noted that, while a probabilistic decision model is necessary for VoI analysis as it quantifies the uncertainty around an adoption decision on the basis of which VoI measures are calculated, 'payback' can be carried out without a model[67;92]. This is possible in situations where a trial is expected to give final—rather than intermediate—outcomes, such as QALYs, which can be easily combined with the cost of the assessed intervention to given summary cost-effectiveness results. In such cases, 'payback' can be calculated within a few weeks. Given the short time frames within which research funding organisations operate, this suggests a notable advantage for 'payback'.

None of the methods requires particular expertise other than an understanding of main concepts in economic evaluation, fair familiarity with decision analytic modelling and basic programming skills. 'Payback' analysis is straightforward to comprehend and undertake, involving only simple calculations which can be performed in commonly used spreadsheet applications. While VoI concepts are more complex and require some form of programming, there exists a wealth of resources, including textbooks and published articles, which offer guidance on how to carry out value of information analysis.

Previous work has looked into the time frames within which VoI (EVPI and EVPPI) and 'payback' (PATHS) analysis can be carried out. In their study, Claxton and colleagues[110] found that modelling and VoI (EVPI and EVPPI only) would take a team of researchers with different levels of experience approximately 10 to 12 weeks to carry out. Townsend *et al*.[67] found that the PATHS analysis can be undertaken within 1 to 4 weeks, depending on the complexity of the project. These estimates are in broad agreement with observations from the present study, considering that the latter was undertaken by a single researcher and

required building new models from the beginning. Evidently, if a systematic review and/or a decision model are already available for use, the time required for conducting the analyses would be considerably shorter.

9.3.4. Usefulness of 'payback' and 'value of information' results

As explained in the previous chapter, 'payback' and VoI are able to inform decisions on which topics or proposals to fund, and, when data is available they can assist in addressing questions related to determination of optimal research design and continuation of a phase II trial to phase III. A central question relates to how the generated results should be interpreted and, by extension, how they can be used to inform funding decisions.

'Payback' results show the expected benefits from conducting a research study given specific assumptions about the study's results and their impact on clinical practice. Positive NMBs indicate that a study is worth conducting, while negative NMBs indicate that the possible benefits from conducting research are exceeded by the cost of the study. Despite this relatively simple rule, making decisions in the light of 'payback' results poses challenges. First, because of the way 'payback' results are calculated and presented, decision-makers will need to make a judgement around which 'combination' (e.g. 'optimistic', 'neutral' or 'pessimistic') is more likely to transpire. Such judgements may be difficult, especially in cases where, as part of the same analysis, a particular scenario may show net losses (i.e. negative $INMB_{research\ vs.\ no\ research}$) associated with research while another scenario may show net gains (i.e. positive $INMB_{research\ vs.\ no\ research}$). Second, although in theory positive INMBs indicate that a proposal is a worthwhile investment, it is unclear whether there is a level of expected

returns—for example, NMBs less than £50,000—below which proposals would be unlikely to be funded. In the absence of clear rules, 'payback' results would still leave room for different interpretations and value judgements.

VoI measures that quantify the potential gains from 'perfect' information (such as EVPI and EVPPI) show the maximum expected benefits from research. Such information can be used as a 'hurdle', to rule out proposed research studies which are not potentially worthwhile (i.e. where the cost of research is greater than the expected maximum NMBs from further research). However, EVPI results can only provide a 'necessary' condition, that is, even if a research study passes the first hurdle and is deemed potentially worthwhile, this does not automatically imply that the study will be actually worth conducting[51;123]. This is because the actual (marginal) benefits from conducting the study may turn out to be outstripped by its cost. To decide on whether a particular study should be funded, one must know whether the marginal benefits of a research study exceed its cost.

Such indications can be given by EVSI and ENBS. These measures have been advocated as the most appropriate means of assisting priority-setting[55;123] and have been suggested as a toolbox for determining the design of clinical trials[115;262;287]. However, EVSI and ENBS have notable limitations.

First, EVSI is challenging to undertake on the basis of model results, as such analysis is particularly intricate in all but the most straightforward situations, where the considered 'prior' evidence and the simulated trial results are expressed as conjugate distributions[51]. Second, although in principle the method can indicate a study's optimal design, in practice

this is complex and particularly demanding. This is because establishing the optimal design for one aspect of the trial (e.g. optimal sample size) requires making sure that other design aspects (e.g. allocation of patients to trial arms) are also optimised. To date, the majority of the applications of non-parametric, model-based EVSI analyses—including the present EVSI analysis for HRPC and NSCLC—are based on the assumption that equal allocation of patients between trial arm is optimal (i.e. the marginal benefits of allocating a patient in one arm are the same as the marginal benefits of allocating the same patient in another arm). However, as Briggs *et al.*[51] point out, this is not usually the case, and different (unequal) allocation of patients is likely to result in greater benefits. In essence, recognising this would require assessing all possible sample sizes (i.e. by starting from no participants in the trial and adding one participant at a time) across all possible ways of allocating participants to trial arms, which makes the method highly complex and computationally demanding[51;262]

Third, even if the optimal design characteristics of a trial as a whole were specified, it is uncertain whether funding organisations would be inclined to use this information to determining the requirements (e.g. sample size) of studies for commissioning, or as a 'gauge' against which to assess the design of submitted research proposals. This is especially true in cases where these indications may contradict widely used and accepted ways of determining trial design characteristics (e.g. traditional power and sample size calculations). For example, according to widely-used power calculations based on minimising the chance of type II error, the minimum sample size for the TRAPEZE phase III study was determined at 300 patients per arm. However, the results of EVSI analysis in Chapter 7 suggested that recruiting 200 patients per arm is preferable. It is unclear whether and to what degree funding

organisations would be prepared to follow EVSI indications and fund a trial which, according to established methods, would appear underpowered.

A common criticism of VoI is that calculations in all concepts assume that the estimated NMBs from a decision represent the true NMBs that would be realised in the population. This overlooks the fact that, in practice, the predicted NMBs are rarely realised due to sub-optimal implementation. To address this criticism, Fenwick *et al*.[187] proposed a conceptual framework comprising a series of measure of the expected value of implementation. This work can, in principle, indicate the situations where undertaking research may be less beneficial than improving adherence to recommended treatments. However, the framework does not appear to answer the question *'what is the expected value of conducting research to generate information around a treatment adoption decision, given that the treatment adoption decision may not be perfectly implemented?'* As it stands at the moment, the framework is limited by the restrictive assumption that acquiring better information will have no effect on implementation, while it is unclear whether the results of value of implementation analysis can be of help to funding organisations with no remit—or budget—to fund implementation strategies.

Notwithstanding these issues, it is thought that there is value in using 'payback' and VoI as a means of assisting research funding decisions. However, given the above limitations, it is advocated that the most appropriate use of such results would be as indications for ruling out research which is not expected to bring about additional benefits, rather than as strict directions on which research programme should be funded.

9.3.5. Suitability of 'payback' and VoI in different contexts

The extent to which 'payback' and VoI can be compared against each other in order to establish which framework is 'overall superior' is unclear. The frameworks are based on sound underpinning principles and share the same overarching objective—to provide evidence that will inform research funding decisions. However, due to differences in the methodologies they use, the frameworks present distinctive characteristics, strengths and limitation. Although, as explained in Chapter 8, both frameworks are able to provide information that can be of use in the main streams of publicly funded research, there exist situations where each of 'payback' and VoI may be preferable.

VoI analysis aims to quantify the loss of benefits due to uncertainty around treatment adoption decisions, and identify the situations where further information to support such decisions would be beneficial. Given this, VoI is closely connected with treatment adoption decisions. Undertaken in conjunction with decision modelling, VoI has been advocated as a unified framework—often called decision-analytic value of information DA-VOI[111;120]—which aims to address treatment-related and research-related prioritisation simultaneously.

With this in mind, the framework is well placed to assist in situations where recommendations on conducting further research may be made conditional to, or even in parallel with, treatment adoption decisions[288]. This is typically the case for new treatments or devices which are considered for adoption by the National Institute for Health and Clinical Excellence (NICE). Indeed, recent work[122;289] has advocated that treatment approval decisions made by NICE need to consider whether further research is required in order to

substantiate a choice between policy options such as a) approval of a new treatment for the whole of the eligible population, without requesting further research; b) approval of the treatment only for those who already receive it in the context of research (i.e. patients in clinical trials), or c) approval under the condition that more research will be conducted. The latter option is particularly relevant in cases where conducting research after approval is granted may be infeasible or unethical[289;290]. In this context, VoI can be carried out as an adjunct to the NICE Technology Appraisal process looking at new medicines, devices or procedures, so that the whole exercise provides the necessary information to guide joint treatment approval and research recommendations. In practical terms, the fact that decision models developed for the purposes of technology appraisals can be readily accessible suggests that an important prerequisite for carrying out VoI analysis is already present.

On the other hand, 'payback' can provide timely indications on whether research and change in clinical practice would be beneficial, by identifying the situations where the benefits that would accrue from research and practice change would be enough to compensate for the cost of research. As the framework links research with the benefits of a change in clinical practice, it is well placed to assess research questions related to whether or not current practice is satisfactory, or a change would be beneficial. Given this, the framework appears to be in a stronger position to evaluate the benefits from research on topics related to treatments, services and arrangements which are currently in use at different rates, but for which good quality evidence is lacking. Such questions are particularly pertinent to the health care system and are routinely suggested as topics for research by the NHS and its stakeholders[40;41]. In addition, the fact that decision models are typically not

available for topics suggested to NIHR through sources other than the NICE Technology Appraisal programme is a practical hindrance for using VoI, but it does not impede 'payback', which, as Townsend *et al*.[67] have demonstrated, can be also used when a decision model is not available. Owing to the above, 'payback' appears to be particularly useful in assessing a wide range of topics. This, together with the fact that 'payback' can be carried out in situations where a decision model is not available make the method particularly practical and versatile.

9.3.6. Implementation of the approaches in research priority-setting

If analytic approaches are to be used routinely in research priority-setting, a number of practical questions around their use are expected to arise. A first, important question relates to the role of analytic approaches, in relation to the way in which results from 'payback' and VoI analyses should be considered and used in research priority-setting. Different options exist. Under one option, results of analytic frameworks could be used in a 'prescriptive' manner, to replace the existing deliberative processes. In such a case, these results would be the only determinant of funding decisions. As expected, under this option there would be no need for panel discussions, and deliberative approaches would be rendered obsolete. However, such a radical approach would be undesirable, as there exist ethical and distributional considerations which, while they are not reflected in the results of analytic approaches, they may be considered pertinent. For instance, analytic approaches may find the NMBs from research on particular topics to be low, but society may consider the topic as a priority. This may be the case where research on a topic may facilitate a more equitable distribution of health care among the population, or target 'rare' diseases for which the

aggregate gains in the population may be low, but for which research is needed to establish the most appropriate care.

A more pragmatic approach would involve using the results of analytic approaches as additional input in the process of decision making, alongside other relevant considerations and criteria. Such an approach would be more aligned to the way evidence from analytic methodologies such as cost-effectiveness, cost-utility and cost-benefit analyses are used in the process of making decisions for treatment recommendations[291;292]. Under this, more realistic, scenario there would still be a clear role for deliberative approaches, which would be useful at different stages of the prioritisation process. At early stages, panel discussion would continue to take place in order to filter out proposals and topics which are out-of-remit, irrelevant, or for which there is wide agreement that there is little need for further research. Most importantly, deliberative methods would still be needed at the last stages of the prioritisation process, to combine the output of analytic approaches with other criteria and judgments, and to ensure that all relevant ethical and distributional considerations are discussed and taken into account in final funding recommendations.

A further question relates to who should be tasked with undertaking the additional analyses. In the context of 'proactive' funding streams, where institutions such as the NIHR commission research on proposed topics, 'payback' or VoI analysis could be undertaken either internally, within the organisation, or externally, by academic teams commissioned to carry out this work on behalf of the organisation. The latter is more consistent with current arrangements where primary and secondary research commissioned on behalf of different NHS stakeholders is undertaken by contracted teams of researchers[293]. The option of

institutions undertaking ‘payback’ and VoI analyses for their own use may also be feasible, although, given the fact that this would require significant increases in organisations’ capacity to deal with this additional work, it is thought that this option would represent a long-term, rather than a short-term goal.

In ‘researcher-led’ streams, where teams of researchers put forward proposals for research on topics of their choice, these additional analyses can be undertaken by either the team of researchers who submit the proposal, or by commissioned academic institutions. In the former case, researchers can present ‘payback’ or VoI results as part of their research proposal. In the short term, commissioning independent academic institutions to carry out this work, instead of asking researchers who are involved in the specific proposal to undertake the analysis, is thought to offer notable advantages: it would remove the burden from researchers who may be lacking the expertise to carry out these additional tasks, while it would provide reassurance that the analyses would not come from teams of researchers who have a vested interest in showing that research is worthwhile and should be conducted. On the other hand, such an option would inevitably imply that the costs of undertaking the analysis would be fully borne by the organisation. In the long term, it would be beneficial for ‘payback’ and VoI analyses to be undertaken by researchers. This would mitigate the cost of commissioning the analyses, while, in parallel, it would boost the interest in the frameworks which would be likely to lead to further development and improvements in their methods.

It is thought that an important step to this end would involve research funding organisations adopting a proactive stance, where explicit analyses of the value of proposed studies are formally requested in commissioning briefs and calls for research, and are considered as an

essential part in an application for funding. As explained above, it is thought that 'payback' analysis would be the method of choice for proposals relating to existing treatments and procedures which are commonly used in practice, in order to establish the benefits from research and change in current clinical practice, while VoI would be preferable for assessing research proposals looking at new treatments or existing treatments proposed for different indications, to suggest whether further research is needed to substantiate adoption decisions.

It is also believed that a more widespread use of analytic methods will be difficult without making the methods acceptable to committees and decision makers. It is thought that such a task would require explaining the principles, methods, strengths and weaknesses of the approaches to committee members, as well as by communicating to them the benefits of using explicit, systematically-produced evidence of the value of research, possibly through open discussion and workshops.

Advocates of 'payback' and VoI have highlighted that, at present, limitations in capacity make it difficult to carry out these analysis for every proposal and topic considered by a funding organisation[67;110]. Thus, some judgement would be needed to determine which proposals should be subjected to analytic assessment. The obvious solution would be to carry out 'payback' or VoI analysis only on proposals which have reached the last stages in the prioritisation process. In this way, it will be assured that time and effort is not spent on proposals which are out-of-remit, or for which experts and reviewers agree that they do not represent good candidates for research.

Alternatively, the analyses can be restricted to proposals which are deemed particularly costly. As Townsend *et al*.[67[p.54]] explain, a monetary cut-off point could be used, above which all proposed trials should be evaluated using analytic approaches: *"The authors recommend that formal analysis of potential payback, along these lines* [i.e. the developed PATHS model]*, should be undertaken as part of an on-going evaluation for projects costing over a certain threshold of, say, £250,000. For very expensive projects, some formal value of information analysis might also be routinely appropriate."* Using a monetary threshold to identify costly trials is a unambiguous and straightforward criterion, although complications would be expected to arise if a universal threshold was to be determined, given that the cost of research is expected to vary across disease areas. Indeed, studies in disease areas with requirements for long follow up durations and large numbers of participants would inevitably be more expensive than smaller and shorter studies which may be adequate in other disease areas. Owing to this, it may be preferable for any monetary cut-off points to be determined by taking into account particular aspects and characteristics of research in different disease areas.

It is expected that modifications in the existing prioritisation processes may be needed to account for the additional time required for undertaking the analyses. On the basis of the observations made in the present study, the time for conducting these analyses is not expected to be substantial and is unlikely to result in significant delays in decision-making. It is also anticipated that additional resources will be required to fund the analyses *per se*. These resources will have to come out of a health care budget which is already stretched and, although as of 2013 it will not be subject to cuts, it is planned to remain at current

levels[294]. Nonetheless, the fact that further analyses aim to provide assurance that scarce resources are spent more prudently is consistent with the policy aims of containing unnecessary spending and improving efficiency[295]. At the same time, it is thought that the cost of undertaking these additional analyses are likely to be counterbalanced by the potential cost-savings from not conducting a large and lengthy trial which would provide limited returns.

A further question relates to which organisations would be better placed to use these analytic approaches. Both 'payback' and VoI are underpinned by the notion that the value of evaluative research lies in its potential to inform treatment adoption decisions which will provide benefits to the population. According to this stance, benefits from research other than those accruing from improved decision-making (e.g. advances in knowledge) are secondary and, therefore, are not accounted for in the results[67]. Owing to this, it is thought that organisations with greater potential to benefit from the use of analytic approaches are those which fund research with a view to answering specific questions on treatment provision within the health care system, such as the NIHR HTA. As an implication, funders whose primary aim is to produce scientific knowledge *per se*, such as medical charities, may be reluctant to take 'payback' and VoI results into account [personal communication with Ms J. Hearn, Head of Clinical Trials, Cancer Research UK, 29-06-2011]. Nonetheless, on the premise that such organisations seek to allocate their research budget in a way that provides the greatest benefit to the population, it is argued that indications of the 'cost-effectiveness' of research are relevant and should be accounted for in decision-making, even if they are deemed as secondary considerations.

9.4. Conclusions and policy recommendations

It is widely agreed that decisions on how to allocate limited public resources need to be made in the light of explicit evidence on competing activities' potential benefits and costs[128;132]. In recognition of this, analytic methods—such as cost-effectiveness and cost-benefit analyses—are used widely and play an important role in informing treatment adoption decisions[8;156;283]. In the area of health care research, the crucial role of providing explicit information on research programmes' cost and potential benefits can be fulfilled by analytic frameworks, the use of which is equally justified.

At the moment, explicit and systematically-produced evidence on the potential cost and benefits of research proposals is neither generated routinely, nor is it taken into account in research funding decisions. Such decisions are made exclusively by using 'deliberative' processes, where panels of experts assess and infer the merits of proposals through discussion and deliberations. Two main analytic frameworks have been proposed to generate explicit evidence on the value of research proposals and enhance the transparency of research prioritisation decisions—'payback of research' and 'value of information'. Although they make use of different methodologies, the frameworks have similarities and share a common aim: to provide prospective information that will guide and inform research funding decisions. In doing so, the 'payback' framework recognises that research is worth conducting insofar as its results can trigger a beneficial change in clinical practice, and infers the value of research from the estimated benefits due to change in practice. On the other hand, VoI stipulates that the value of research lies in its ability to reduce uncertainty about

treatment adoption decisions, and thus, it estimates the value of research according to decreases in uncertainty and reductions in the expected opportunity loss.

Although there exists a fair amount of literature explaining the mechanics of the frameworks (particularly VoI), there is little discussion around their potential role and value in priority-setting. Such evidence is crucial in establishing whether analytic approaches should be introduced and used to guide research funding decisions. With this in mind, the present work aimed to give a comprehensive assessment of the frameworks' strengths, limitations and usefulness, and provides recommendations around their potential role and use.

Both the frameworks are based on well-grounded principles and use sound logic. In general, their methodologies are clear and coherent, although, there exist aspects where further methodological research would be needed. Key aspects relate to the specification of possible outcomes in 'payback' methods, further development of concepts related to implementation, and research around the practicalities of conducting EVSI analysis.

As expected, both the methods make use of assumptions, some of which are inherent to their methodologies while others are due to lack of data. Key assumptions needed to operationalize the methods relate to possible results of further research, the existing uncertainty around parameters and the future number of patients who would stand to benefit from research. The use of assumptions instead of empirical data adds an extra layer of uncertainty in the results. However, given the fact that the purpose of these analytic frameworks is to predict the value of research prospectively, before research takes place, the use of assumption is, to a large extent, unavoidable. The analyses are relatively

straightforward to undertake, although some concepts such as the expected value of perfect information for parameters (EVPPI) and the expected value of sample information (EVSI) are more complex and may require longer time frames to complete. Observations made in the context of the present empirical analyses suggest that the expertise and time required for undertaking 'payback' and VoI is unlikely to constitute a significant burden to funding organisations and researchers, and are comparable to those needed for conducting model-based economic evaluations, which are nowadays well established and routinely undertaken.

The frameworks are valuable in that they provide straightforward indications on whether the return to investments in evaluative research would justify their cost. Given this, they can suggest topics on which research would be beneficial, and indicate whether specific proposals are worth undertaking. In this way, the approaches can be helpful in commissioned streams, which aim to prioritise topics for research proposed by different stakeholders within the health care system, and 'researcher-led' streams, which focus on prioritising proposals for clinical trials submitted directly by researchers.

In light of the above, it is thought that analytic approaches such as 'payback of research' and VoI are valuable and have a clear role in assisting priority-setting for evaluative research. The frameworks offer explicit estimates of the expected value of research and, in this way, they can provide additional assurance that scarce resources are committed to research which is likely to confer benefits to the population.

While both ‘payback’ and VoI can be used in a variety of occasions, each of them appears to be better placed to answer specific types of questions, and thus they may be preferable in different situations. VoI has been seen as a framework that is intertwined with treatment approval decisions and, as such, the method appears to have an advantage in dealing with situations where recommendations for further research need to be made for new treatments, often in conjunction with treatment approval decisions. This is typically the case when new treatments and technologies are assessed through technology appraisal pathways, to determine whether they should be covered and made available to the public. On the other hand, the principles underlying ‘payback’, and in particular the framework’s ability to take into account the benefits expected to be brought about by informed changes in the use of treatments suggest that the framework is placed favourably to help in situations where research is proposed to provide evidence around existing treatments and technologies which are currently in use, but for which evidence on their appropriateness is weak or inconclusive. Such questions are commonplace and constitute a large part of the topics and questions that the NHS is interested in addressing. This, together with the fact that the methods is less reliant on the existence of a decision analytic model make ‘payback’ applicable to a wider range of situations.

It is important to recognise that analytic methods present specific limitations and, with this is mind, it is thought that the results of analytic approaches should be seen as input in the process of prioritisation, rather than as the sole criterion for a funding decision. Results can be combined with other considerations which are deemed relevant and are raised through discussion and deliberations. It is thought that deliberative approaches will remain a useful

component in the prioritisation process, but they should be used in conjunction with analytic methods—the latter to provide explicit evidence of the value of research, and the former to combine this evidence with further consideration and criteria.

It is thought that a proactive stance will be needed in order for analytic approaches to become an internal part of the prioritisation process. One the one hand, this would involve funding organisations requesting formal analyses of the value of research as part of the submitted proposals. On the other hand, a more prominent role of analytic approaches would require making such approaches more acceptable to experts and committees, by conveying the benefits of using explicit, systematically produced evidence in the process of decision making.

Inevitably, the likely introduction and use of the approaches would be expected to raise a number of practical questions. Key questions would relate to who should be tasked with undertaking these extra analyses and in what cases analytic approaches should be undertaken and how their results should be used. It is recommended that, in the short term, it would be beneficial for this work to be commissioned to contracted academic institutions, although, in the long term, there would be advantages in requesting researchers to undertake this analysis and present the results as part of their proposal, to substantiate their case for funding. Given current capacity constraints, it would appear infeasible to carry out 'payback' or VoI analysis for every proposal submitted for funding. Current processes where research proposals are selected through different prioritisation stages can be used to filter specific proposals which are considered as good candidates for research, but for which further evidence would be needed to support a funding decision.

In summary, it is thought that there is a clear role for analytic approaches in the process of priority-setting. Although the approaches are not a panacea, it is believed that their introduction and use would enhance the explicitness of the process, and would provide additional assurance that valuable research resources are allocated in an efficient way.

9.5. Key strengths and limitations

Work undertaken for the aims of this thesis presents certain strengths. First, literature reviews carried out throughout this work followed a systematic approach and were undertaken through methods recommended in published guidelines[56]. Reviews involved searches in various sources (e.g. electronic bibliographic databases, portals of funding organisations, key articles' reference list and the internet in general) in order to minimise the chance that relevant information is missed, while the selection of relevant literature followed an iterative approach and was carried out by using predetermined inclusion and exclusion criteria.

A further strength relates to the use of decision analytic models, which were built to synthesise the available information and served as vehicles for applying the 'payback' and VoI to the specific case studies. Although the models did not aim to inform 'real world' treatment adoption decisions, they were developed with the intention to be as comprehensive as possible. Input for the analyses was identified through systematic reviews of the literature, while the methods used in structuring and populating the models were in agreement with guidelines on 'good practice' in modelling[296;297]. As suggested in the literature, both deterministic and probabilistic sensitivity analyses were undertaken[114;170;181].

In addition, the case studies used throughout this work represented actual research proposals submitted for funding. Rather than comparing the expected benefits from research with vague estimates of the cost of further evaluative research as it often the case in VoI analysis[123], comparisons in this study were made against specific proposals of specific design and cost.

'Payback' analysis was carried out by using the latest and most comprehensive model available in the literature—the Preliminary Assessment of Technology for Health Services (PATHS)[67]. The application of 'payback' was underpinned by the acknowledgement that, often clinical research will give raw, patient-level evidence (e.g. probability of survival at different points in time), and in order for this information to have an impact on clinical practice, this evidence needs to be translated into final endpoints. Base case analyses using the PATHS model were supplemented by a series of sensitivity analyses, which were undertaken to explore different scenarios and investigate the effect of different assumptions on the results.

VoI analysis undertaken as part of this work involved all the concepts which consist the framework, including measures of the value of perfect information (EVPI and EVPPI), sample information (EVSI and ENBS) and implementation. Undertaking these analyses is seen as a particular strength, as they offered an insight into the advantages and challenges associated with each VoI concept separately, and enabled a comprehensive assessment of the framework as a whole.

Last but not least, this assessment looks at the frameworks of interest from different viewpoints, and discusses their strengths, weaknesses and potential value with regards to a number of relevant aspects, related to their theoretical validity, methodological strengths and weaknesses, ease of use, sensitivity to assumptions and usefulness within priority-setting. In addition, this thesis provides specific recommendations on the use and role of analytic approaches in research priority-setting, and gives suggestions around different aspects of introducing and employing these approaches in practice.

Despite the above, the study has certain limitations. First, it is possible that the literature review conducted to identify approaches to priority-setting may have missed existing analytic models. This is likely to be the case if there exist analytic models which were either developed for specific research funding organisations and were never reported or discussed in journal articles, or if developed models were published in languages other than English. Comparing the models identified in the present project with those in existing reviews[55;67;80] showed that all the models in these reviews have been also included in this thesis.

Second, for pragmatic reasons, the analyses were carried out after the decision about funding the BTOG-2 and TRAPEZE phase III trials were made. To account for this, care was taken so that the analyses related to the specific points in time when the funding decisions were considered, by using information that was available prior to the funding decisions. Although evidence for parameters such as costs and effectiveness existed in the pre-trial literature, other inputs for the analysis (for example, uptake of Gem+Cisp and Gem+Carb in clinical practice prior to 2004) were scarce. As a result, such information was based on

expert opinion, rather than empirical data, which has inevitably introduced additional uncertainty in the analysis.

Third, both the case studies used in this analysis were concerned with 'clinical interventions', in the form of cancer medications. Thus, it is possible that using 'payback' and VoI on research proposals for other forms of treatment (e.g. public health interventions, screening programmes etc.) might have revealed additional strengths and limitations of the frameworks. This may be particularly the case for public health interventions, which present a series of distinctive characteristics as compared to drugs, devices or medical procedures[298].

Public health interventions are, for instance, different to clinical interventions in that the benefits from such programmes are often measured in outcomes other than QALYs[298]. While there exist indications of the decision maker's willingness to pay for a QALY[146], it is more difficult to speculate on the possible willingness to pay for an outcome expressed in natural units (e.g. a unit reduction in body mass index). As calculations of NMBs require an indication of the willingness to pay for a unit of benefit, estimating the expected benefits from research on public health interventions and translating these into NMBs would be inevitably challenging. Further, benefits from public health interventions typically accrue not only to the health care sector, but also to other areas of activity and economic sectors. Nonetheless, neither prospective 'payback' nor VoI is designed to capture and measure potential gains such as, for example, improved educational attainment arising from improved health as a result of a public health intervention. Notwithstanding these particularities, 'payback' and VoI have been in the past applied to case studies representing different types of health care programmes (for example community-based midwifery

postnatal support[67] and screening for macular degeneration[110]), and it is thought that the main conclusions with regards to the frameworks' value and potentials hold irrespective of the case studies and treatments under assessment.

An additional limitation relates to the specification of time horizon over which the information provided by the proposed trials is expected to be useful. In the present study, time horizons were based on expert opinion (Professors L. Billingham and N. James), who were asked to speculate on the number of years before the treatments for NSCLC and HRPC may become obsolete. Although such an approach is in agreement with the way time horizons have been typically set in the literature[299], it is, in essence, opaque and arbitrary. A possible way of obtaining an estimate of an intervention's expected time horizon could combine expert opinion with historic data on the average 'lifetime' of pharmaceuticals and indications of the rate at which technological innovations are expected to emerge[300]. Although there is no guarantee that such an approach would necessarily provide more accurate estimates of the required time horizon, it would nevertheless look into this issue in a more systematic and objective way.

Last, as explained in previous chapter, conducting VoI and 'payback' analyses requires, as a starting point, existing evidence about the decision problem of interest to have been identified and summarised. Where appropriate, this may involve techniques such as meta-analysis, as well as indirect and mixed treatment comparisons[114;301]. The latter methods use evidence from trials that form networks where treatments of interest may be compared indirectly through comparisons against other treatments. Indirect and mixed treatment

comparisons are particularly useful in situations where no head-to-head comparisons of treatments are available[302].

In the context of the NSCLC case study, only one trial[208] reporting a head-to-head comparison for NSCLC (i.e. Gem+Cisp vs. Gem+Carb) was identified; nonetheless there exist studies which compared Gem+Cisp or Gem+Carb against other treatments (including etoposide, mitomycin, vinblastine, vinorelbine and paclitaxel). In principle, indirect treatment comparisons could have been undertaken to account for any additional information that could have been drawn out from the existing literature. However, this type of analyses are appropriate in situations where there is certainty that specific conditions, such as homogeneity in the populations, similarity in disease severity and treatment schedules, hold across studies[302;303]. In addition, due to the fact that the role of the modelling exercise in this study is to provide a vehicle for the application of the analytic approaches rather than for actual decision making, these extensive analyses were felt to be beyond the purposes of this study.

9.6. Contribution of this work

As explained in Chapter 2, the existing literature on analytic frameworks for priority-setting is restricted; the only study[116] which undertook an empirical application of VoI and 'payback' with a view to discussing their potentials and limitations has looked at the methods only partially, and therefore its discussion and conclusions are inevitably limited. Thus, a need has been identified for an exploration that builds on and extends the available literature, by providing an in-depth discussion around the value and potential role of analytic approaches

in research priority setting. To this end, the present thesis contributes to the literature in the following ways.

First, the study explored all the concepts consisting the analytic frameworks of interest, by undertaking a 'hands-on' application of basic concepts relating to the expected value of perfect information (EVPI and EVPPI)[97], extensions of these concepts concerned with the value of sample information (EVPI and ENBS)[115;304] as well as recently suggested concepts related to implementation[187]. Observations made throughout these empirical applications gave an insight into the feasibility, strengths and weaknesses of these concepts, and enabled a more complete assessment of the frameworks. This complements and extends the work carried out by Fleurence[116], where VoI was represented only by EVPI.

On the basis of these observations, this work assessed aspects deemed relevant to potential users, including their feasibility and ease of use, robustness and reliance on assumptions, theoretical and methodological soundness, ability to inform funding decisions, and potential to fit into the existing priority-setting framework. With this in mind, the study offered specific recommendations around the role of the frameworks and their results in assisting research funding decisions, and made suggestions on practical aspects of incorporating and using the frameworks within the existing research-funding framework.

In addition, it is thought that the current work has made the following methodological contributions. First, it highlighted a way of incorporating and using probabilistic analysis for hypothesised uptake rates and likelihood weights in 'payback' calculations. Although this addition does not by any means represent a significant 'breakthrough' in the methodology, it

is thought to be a useful improvement which offers a more explicit representation of the uncertainty in key parameters entering the analysis. Secondly, the present work highlighted a way of extending 'payback' to assess proposals for multiple-treatment comparisons, and demonstrated its use by applying it on the four-treatment comparison TRAPEZE case study.

With regards to VoI, this study has illustrated the application of value of implementation concepts under the more realistic assumption that acquisition of information about treatments' cost-effectiveness has a direct effect on their uptake in clinical practice. As Fenwick *et al*.[187] explain, such an extension is essential in enabling the concept of value of implementation to represent 'real world' situations in a more accurate and realistic way. Moreover, the analysis indicated a relevant measure of the value of implementation, in the form of the 'expected value of information-implementation trade-off'. The measure compares the expected value of acquiring further information against the value of investing in implementation, and, in principle, it can indicate the situations where pursuing treatment implementation may be preferable.

Last, the work demonstrated a simple and effective way of conducting expected value of sample information analysis. A common difficulty in conducting EVSI analysis relates to combining prior and sample (simulated) information when such information is expressed in terms of dissimilar distributions which are non-conjugate and cannot be combined using simple analytic solutions. In this work, this problem was evaded by expressing the simulated sample information in the same units as the prior information (i.e. counts of patient transitions to difference health states in HRPC and proportions of patients in specific health states at different points in time in NSCLC). This was achieved by following up individuals in a

hypothetical trial, which mimicked the proposed trial, one at a time and simulating their patient histories. As this work was carried out in a widely-available spreadsheet application, it is believed that it contributes to making the mechanics of conducting EVSI analysis more transparent and accessible to a wider base of researchers.

All in all, it is thought that this work helps to resolve uncertainties around the strengths and weakness of analytic approaches, sheds light on different aspects related to their potential role and value in research priority-setting, makes policy recommendations around their use and offers suggestions on practical aspects of introducing and using analytic methods to inform research funding decisions.

9.7. Recommendations for further research

This study identified a number of areas where further research would be valuable. With regards to 'payback', further methodological research would be needed to establish more explicit and systematic ways of determining possible research outcomes (also called 'Delta results'[90] or exemplar outcomes[92]), given the fact that such outcomes have an important impact on the final results. Additional research would also be needed to look into appropriate ways of obtaining robust estimates for other key uncertain parameters entering the analysis, most importantly a treatment's future uptake in clinical practice and the likelihood of a proposed trial showing the specified results. For the former, this may involve formal methods of eliciting expert opinion from adequately large groups of researchers and decision-makers. For the latter, there may be scope for obtaining likelihood weights by combining expert opinion with existing evidence (e.g. existing results of other studies, phase

II data), possibly by using Bayesian processes[305;306]. Obtaining better estimates of such parameters would allow more confidence in 'payback' results and would strengthen the case for the use of the framework. Moreover, it is expected that further research to ascertain the most appropriate methodology for applying 'payback' for multiple treatment comparisons would be highly useful.

With relation to VoI, it is thought that further research would be needed around EVSI, with a view to making this analysis more practical and applicable to a wider range of situations. In addition, there is a need for further development in the concepts of value of implementation. If the aim of this framework is to address the fact that, without taking into account imperfect implementation, the predicted benefits from research and decision making will be overestimated, the framework should be expanded. A first useful contribution would be to further develop the methodology so that it accounts for the realistic situation that acquisition of information about a treatment would be expected to have a direct impact on the treatment's implementation. Indications of how this may be done have been given in this work.

The importance of obtaining valid estimates of the expected lifetime of the produced information has been highlighted in a number of instances throughout this thesis. Although they are crucial factors in determining the value of research, time horizons are currently determined in a way that is, in essence, ambiguous. Thus, there would be important benefits from developing a systematic algorithm or process for estimating the time horizon of information. As pointed out earlier, such an algorithm may need to take into account

information from expert opinion, observations on the lifetime of treatments and projections of the trajectory of technological innovations.

Finally, it is thought that a stream of research, workshops and consultations would be needed to determine the most appropriate ways of introducing analytic methods into existing research funding processes. It is envisaged that potential users of the frameworks will need to establish ways for incorporating analytic methods which are both efficient and acceptable to decision makers. As a first step, this may involve looking into process-related issues that are specific to funding organisations, such as how the steps they currently follow may be modified to incorporate analytic approaches, and whether and to what extent current timelines should be extended to account for the additional analyses.

Equally importantly, funding organisations will have to address more fundamental issues around how 'payback' and 'value of information' results would be combined with other criteria considered important to the organisation, and what 'weight' should be attached to suggestions and indications obtained through analytic methods.

Appendices

Appendix 1. Literature review on priority-setting for research

Appendix 1.A. Search strategies for identifications of literature on priority-setting for research

1.A.A Search strategies used in MEDLINE and EMBASE

Strategies 1.1 to 1.5 below were used for searches in MEDLINE and EMBASE (Ovid interface; 1946 to January Week 1 2010).

Search strategy 1.1

1. "priority-setting".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
2. (assess$ adj3 priorit$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
3. (research adj3 priorit$ adj3 set$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
4. (prioritisation or prioritization).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
5. Research/cl, ec, mt, sn [Classification, Economics, Methods, Statistics & Numerical Data]
6. priorities, research.mp.
7. (priorit$ adj4 research).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
8. (research adj3 resourc$ adj3 allocat$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
9. (research adj3 fund$ adj3 allocat$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
10. *Research Support as Topic/

11. research fund$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
12. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
13. (health adj3 technology adj3 assessment).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
14. (evaluative adj3 research).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
15. (assessment adj3 health adj5 technolog$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
16. *Technology Assessment, Biomedical/cl, ec, mt, og, st, sn, td [Classification, Economics, Methods, Organization & Administration, Standards, Statistics & Numerical Data, Trends]
17. Evaluation Studies/
18. Research Support as Topic/
19. clinical trial/ or clinical trial, phase i/ or clinical trial, phase ii/ or clinical trial, phase iii/ or clinical trial, phase iv/ or controlled clinical trial/ or multicenter study/ or randomized controlled trial/
20. 13 or 14 or 15 or 16 or 17 or 18 or 19
21. Research Support as Topic/cl, ec, mt, og, sn [Classification, Economics, Methods, Organization & Administration, Statistics & Numerical Data]
22. ((fund$ or financ$ or support$ or cost$ or subsid$ or budget$) and (organi$ation or societ$ or body or bodies or instit$)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
23. NIHR.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
24. (research adj3 fund$ adj3 (organi$ation or institut$ or body or bodies or society or charit$)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
25. 21 or 22 or 23 or 24
26. 12 and 20 and 25
27. ((health technol$ assess$ or research or clinical trial$) and priorit$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
28. 26 and 27

Search strategy 1.2

1. "priority-setting".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
2. (assess$ adj3 priorit$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
3. (research adj3 priorit$ adj3 set$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
4. (prioritisation or prioritization).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
5. Research/cl, ec, mt, sn [Classification, Economics, Methods, Statistics & Numerical Data]
6. priorities, research.mp.
7. (priorit$ adj4 research).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
8. (research adj3 resourc$ adj3 allocat$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
9. (research adj3 fund$ adj3 allocat$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
10. *Research Support as Topic/
11. research fund$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
12. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
13. (health adj3 technology adj3 assessment).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
14. (evaluative adj3 research).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
15. (assessment adj3 health adj5 technolog$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
16. *Technology Assessment, Biomedical/cl, ec, mt, og, st, sn, td [Classification, Economics, Methods, Organization & Administration, Standards, Statistics & Numerical Data, Trends]
17. Evaluation Studies/
18. Research Support as Topic/

19. clinical trial/ or clinical trial, phase i/ or clinical trial, phase ii/ or clinical trial, phase iii/ or clinical trial, phase iv/ or controlled clinical trial/ or multicenter study/ or randomized controlled trial/
20. 13 or 14 or 15 or 16 or 17 or 18 or 19
21. Research Support as Topic/cl, ec, mt, og, sn [Classification, Economics, Methods, Organization & Administration, Statistics & Numerical Data]
22. ((fund$ or financ$ or support$ or cost$ or subsid$ or budget$) and (organi$ation or societ$ or body or bodies or instit$)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
23. NIHR.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
24. (research adj3 fund$ adj3 (organi$ation or institut$ or body or bodies or society or charit$)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
25. 21 or 22 or 23 or 24
26. 12 and 20 and 25
27. Health Priorities/
28. 26 and 27

Search strategy 1.3

1. "priority-setting".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
2. "research priorities".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
3. (assess$ adj3 priorit$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
4. (research adj3 priorit$ adj3 set$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
5. "priority-setting".mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
6. (prioritisation or prioritization).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
7. research/cl, ec, mt, sn [Classification, Economics, Methods, Statistics & Numerical Data]

8. priorities, research.mp.
9. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10. comparative Effectiveness Research/
11. evidence-Based Medicine/ or exp Technology Assessment, Biomedical/
12. behavioral research/ or biomedical research/ or empirical research/ or peer review, research/ or research design/
13. (health adj3 technolog$ adj3 assess$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
14. ((economic or effectiveness) adj3 (evaluat$ or assess$)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
15. drug Approval/ or Device Approval/
16. "costs and Cost Analysis"/ or Cost-Benefit Analysis/
17. 10 or 11 or 12 or 13 or 14 or 15 or 16
18. 9 and 17
19. (priorit$ adj4 research).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
20. 18 and 19

Search strategy 1.4

1. exp Health Priorities/cl, ec, og, st, td [Classification, Economics, Organization & Administration, Standards, Trends]
2. exp Technology Assessment, Biomedical/cl, ec, og, st, td, ut [Classification, Economics, Organization & Administration, Standards, Trends, Utilization]
3. (health adj3 technolog$ adj3 assess$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
4. 2 and 3
5. 1 and 4

Search strategy 1.5

6. exp Technology Assessment, Biomedical/cl, ec, og, st, td, ut [Classification, Economics, Organization & Administration, Standards, Trends, Utilization]

7. (health adj3 technolog$ adj3 assess$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
8. exp Research Support as Topic/cl, ec, mt, og, st, td [Classification, Economics, Methods, Organization & Administration, Standards, Trends]
9. (research adj3 resourc$ adj3 (fund$ or allocat$)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
10. (research adj3 (priorit$ or priority-sett$ or priority sett$)).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
11. 1 or 2
12. 3 or 4 or 5
13. 6 and 7

1.A.B Search strategies used in Centre for Reviews and Dissemination (CRD) databases

Strategies 2.1 and 2.2 were used for searches in the Database of Abstracts of Reviews of Effects (DARE), NHS Economic Evaluation Database (NHS EED) and Health Technology Assessment (HTA) databases (Ovid interface, inception to January Week 1 2010).

Search strategy 2.1

1. "priorit setting*"
2. research AND priorit*
3. MeSH Research Support as Topic QUALIFIERS CL EC ST TD EXPLODE 1
4. MeSH Health Priorities QUALIFIERS CL EC OG ST TD UT EXPLODE 1 2
5. health NEAR technology NEAR assessmen*
6. MeSH Clinical Trials as Topic QUALIFIERS CL EC TD UT EXPLODE 1 2 3 4
7. MeSH Technology Assessment, Biomedical QUALIFIERS CL EC OG EXPLODE 1 2
8. "research fund*"
9. (fund* OR financ* OR support* OR cost* OR subsid* OR budget*) AND (organi*ation OR societ* OR body OR bodies OR instit*)
10. 1 OR 2 OR 3 OR 4

11. 5 OR 6 OR 7
12. 8 OR 9
13. 10 AND 11 AND 12

Search strategy 2.2

1. priorit* AND health
2. priorit* AND health AND research
3. research NEAR fund* NEAR allocat*
4. MeSH Research QUALIFIERS CL EC OG EXPLODE 1
5. 1 OR 2 OR 3
6. 4 AND 5

1.A.C Search strategy used for searches in EconLit and CINAHL

Strategies 3.1 and 3.2 were used for searches in EconLit and CINAHL databases (inception to January Week 2 2010) respectively, through the EBSCO portal.

Search strategy 3.1

1. research AND priorit*
2. fund* OR finance* OR allocate* OR cost OR assess*
3. research fund*
4. health N3 technology N3 assessment
5. biomedical research
6. clinical trial*
7. research support
8. evaluat* N3 research
9. research N3 fund* N5 decision*
10. National institute for Health Research OR NIHR

11. Health N Technology N Assessment OR HTA
12. Medical Research Council OR MRC
13. research AND (council OR organi?ation OR body OR committee OR institute)
14. health
15. S1 or S2 or S3
16. S4 or S5 or S6 or S7 or S8
17. S9 or S10 or S11 or S12 or S13
18. S14 and S15 and S16 and S17
19. research AND priorit*
20. fund* OR finance* OR allocate* OR cost OR assess*
21. research fund*
22. health N3 technology N3 assessment
23. biomedical research
24. clinical trial*
25. research support
26. evaluat* N3 research
27. research N3 fund* N5 decision*
28. national institute for Health Research OR NIHR
29. Health N Technology N Assessment OR HTA
30. medical Research Council OR MRC
31. research AND (council OR organi?ation OR body OR committee OR institute)
32. health
33. S19 or S20 or S21
34. S22 or S23 or S24 or S25 or S26
35. S27 or S28 or S29 or S30 or S31
36. S32 and S33 and S34 and S35

Search strategy 3.2

1. research AND priorit*
2. fund* OR finance* OR allocate* OR cost OR assess*

3. research fund*
4. health N3 technology N3 assessment
5. biomedical research
6. clinical trial*
7. research support
8. evaluat* N3 research
9. research N3 fund* N5 decision* National institute for Health Research OR NIHR
10. Health N Technology N Assessment OR HTA
11. Medical Research Council OR MRC
12. Research AND (council OR organi?ation OR body OR committee OR institute) (27889)
13. health
14. S1 or S2 or S3
15. S4 or S5 or S6 or S7 or S8
16. S9 or S10 or S11 or S12 or S13
17. S14 and S15 and S16 and S17
18. research funding OR research priority-setting OR reserach priorities OR research funding decisions
19. S18 and S19

Table 1.a: Articles identified through searches in bibliographic databases

Searches in databases	MEDLINE	EMBASE	NHS CRD	EconLit	CINAHL	Total number
Search 1.1	291	315	-*	-	-	606
Search 1.2	94	25	-	-	-	119
Search 1.3	301	438	-	-	-	739
Search 1.4	4	0	-	-	-	4
Search 1.5	46	20	-	-		66
Search 2.1	-	-	33	-	-	33
Search 2.2	-	-	2	-	-	2
Search 3.1	-	-	-	51	-	51
Search 3.2	-	-	-	-	17	17
Total hits	736	798	35	51	17	1637
Duplicates within database	153	39	0	0	0	192
Unique articles within database	583	759	35	51	17	1445
Duplicates between databases	160					
Total number of unique articles	**1285**					
*Hyphens indicate that the search strategy was not applied to the specific database.						

Table 1.b: Additional articles identified through searches other than in bibliographic databases

Additional articles identified through additional searches*	
References citing key articles	1
References cited in key articles	7
References identified through 'related articles' search in MEDLINE	5
References identified through searches in key journals	1
References identified through searches on the internet (using Google Scholar® and Dogpile®).	1
References identified subsequently, while conducting other searches	1
Total number of relevant articles	**16**
* Articles identified through searches other than in bibliographic databases were assessed for inclusion upon identification. This table givens the number of included articles only.	

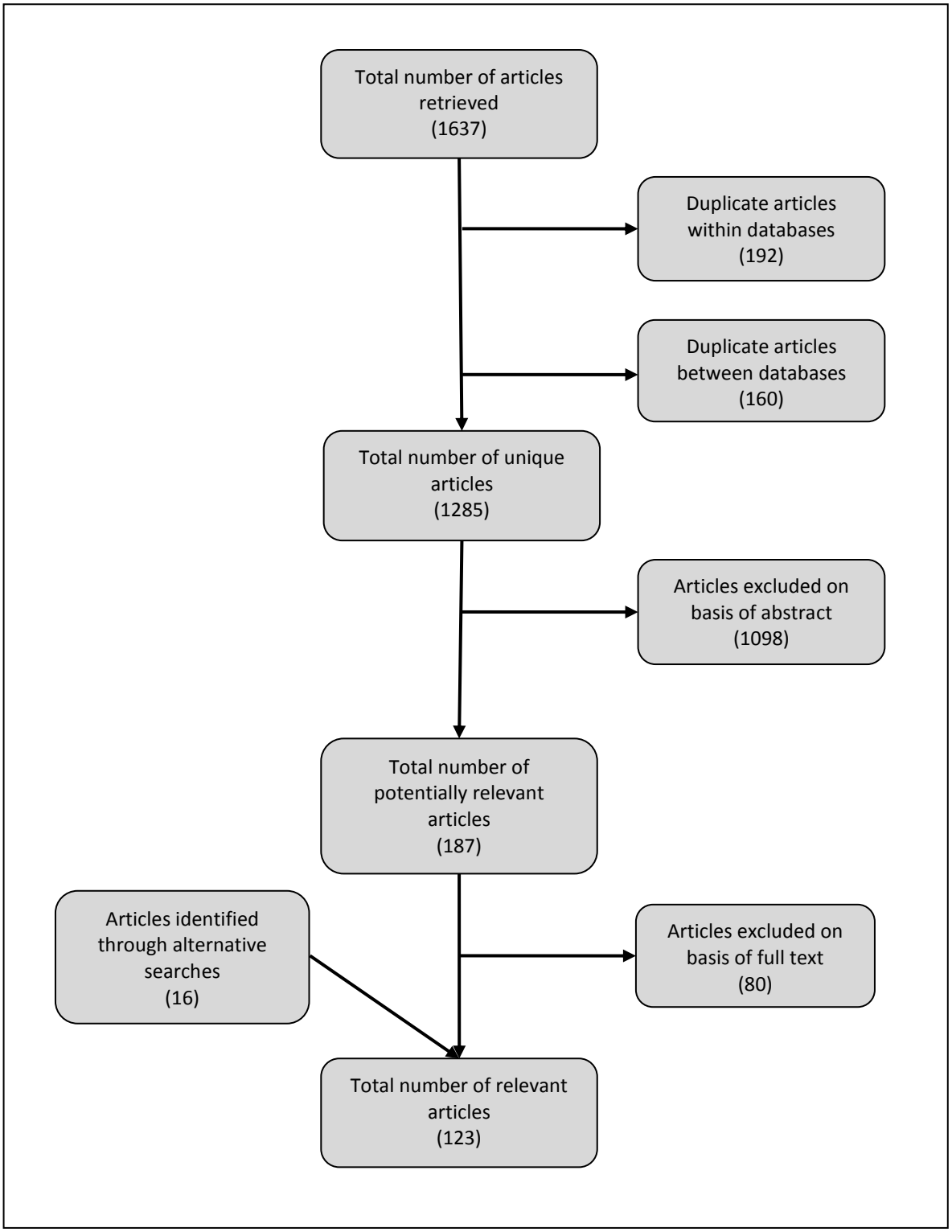

Figure 1.a: Flow chart showing the process of identifying and selecting literature on priority-setting for research

Appendix 2. Searches for literature related to 'payback of research' and 'value of information'

Appendix 2.A. Searches strategies for identification of literature relevant to 'payback of research'

2.A.A Search strategies used in electronic bibliographic databases

Strategies 1 to 3 below were used for searches in MEDLINE, EMBASE, HMIC and Books@Ovid (Ovid interface; 1946 to May Week 1 2010).

Search strategy 1

1. (Buxton M or Townsend J or Hanney S or Harper G$).m_auts.
2. (health and research).mp. [mp=tx, bt, ti, ab, sh, hw, tn, ot, dm, mf, nm, ui]
3. 1 and 2

Search strategy 2

1. payback.mp. [mp=tx, bt, ti, ab, sh, hw, tn, ot, dm, mf, nm, an, hn, ui]
2. limit 1 to abstracts [Limit not valid in Books@Ovid; records were retained]
3. (research and health).mp. [mp=tx, bt, ti, ab, sh, hw, tn, ot, dm, mf, nm, an, hn, ui]
4. 2 and 3

Search strategy 3

1. PATHS model.mp. [mp=tx, bt, ti, ot, ab, hw, sh, tn, dm, mf, nm, ui]
2. payback of research.mp. [mp=tx, bt, ti, ot, ab, hw, sh, tn, dm, mf, nm, ui]
3. (preliminary and assessment and research).mp. [mp=tx, bt, ti, ot, ab, hw, sh, tn, dm, mf, nm, ui]

4. ('cost-benefit of research' or 'cost benefit of research').mp. [mp=tx, bt, ti, ot, ab, hw, sh, tn, dm, mf, nm, ui]
5. ('cost-effectiveness of research' or 'cost effectiveness of research').mp. [mp=tx, bt, ti, ot, ab, hw, sh, tn, dm, mf, nm, ui]
6. (prospective and payback) .mp. [mp=tx, bt, ti, ot, ab, hw, sh, tn, dm, mf, nm, ui]
7. 'research priorit*'.mp. [mp=tx, bt, ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, ps, rs, an, ui]
8. 1 or 2 or 3 or 4 or 5 or 6
9. 7 and 8

2.A.B Additional searches

Additional searches involved searches for articles citing key publications, scanning the reference lists of key publications, 'related articles' searches in MEDLINE, searches in the reference lists of key journals, searches through Google Scholar® and searches on Brunel University's website (url: http://www.brunel.ac.uk/about/acad/herg/publications/payback)

Table 2.a: Number of articles related to 'payback' identified in bibliographic databases

Search strategies used in searches in databases	Number of hits in MEDLINE, EMBASE, HMIC and Books@OVID
Search strategy 1	169
Search strategy 2	74
Search strategy 3	9
Total hits	252
Duplicates	85
Total number of unique articles	**167**

Table 2.b: Articles on 'payback' identified through searches other than in bibliographic databases

Articles identified through additional searches	
References citing key articles	10
References cited in key articles	2
References identified through 'related articles' search in MEDLINE	8
References identified through searches in Google Scholar®	3
References identified subsequently, while conducting other searches	2
References identified on Brunel University website	12
Total number of relevant articles	**37**
* Articles identified through searches other than in bibliographic databases were assessed for inclusion upon identification. This table givens the number of included articles only.	

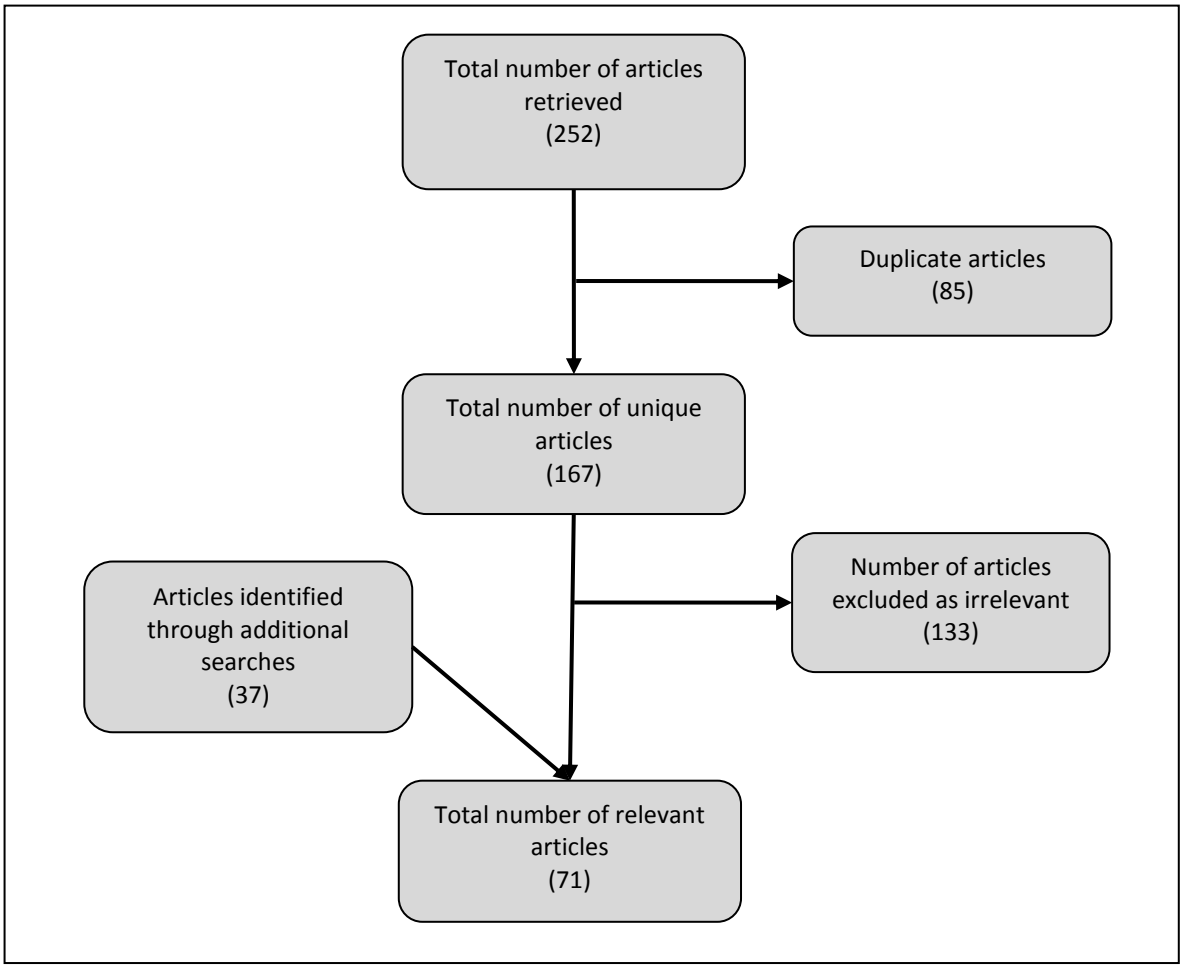

Figure 2.a: Flow chart showing the process of identifying and selecting literature related to 'payback of research'

Appendix 2.B. Summary characteristics of practical applications of PATHS

Table 2.c: Characteristics of PATHS applications in Townsend *et al*.[67]

Characteristic	Study: Postnatal midwifery support study	Study: Infusion protocol in adult pre-hospital care	Study: Small aneurysm trial	Study: -interferon for multiple sclerosis
Clinical area	Postnatal support service for new mothers	Pre-hospital protocol for severely injured adults	Optimal management of patients with small abdominal aortic aneurysm	Use of -interferon for the treatment of patients with relapsing remitting RR-MS) and secondary progressive multiple sclerosis (SP-MS).
Comparators	Postnatal support service (10 additional days) from community midwifery support workers (CMSW) Support service (10 visits within 1 month post delivery) (current practice)	Field stabilisation following severe injury Immediate transportation to hospital	Watchful waiting Elective early surgery	-interferon Standard treatments

Characteristic	Study: Postnatal midwifery support study	Study: Infusion protocol in adult pre-hospital care	Study: Small aneurysm trial	Study: -interferon for multiple sclerosis
Exemplar outcomes	Outcome A: CMSW service is effective and cost-effective (25-point improvement in the GHP profile of the SF-36 Outcome B: CMSW is associated with non-significant improvement in GHP between intervention and control Outcome C: CMWS service is less effective that current practice	Outcome A: Immediate transportation to hospital more effective than comparator Outcome B: No concluding evidence on effectiveness of interventions Outcome C: Field stabilisation more effective than comparator	Outcome A: Early surgery is effective and cost-effective Outcome B: Surveillance associated with higher quality of life Outcome C: No significant difference between elective surgery and watchful waiting exists	Outcome A: -interferon does not offer net clinical benefits and it is not cost-effective Outcome B: -interferon offers net clinical benefits but it is not cost-effective Outcome C: -interferon offers net clinical benefits and it is cost-effective
Assumed change in clinical practice	Following outcome A, provision of CMSW increases to 25% Following Outcome B, service will be provided to 5% of the eligible population Following Outcome C, service is provided to only 2.5% of eligible population	Following outcome A, 90% of the eligible population is immediately transported to hospital Following outcome B, up to 70% are immediately transported to hospital Following outcome C, 90% of the patients receive field stabilisation	Following outcome A, early surgery rates increases to 95% of the eligible population Following outcome B, early surgery rates drop to 85% Following outcome C, early surgery rates drop to 50% of the eligible population	Following Outcome A, -interferon is prescribed to a very limited number of patients Following Outcome B, -interferon is prescribed to a limited number of patients on the (decision made on basis of cost-effectiveness), or no restrictions apply on prescription of -interferon (decision made on basis of clinical benefits) Following Outcome C, -interferon is prescribed to a greater number of patients.

Characteristic	Study: Postnatal midwifery support study	Study: Infusion protocol in adult pre-hospital care	Study: Small aneurysm trial	Study: -interferon for multiple sclerosis
Results	Expected costs and benefits of research: between £2 ('pessimistic' combination) and £3.50 ('optimistic' combination) per point improvement in General Function Scale of SF-36. EICER deemed bellow the expert decision-makers' threshold for adoption (£150/25-point change) and thus the study appears 'good value for money'	The 'negative' and 'positive' scenarios give EICERs between £2,500 and £4,000 per life saved (for a 40 years old patient). According to the authors, this is low compared with nearly all the current treatment costs and so would be cost-effective.	The authors consider the study marginally cost-effective, at £175, 000 (£20,000 per life-year saved).	Most likely scenario gave cost per QALY of £1 million. Authors concluded the study is highly unlikely to prove cost-effective.
What is the decision rule?	Comparison between EICER and local decision-makers' threshold of accepting the technology The particular trial appears to be cost-effective	Comparison between EICER and local decision-makers' threshold of accepting the technology The particular trial appears to be cost-effective	Comparison between EICER and local decision-makers' threshold of accepting the technology The particular trial appears to be marginally cost-effective	Comparison between EICER and local decision-makers' threshold of accepting the technology The trial is highly unlikely to be cost-effective

Table 2.d: Characteristics of PATHS applications in Fleurence[116]

Characteristic	Study: RECORD Trial	Study: Vitamin D and Calcium Trial	Study: Hip Protector Trial	Study: Pressure Trial
Clinical area	Preventive treatment against osteoporotic fractures in elderly people	Preventive treatment against osteoporotic fractures in elderly people	Prevention of hip fractures in elderly people	Prevention and treatment of pressure ulcers

Characteristic	Study: RECORD Trial	Study: Vitamin D and Calcium Trial	Study: Hip Protector Trial	Study: Pressure Trial
Comparators	Vitamin D and Calcium (VDC) Vitamin D (VD) Calcium (C) Placebo	Vitamin D and Calcium tablet daily (VDC) Information leaflet	Hip protectors (HP) No treatment	Alternating pressure overlay (AO) Alternating pressure replacement mattress (AR) Standard care (high-specification foam mattresses)
Exemplar outcomes	Outcome A ('positive'): The trial shows that VDC, VD and C are effective and cost-effective in preventing fractures Outcome B ('inconclusive'): The trial shows no difference between VDC, C and VD and placebo in preventing fractures Outcome C ('negative'): The trial shows that VDC, C and VD are less effective than placebo in preventing fractures	Outcome A ('positive'): The trails shows that VDC is effective and cost-effective for the prevention of fractures Outcome B ('inconclusive'): The trial shows no difference between VDC and no treatment for the prevention of fractures Outcome C ('negative'): The trial shows that VDC is less effective than no treatment for the prevention of fractures.	Outcome A ('positive'): The trial shows that HP is effective and cost-effective for the prevention of fractures. Outcome B ('inconclusive'): The trial shows no difference between HP and no treatment for the prevention of fractures Outcome C ('negative'): The trial shows that HP is less effective than no treatment for the prevention of fractures.	Outcome A ('positive'): The trial shows that AO and AR are effective and cost-effective for the prevention and treatment of pressure ulcers Outcome B ('inconclusive'): The trial shows no difference between AO and AR for prevention and treatment of pressure ulcers. Outcome C ('negative'): The trial shows that AO and AR are less effective than standard care for the prevention and treatment of pressure ulcers

Characteristic	Study: RECORD Trial	Study: Vitamin D and Calcium Trial	Study: Hip Protector Trial	Study: Pressure Trial
Assumed change in clinical practice	Following outcome A, implementation of VDC, C and VD extends to 20%, 50% and 5% of patients at risk, accordingly. Following outcome B, current practice continues as it is. VDC, VD and C are given to 30%, 15% and 5% of the population at risk, respectively. Following outcome C, VDC, VD and C are given to 5%, 5% and 5% of the population at risk, respectively.	Following outcome A, implementation of VDC extends to 75% of patients at risk Following outcome B, current practice continues, with VDC provided to 30% of the population at risk Following outcome C, VCD is given to 5% of the population at risk.	Following outcome A, implementation of HP extends to 50% of the population at risk Following outcome B, current practice continues as it is. HP is provided to 1% of the population at risk Following outcome C, HP is not provided to the population at risk.	Following outcome A, implementation extends to 5%, 10% and 85% of the population for AO, AR and standard care, respectively, in the first year and 15%, 20% and 65% in the subsequent years Following outcome B, the rate of use is 5%, 10% and 85% for AO, AR and standard care, respectively. Following outcome C, the first year rate of the treatments' use is 5%, 10% and 85% for AO, AR and standard care, respectively. In years 2 and 3 use changes to 1%, 1% and 98% for AO, AR and standard care respectively. After this, AO and AR are no longer used and the use of standard care reaches 100%.

Characteristic	Study: RECORD Trial	Study: Vitamin D and Calcium Trial	Study: Hip Protector Trial	Study: Pressure Trial
Results	Results were presented in the form of Net Monetary Benefits using a decision-maker's ceiling ratio of £30,000 per QALY.			
	Expected NMBs of research were £186.19 million for an 'optimistic' combination, £145.8 million for a 'neutral' combination and £145.52 million for a 'pessimistic' combination.	Expected NMBs of research were £133.39 million for an 'optimistic' combination, £108.04 million for an 'inconclusive' combination and £112.19 million for a 'pessimistic' combination.	Expected NMBs of research were £7.74 million for an 'optimistic' combination, £48.42 million for a 'neutral' combination and £24.6 million for a 'pessimistic' combination.	Expected NMBs were £13.81 million for an 'optimistic' combination, -£62.76 million for a 'neutral' combination and -£24.6 million for a 'pessimistic' combination.
What is the employed decision rule? **Is the research proposal recommended for funding?**	Positive expected Net Monetary Benefits. The particular trial appears to be cost-effective.	Positive expected Net Monetary Benefits. The particular trial appears to be cost-effective	Positive expected Net Monetary Benefits. Positive NMB suggest the trial appears to be cost-effective.	Positive expected Net Monetary Benefits. The trial is cost-effective only under the 'optimistic' combination.

Appendix 2.C. Search strategies for identification of literature relevant to 'value of information'

2.C.A Searchers used in electronic bibliographic databases

Strategies 1 to 3 below were used for searches in MEDLINE, EMBASE, HMIC and Books@Ovid (Ovid interface; 1946 to May Week 2 2010).

Search strategy 1

1. (claxton or claxton k or claxton kp).au.

Search strategy 2

1. ("value of information" or "expected value of information" or "expected value of perfect information" or "expected net benefit of sampling" or "expected value of perfect information for parameters" or "expected value of implementation").mp. [mp=tx, bt, ti, ab, sh, hw, tn, ot, dm, mf, nm, an, hn, ui]
2. (health and research).mp. [mp=tx, bt, ti, ab, sh, hw, tn, ot, dm, mf, nm, an, hn, ui]
3. 1 and 2

Search strategy 3

1. (("EVPI" or "VoI" or "EVSI" or "EVPPI") and information and value and cost-effectiveness).mp. [mp=tx, bt, ti, ab, sh, hw, tn, ot, dm, mf, nm, an, hn, ui] (46)

2.C.B Additional searches

Additional searches were carried out for articles citing key publications, articles in the reference lists of key publications, 'related articles' identified through MEDLINE, reference lists of key journals, as well as free searches through Google Scholar®.

Table 2.e: Number of articles related to VoI identified in bibliographic databases

Search strategies used in searches in databases	Number of hits in MEDLINE, EMBASE, HMIC and Books@OVID
Search strategy 1	130
Search strategy 2	249
Search strategy 3	46
Total hits	425
Duplicates	201
Total number of unique articles	224

Table 2.f: Articles related to VoI identified through searches other than in bibliographic databases

Articles identified through additional searches	
References citing key articles	5
References cited in key articles	2
References identified through 'related articles' search in MEDLINE	3
References identified through searches in Google Scholar®	0
References identified subsequently, while conducting other searches	0
References identified on University of York website	3
Total number of relevant articles	**13**
* Articles identified through searches other than in bibliographic databases were assessed for inclusion upon identification. This table givens the number of included articles only.	

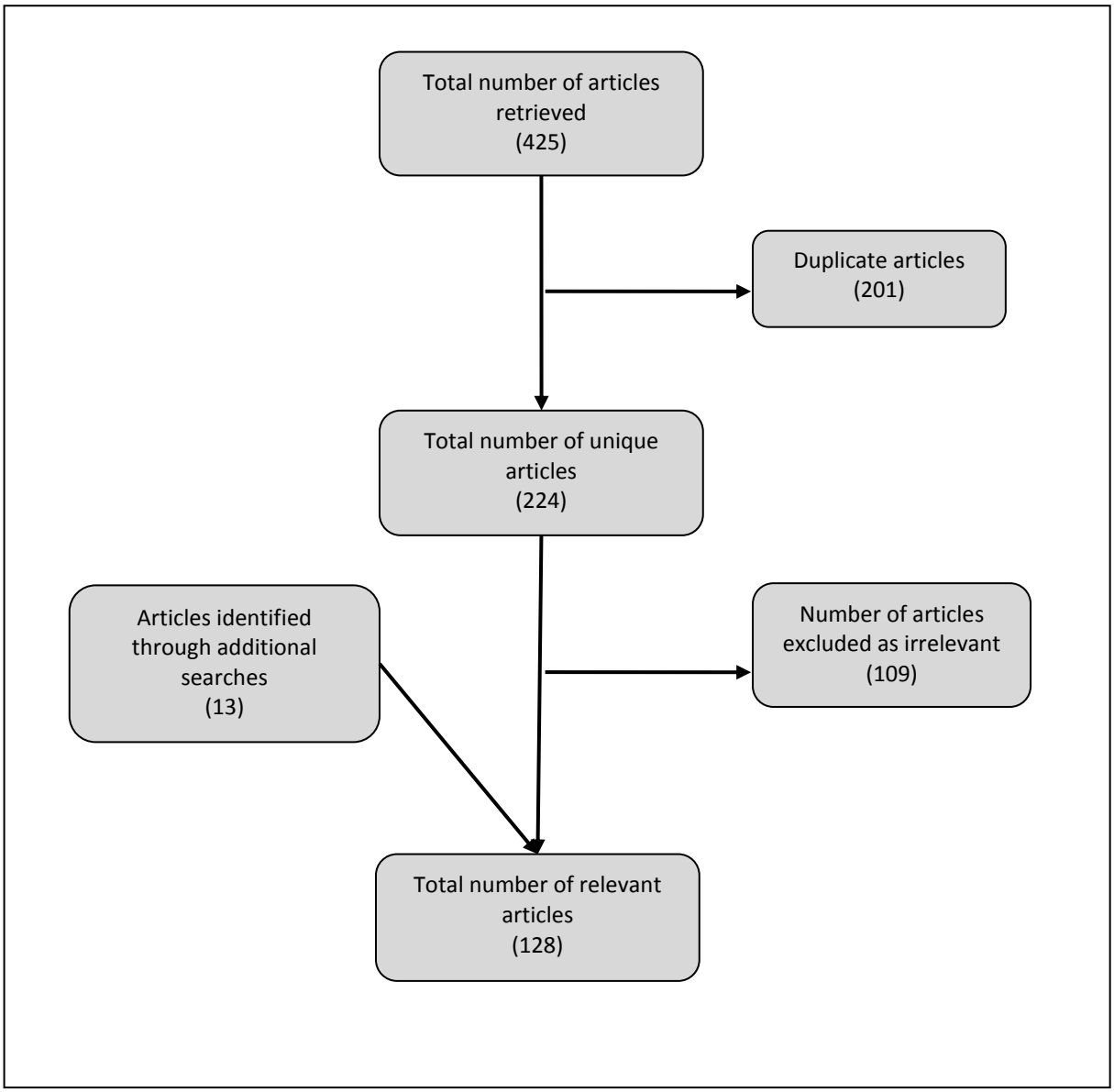

Figure 2.b: Flow chart showing the process of identifying and selecting literature related to 'value of information'.

Appendix 3. Literature review and parameters for NSCLC

Appendices 3.A. and 3.B. provide information on the literature review for evidence on the effectiveness, costs and cost-effectiveness of chemotherapy treatments for NSCLC. Appendices 3.C. and 3.D. provide information related to the NSCLC decision model.

Appendix 3.A. Search strategies for evidence on NSCLC

Search strategies 1 to 3 below were used for searches in MEDLINE (PubMed interface; 1946 to September Week 2 2011).

Search strategy 1

1. ((("Carcinoma, Non-Small-Cell Lung/drug therapy"[Mesh] OR "Carcinoma, Non-Small-Cell Lung/economics"[Mesh] OR "Carcinoma, Non-Small-Cell Lung/epidemiology"[Mesh]))) AND "Cisplatin"[Mesh] AND "Carboplatin"[Mesh] AND "gemcitabine "[Substance Name]

Search strategy 2

1. "Carcinoma, Non-Small-Cell Lung"[Mesh] AND "Cost-Benefit Analysis"[Mesh]

Search strategy 3

1. non-small AND cell AND lung AND cancer
2. metastatic AND lung AND cancer
3. cisplatin OR carboplatin OR gemcitabine
4. "cost-effectiveness" OR "cost effectiveness" OR "model"
5. #1 OR #2

6. #3 AND #4 AND #5

Search strategy 4

Search strategy 4 involved searches in published NICE clinical guidelines on non-small cell lung cancer, searches for reviews in the Cochrane Library, 'related articles' searches in PubMed and ISI Web of Science, searches in the reference lists of identified articles, and general searches through Google Scholar®.

Table 3.a: Identified articles on NSCLC by search strategy

Searches in databases	**MEDLINE**	**General searches** (NICE; Cochrane Library; CRD; ISI Web of Science; Google Scholar®; 'related article' searches; reference lists of identified articles)
Search strategy 1	84	-
Search strategy 2	157	-
Search strategy 3	65	-
Search strategy 4	-	26
Total hits	306	26
Duplicate artic**les across searches**	34	0
Unique articles within databases	272	26
Duplicate articles between databases	12	
Total number of unique articles	**286**	
Note: dashes indicate that the particular search strategy was not applied to the specific database.		

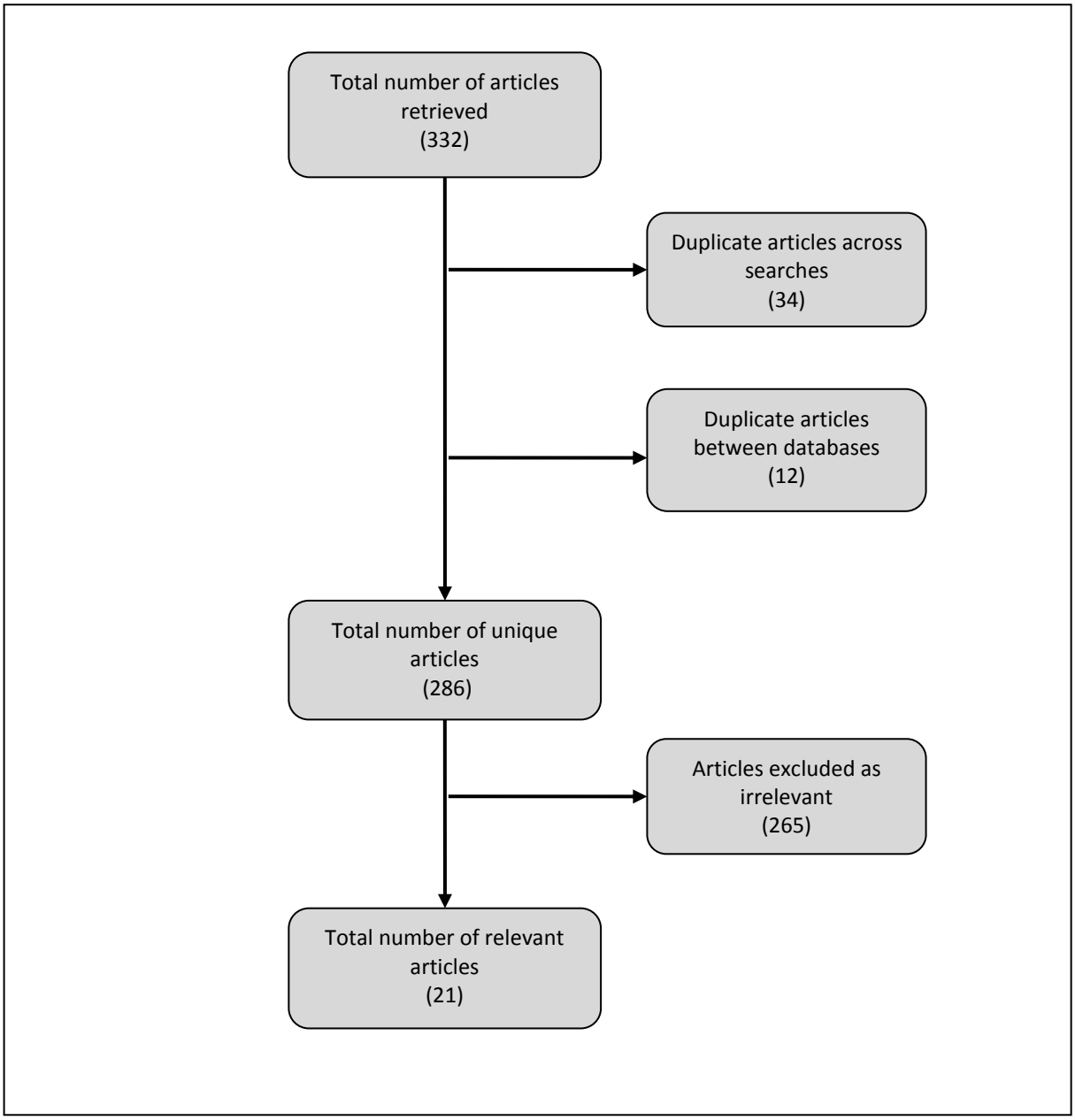

Figure 3.a: Flow chart showing the process of identifying and selecting studies on chemotherapy for NSCLC

Appendix 3.B. Summary of identified studies in NSCLC

Table 3.b: Summary of identified clinical effectiveness studies (NSCLC)

Reference	Study type	Intervention	Comparison	Outcome measures	Authors' conclusion
Cardenal eta l(1999)[226]	RCT	Gemcitabine + cisplatin	Etoposide + cisplatin	Response Survival Toxicity Time-to-progression QoL (EORTC QLQ-C30-LC13)	Gemcitabine + cisplatin showed superior response and superior (longer) time-to-progression No significant difference in survival No significant difference in QoL between treatments The treatments showed similar toxicity
Crino *et al.* (1999)[227]	RCT	Gemcitabine + cisplatin	Mitomycin + ifosfamide + cisplatin	Response Survival Toxicity Time-to-progression QoL (EORTC QLQ-C30-LC13)	No significant difference in survival and time-to-progression between treatments Gemcitabine + cisplatin showed superior (higher) response rate Similar toxicity between treatments No significant difference in QoL between treatments

Reference	Study type	Intervention	Comparison	Outcome measures	Authors' conclusion
Comella *et al.* (2000)	RCT	Gemcitabine + cisplatin + vinorelbine	Gemcitabine + cisplatin Vinorelbine + cisplatin	Response Survival Toxicity Time-to-progression QoL (EORTC QLQ-C30-LC13)	Gemcitabine + cisplatin + vinorelbine showed superior response and survival Similar toxicity between treatments
Sandler *et al.* (2000)[307]	RCT	Gemcitabine + cisplatin	Cisplatin	Response Survival Toxicity Time-to-progression QoL (FACT-L)	Gemcitabine + cisplatin showed superior response and survival and time-to-progression Similar toxicity between treatments Similar QoL between treatments
Clegg *et al.* (2001)[201]	Systematic review of RCTs	Gemcitabine+ cisplatin	Various doses and treatment schedules of: Vinorelbine + cisplatin Paclitaxel + cisplatin Docetaxel Best supportive care	Response Survival Toxicity QoL	Active treatments (i.e. treatments other than best supportive care) showed modestly superior response and survival QoL was similar between treatments Active treatments showed similar toxicity compared to best supportive care
Giaccone *et al.* (2002)[308]	RCT	Paclitaxel + cisplatin	Gemcitabine + cisplatin Paclitaxel + gemcitabine	Response Survival	Paclitaxel + cisplatin showed superior response and survival Similar toxicity between treatments

Reference	Study type	Intervention	Comparison	Outcome measures	Authors' conclusion
Grigorescu (2002)[309]	RCT	Vinblastine + cisplatin	Gemcitabine + carboplatin	Response Survival Toxicity	Gemcitabine + cisplatin showed superior response and survival Similar toxicity between treatments
Scagliotti *et al.* (2002)[231]	RCT	Gemcitabine + cisplatin	Paclitaxel + carboplatin Vinorelbine + cisplatin	Response Survival curves Toxicity Time-to-progression QoL (EORTC QLQ-C30-LC13)	No significant difference in response, survival and time-to-progression Differences in toxicity between treatments (vinorelbine + cisplatin) showed higher rates of neutropenia; gemcitabine + cisplatin showed higher rates of alopecia and neurotoxicity) No difference in QoL between gemcitabine + cisplatin and vinorelbine + cisplatin. Compared to vinorelbine + cisplatin, paclitaxel + carboplatin showed superior QoL in functioning, fatigue and nausea dimensions, and inferior QoL related to peripheral neuropathy and alopecia.
Schiller *et al.* (2002)[217]	RCT	Gemcitabine + cisplatin Docetaxel + cisplatin Paclitaxel + carboplatin	Paclitaxel + cisplatin	Response Survival Time-to-progression Toxicity	No significant difference in response and survival between treatments Gemcitabine + cisplatin showed superior (longer) time to progression Gemcitabine + cisplatin showed inferior toxicity

Reference	Study type	Intervention	Comparison	Outcome measures	Authors' conclusion
Alberola *et al.* (2003)[209]	RCT	Gemcitabine + cisplatin + vinorelbine Gemcitabine + vinorelbine + isofamide	Gemcitabine + cisplatin	Response Survival Time-to-progression Toxicity	Gemcitabine + cisplatin + vinorelbine showed inferior response No significant difference in survival and disease progression between treatments Gemcitabine + cisplatin + vinorelbine showed inferior (higher) toxicity
Danson *et al.* (2003)[228]	RCT	Gemcitabine and carboplatin	Mitomycin + ifosfamide + cisplatin Mitomycin + vinblastine + cisplatin	Response Survival Time-to-progression QoL (EORTC C30 and HADS)	No significant difference in survival, response and time-to-progression between treatments Similar toxicity between treatments No significant difference in QoL between treatments
Gridelli *et al.* (2003)[229]	RCT	Gemcitabine + vinorelbine	Gemcitabine + cisplatin Vinorelbine + cisplatin	Response Survival Toxicity Time-to-progression QoL (EORTC QLQ-C30-LC13)	No significant difference in response and survival between treatments Similar toxicity between treatments Gemcitabine + vinorelbine showed superior QoL
Smit *et al.* (2003)[206]	RCT	Gemcitabine + cisplatin	Paclitaxel + cisplatin Paclitaxel + carboplatin	Response Survival Toxicity Time-to-progression QoL (EORTC QLQ-C30-LC13)	No significant difference in survival, response and time-to-progression between treatments Similar toxicity between treatments No significant difference in QoL between treatments

Reference	Study type	Intervention	Comparison	Outcome measures	Authors' conclusion
Zatloukal *et al.* (2003)[208]	RCT	Gemcitabine + cisplatin	Gemcitabine + carboplatin	Response Survival Toxicity Time-to-progression	No significant difference in response, time-to-progression and survival between treatments Similar toxicity between treatments
RCT: randomised controlled trial					

Table 3.c: Summary of identified economic studies (NSCLC)

Study	Type of economic evaluation	Intervention	Comparison	Authors' conclusion
Clegg *et al.* (2001)[201]	CEA based on evidence from multiple studies and decision analytic modelling	Gemcitabine + cisplatin	Various doses and treatment schedules of: Vinorelbine +cisplatin Paclitaxel+cisplatin Docetaxel Best supportive care	Active treatments result in improved survival at a low cost per life-year gained (£2190 to £10,040 per life-year gained).
Palmer and Brandt (1996)[214]	CEA based on evidence from multiple study	Gemcitabine + cisplatin	Mitomicyn + ifosfamide + cisplatin Etoposide + cisplatin Vinorelbine + cisplatin	No significant difference between the cost-effectiveness of treatments The use of the comparator treatments in place of gemcitabine + cisplatin will result in additional costs.

Study	Type of economic evaluation	Intervention	Comparison	Authors' conclusion
Lees *et al.* (2002)[213]	CEA based on evidence from multiple studies	Gemcitabine + best supportive care Gemcitabine + cisplatin	Best supportive care Standard chemotherapies (etoposide + cisplatin; mitomycin + ifosfamide + cisplatin; mitomycin + vinolastine+platinum) Novel chemotherapies (paclitaxel + cisplatin; paclitaxel + carboplatin; docetaxel + cisplatin; vinorelbine + cisplatin)	Gemcitabine + best supportive care resulted in £5230 per additional progression-free life-year compared to best supportive care alone. Gemcitabine + cisplatin resulted in cost per additional progression-free life year of £1750 and £5680 against etpoposide + cisplatin and mitomycin + vinoblastine + platinum, respectively. Gemcitabine + cisplatin resulted in cost savings in comparison to all novel chemotherapies.
Khan *et al.* (1999)[210]	CMA	Carboplatin in combination with other agents for non-small cell lung cancer, small cell lung cancer and ovarian cancer	Cisplatin in combination with other agents for non-small cell lung cancer, small cell lung cancer and ovarian cancer	In NSCLC, carboplatin-based combinations result in higher total cost.
Sacristan *et al.* (2000)[211]	CEA based on a single study	Gemcitabine + cisplatin	Etoposide + cisplatin	Gemcitabine + cisplatin appear to be cost-effective, resulting in 'favourable' cost-effectiveness ratios
Schiller *et al.* (2004)[212]	CMA	Gemcitabine + cisplatin	Vinorelbine + cisplatin Paclitaxel + cisplatin Paclitaxel + carboplatin Docetaxel + cisplatin	Gemcitabine + cisplatin resulted in lower total cost than vinorelbine + cisplatin, paclitaxel + cisplatin and paclitaxel + carboplatin. Against docetaxel + cisplatin, gemcitabine + cisplatin resulted in similar or lower cost.

Study	Type of economic evaluation	Intervention	Comparison	Authors' conclusion
Minshall and Liepa (1998)[215]	Inference on the cost-effectiveness of gemcitabine by drawing on published economic studies.	Gemcitabine versus cisplatin + etoposide Gemcitabine versus ifosfamide + etoposide Gemcitabine versus cisplatine + etoposide versus ifosfamide + etoposide Gemcitabine versus cisplatin + etoposide versus carboplatin + etoposide Gemcitabine versus best supportive care Gemcitabine + cisplatin versus cisplatin + etoposide versus cisplatin + vinorelbine versus mitocycin + ifosfamide + cisplatin		The review suggests that gemcitabine alone or in combination with other agents may result in cost savings or my even be cost-effective.
Szczepura (2002)[216]	Inference on the cost-effectiveness of gemcitabine by drawing on published economic studies.	Gemcitabine + best supportive care versus best supportive care alone Gemcitabine + cisplatin versus 'traditional combinations' (etoposide + cisplatin; mitomycin + isosfamide + cisplatin; mitomicyn+vinblastine+cisplatin) Gemcitabine + cisplation versus 'newer combination' (paclitaxel + cisplatin; paclitaxel + carboplatin; docetaxel + cisplatin) Gemcitabine + cisplatin versus vinorelbine + cisplatin		Gemcitabine-based therapies appeared to be cost-effective against standard chemotherapies
CEA: cost-effectiveness analysis; CMA: cost-minimisation analysis				

Appendix 3.C. Parameters for estimation of transition probabilities for NSCLC model

Table 3.d: Results of regression model for disease progression for Gem+Cisp (NSCLC)

Regression parameters	Coefficients	Standard error	t stat	P-value	Lower 95% CI	Upper 95% CI
Intercept	-2.99	0.11	-27.63	0.00	-3.22	-2.76
ln(t)	1.4	0.05	29.9	0.00	1.31	1.50

Table 3.e: Parameters for Weibull time-to-progression model for Gem+Cisp (NSCLC)

Parameters of Weibull model	
Alpha (α)	1.40
Beta (β) (21-day cycle)	12.17

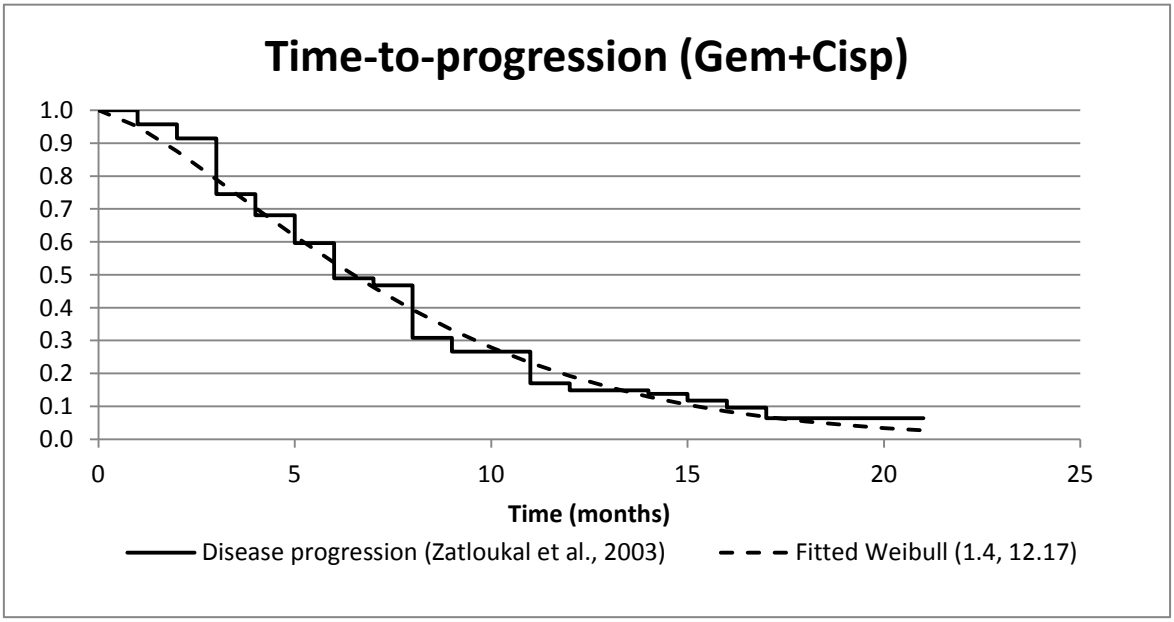

Figure 3.b: Fitted Weibull curve representing time-to-progression for Gem+Cisp (NSCLC)

Table 3.f: Results of regression model for disease progression for Gem+Carb (NSCLC)

Regression parameters	Coefficients	Standard error	t stat	P-value	Lower 95%	Upper 95%
Intercept	-2.47	0.11	-22.56	0.00	-2.70	-2.25
ln(t)	1.29	0.05	27.02	0.00	1.19	1.39

Table 3.g: Parameters for Weibull time-to-progression model for Gem+Carb (NSCLC)

Parameters of Weibull model	
Alpha (α)	1.29
Beta (β) (21-day cycle)	9.91

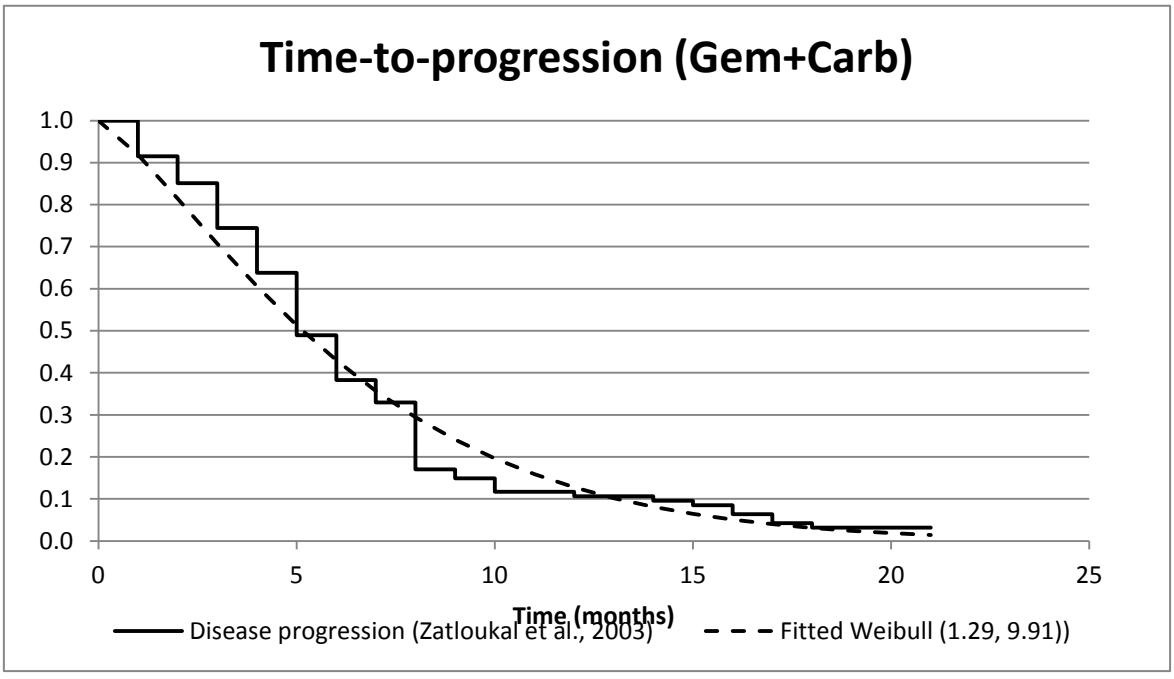

Figure 3.c: Fitted Weibull curve for time-to-progression for Gem+Carb (NSCLC)

Table 3.h: Results of regression model for survival for Gem+Cisp (NSCLC)

Regression parameters	Coefficients	Standard error	t stat	P-value	Lower 95%	Upper 95%
Intercept	-2.81	0.15	-18.93	0.00	-3.15	-2.47
ln(t)	1.10	0.05	20.1	0.00	0.98	1.23

Table 3.i: Parameters for Weibull survival model for Gem+Cisp (NSCLC)

Parameters of Weibull model	
Alpha (α)	1.10
Beta (β) (21-day cycle)	18.4

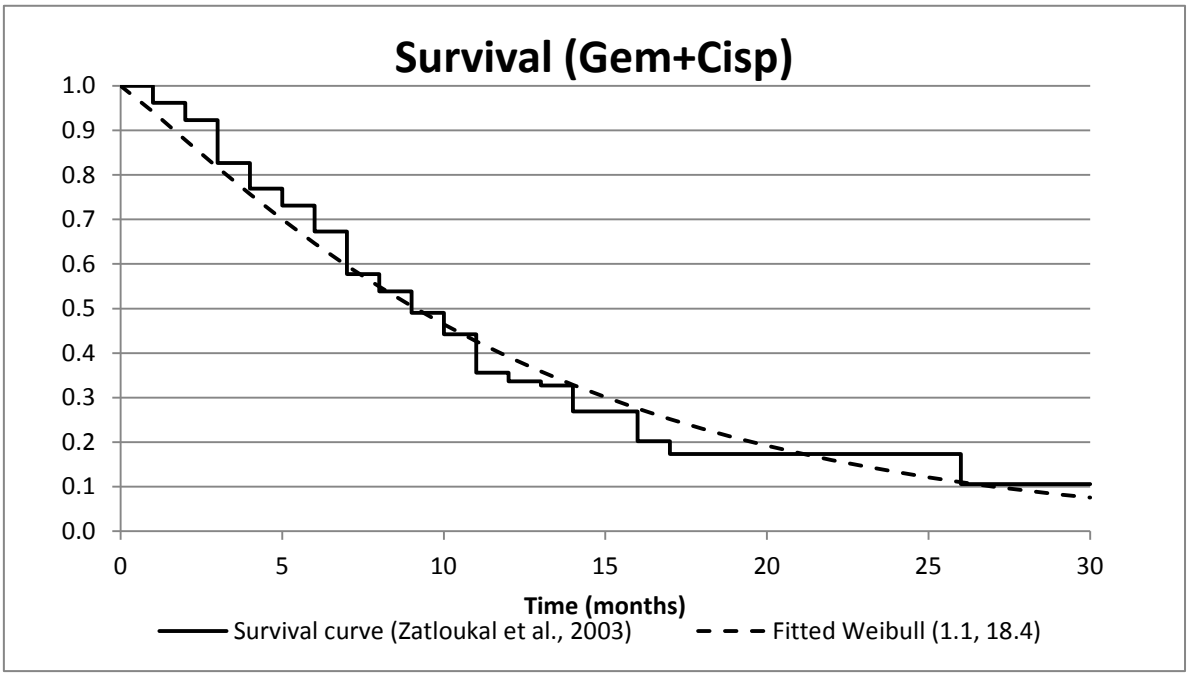

Figure 3.d: Fitted Weibull curve for survival for Gem+Cisp (NSCLC)

Table 3.j: Results of regression model for disease progression for Gem+Carb (NSCLC)

Regression parameters	Coefficients	Standard error	t stat	P-value	Lower 95%	Upper 95%
Intercept	-3.35	0.21	-16.04	0.00	-3.83	-2.87
ln(t)	1.3	0.08	16.82	0.00	1.12	1.48

Table 3.k: Parameters for Weibull survival model for Gem+Carb (NSCLC)

Parameters of Weibull model	
Alpha (α)	1.30
Beta (β) (21-day cycle)	18.99

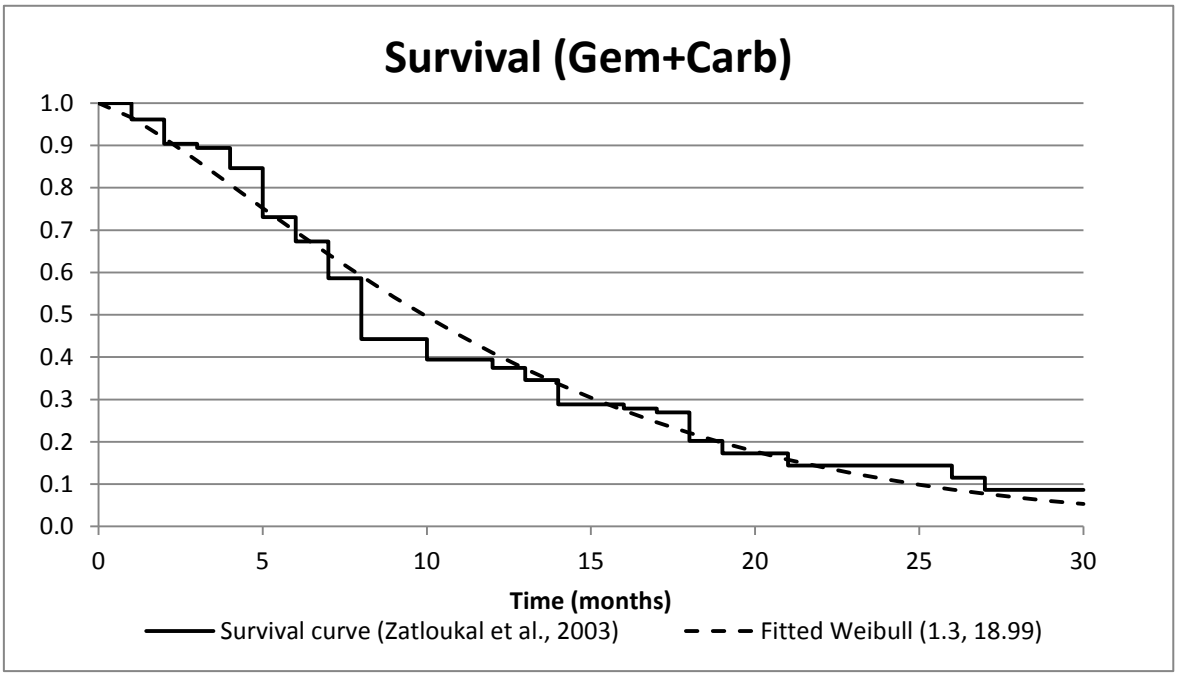

Figure 3.e: Fitted Weibull curve for survival for Gem+Carb (NSCLC)

Appendix 3.D. Parameters for probabilistic sensitivity analysis

Table 3.I: Parameters and assigned distributions for model transitions per treatment (NSCLC)

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Source/comment
Probability of a patient staying in the 'Progression-free' state at each cycle				
Gem+Cisp	Probability of a patient staying at 'Progression-free' state	Alpha and beta parameters of the fitted Weibull time-to-progression model, by varying the model's intercept and regression coefficient	Intercept~normal(-2.99, 0.11) Regression coefficient~normal(1.40, 0.05)	From time-to-progression curve reported in Zatloukal *et al*.[208]
Gem+Carb	Probability of a patient staying at 'Progression-free' state	Alpha and beta parameters of the fitted Weibull time-to-progression model, by varying the model's intercept and regression coefficient	Intercept~normal(-2.48, 0.11) Regression coefficient ~normal(1.29, 0.05)	From time-to-progression curve reported in Zatloukal *et al*.[208]
Probability of a patient dying at each cycle				
Gem+Cisp	Transition probability from 'Progression' to 'Death'.	Alpha and beta parameters of the fitted Weibull survival model, by varying the model's intercept and regression coefficient	Intercept~normal(-2.81, 0.15) Regression coefficient~normal (1.10, 0.06)	From survival curve reported in Zatloukal *et al.*[208]
Gem+Carb	Transition probability from 'Progression' to 'Death'.	Alpha and beta parameters of the fitted Weibull survival model, by varying the model's intercept and regression coefficient	Intercept~normal (-3.35, 0.21) Regression coefficient ~normal(1.30, 0.08)	From survival curve reported in Zatloukal *et al.*[208]

Table 3.m: Parameters and assigned distributions for resource use and cost by treatment (NSCLC)

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Comment/Source
Drug acquisition and administration costs				
Gem+Cisp	Cost of drug acquisition and administration		Gamma(100, 9.46)	Gamma distribution fitted by the method of moments on the basis of: Mean value: £946 (from cost analysis) Standard error: £95 (assumption, 10 percent of mean value)
Gem+Carb	Cost of drug acquisition and administration		Gamma(100, 11.33)	Gamma distribution fitted by the method of moments on the basis of: Mean value:£1133 (from cost analysis) Standard error:£113 (assumption, 10 percent of mean value)
Adverse events-related cost				

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Comment/Source
Gem+Cisp	Cost of adverse events	Probability of patients experiencing different adverse events	Anaemia~beta(10.58, 73.42) Thrombocytopenia~beta(13.78, 70.22) Neutropenia~beta(7.98, 76.02) Granulocytopenia~ beta(19.74, 64.26)	Based on proportions of patients experiencing adverse events as reported in Zatloukal *et al.*[208]
Gem+Carb	Cost of adverse events	Probability of patients experiencing different adverse events	Anaemia~beta(15.84, 72.16) Thrombocytopenia~beta(28.69, 59.31) Neutropenia~beta(12.85, 75.15) Granulocytopenia~ beta(26.66, 61.34)	Based on proportions of patients experiencing adverse events as reported in Zatloukal *et al*.[208]
Cost of other medical resources (same across treatments)				
Gem+Cisp **Gem+Carb**	Cost of other medical resources	Cost of other medical resources	Gamma(16, 45.5)	Gamma distribution fitted on the basis of Mean value:£728 (from Schiller *et al.*[212]) Standard error: £182 (assumption, 25 percent of the mean value)

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Comment/Source
Cost of terminal care (same across treatments)				
Gem+Cisp **Gem+Carb**	Terminal care cost	Terminal care cost	Gamma(16, 91.25)	Gamma distribution fitted on the basis of Mean value:£1460 (from Clegg *et al.*[201]) Standard error: £365 (assumption, 25 percent of the mean value)

Table 3.n: Parameters and assigned distributions for preference-based quality of life scores (NSCLC)

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Comment/Source
Utility values for 'Progression-free' and 'Progression' states				
Gem+Cisp **Gem+Carb**	Utility value of 'Progression- free' state	Utility value of 'Progression-free' state	Normal(0.65,0.08)	Values based on expert opinion (Professor L. Billingham, University of Birmingham)
	Utility value of 'Progression' state	Difference between utilities of 'Progression-free' and 'Progression' states	Normal(0.2, 0.04)	

Appendix 4. Literature review and parameters for HRPC

Appendix 4.A. reports the literature review carried out to retrieve information on the effectiveness, costs and cost-effectiveness of chemotherapy treatments for HRPC. Appendices 4.B. to 4.F. provide information related to parameters used in populating the HRPC decision model.

Appendix 4.A. Search strategies for evidence on HRPC

Search strategies 1 to 5 below were used for searches in MEDLINE (PubMed interface; 1946 to October Week 2 2011).

Search strategy 1

1. (hormone refractory prostate cancer) AND (strontium 89 OR zoledronic acid)

Search strategy 2

1. ((((("Prostatic Neoplasms/drug therapy"[Mesh] OR "Prostatic Neoplasms/economics"[Mesh])) AND "docetaxel "[Substance Name]) AND ("strontium chloride "[Substance Name]) OR "zoledronic acid "[Substance Name]) AND (prostate cancer) AND (cost effectiveness OR cost-effectiveness OR cost-benefit OR cost benefit)

Search strategy 3

1. tax 327 OR tax-327

Search strategy 4

1. ("cost-benefit analysis"[MeSH Terms] OR ("cost-benefit"[All Fields] AND "analysis"[All Fields]) OR "cost-benefit analysis"[All Fields] OR ("cost"[All Fields] AND "effectiveness"[All Fields]) OR "cost effectiveness"[All Fields]) AND ("prostatic neoplasms"[MeSH Terms] OR ("prostatic"[All Fields] AND "neoplasms"[All Fields]) OR "prostatic neoplasms"[All Fields] OR ("prostate"[All Fields] AND "cancer"[All Fields]) OR "prostate cancer"[All Fields]) AND ("clinical trial"[Publication Type] OR "clinical trials as topic"[MeSH Terms] OR "clinical trial"[All Fields]

Search strategy 5

Search strategy 5 was used for searches in the NHS CRD databases (Database of Abstracts of Reviews of Effects (DARE), NHS Economic Evaluation Database (NHS EED) and Health Technology Assessment (HTA) databases) in October 2011.

1. *"prostate" AND "cancer" AND "bone"*

Search strategy 6

Search strategy 6 involved searches for guidelines around chemotherapy for HRPC on the NICE website, 'related articles' searches in PubMed and ISI Web of Science, searches in the reference lists of identified articles and reports, and general searches through Google Scholar®.

Search strategy 7

Search strategy 7 involved searches in the Cochrane Library for systematic reviews in the area of prostatic diseases and urological cancer.

Table 4.a: Identified articles on HRPC by search strategy (HRPC)

Searches in databases	**MEDLINE**	**NHS CRD**	**General searches** (NICE; ISI Web of Science; Google Scholar® reference lists of identified articles)	**Cochrane Collaboration Reviews**
Search strategy 1	85	-	-	-
Search strategy 2	5	-	-	-
Search strategy 3	39	-	-	-
Search strategy 4	105	-	-	-
Search strategy 5	-	36	-	-
Search strategy 6	-	-	6	-
Search strategy 7	-	-	-	4
Total hits	234	36	6	4
Duplicate articles across searches	23	0	0	0
Unique articles within databases	211	36	6	4
Duplicate articles between databases	26			
Total number of unique articles	**231**			
Note: dashes indicate that the particular search strategy was not applied to the specific database.				

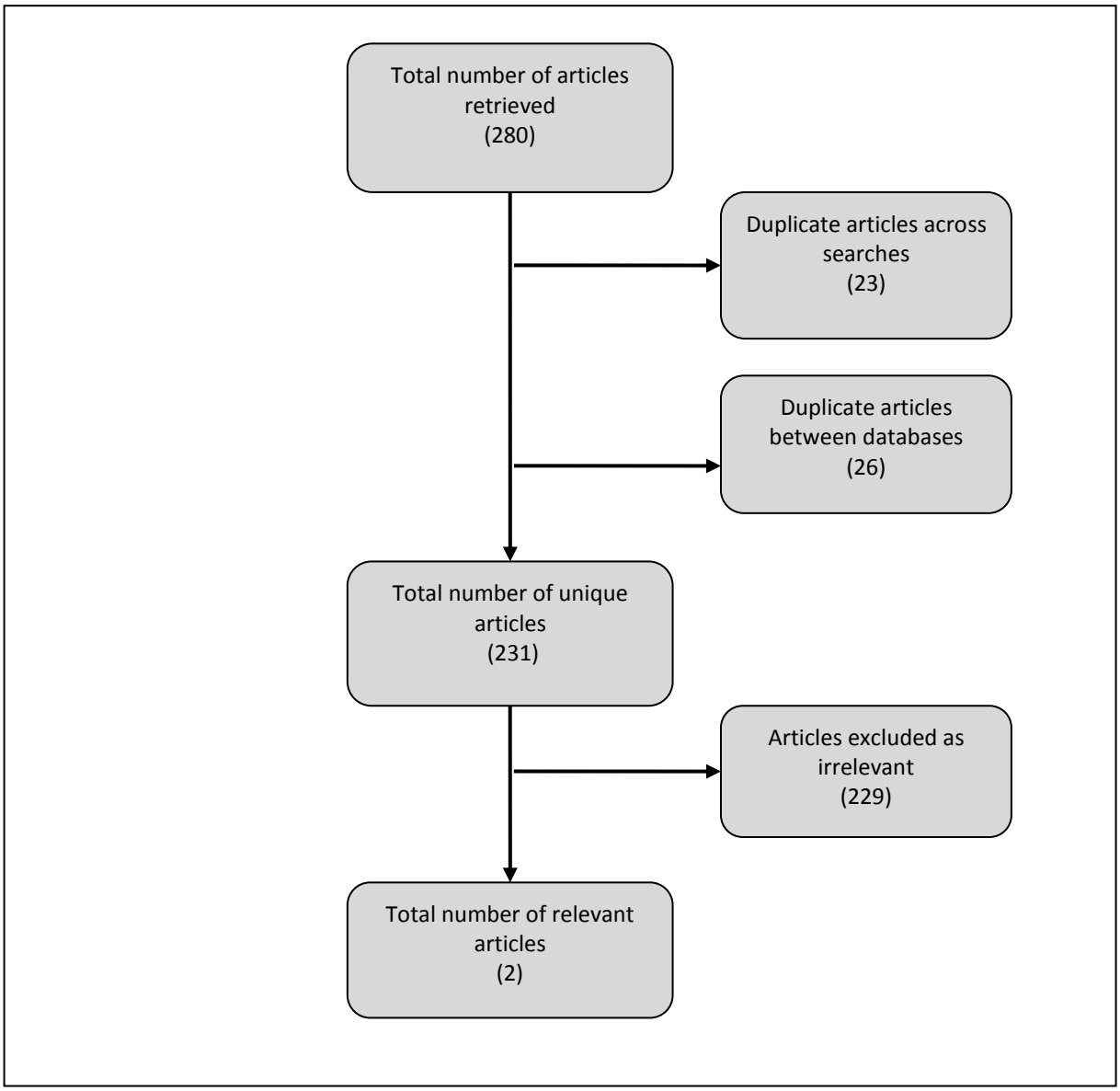

Figure 4.a: Flow chart of identified literature (HRPC)

Appendix 4.B. Data used in estimating transitions probabilities for HRPC model

Table 4.b: Counts of transitions from 'Progression-free, on treatment' to other model states during cycles 1-6 by treatment (HRPC)

From/To	Progression-free, on treatment	Progression-free, not on treatment	Progression	Death	Total
'Progression-free, on treatment'					
DP	204	5	5	5	219
PD+ZA	203	4	4	6	217
DP+Sr89	204	5	3	2	214
DP+ZA+Sr89	218	5	2	3	228

Table 4.c: Counts of transitions between model states from cycle 6 onwards by treatment (HRPC)

From/To	Progression-free, not on treatment	Progression	Death	Total
DP				
Progression-free, not on treatment	371	24	13	408
Progression	-	567	29	596
DP+ZA				
Progression-free, not on treatment	248	26	11	285
Progression	-	590	31	621
DP+Sr89				
Progression-free, not on treatment	461	25	15	501
Progression	-	454	26	480
DP+ZA+Sr89				
Progression-free, not on treatment	479	21	18	518
Progression	-	294	21	315

Appendix 4.C. Comparison of Kaplan-Meier survival curves from data and model

Kaplan-Meier survival curves were plotted to test the fit between survival as given by a) the observed patient-level data (solid line), and b) the HRPC model, using the estimated non-time-dependent transition probabilities (dotted line). Close fit between the curves suggests that the estimated non-time-dependent transition probabilities are appropriate for representing the observed survival data.

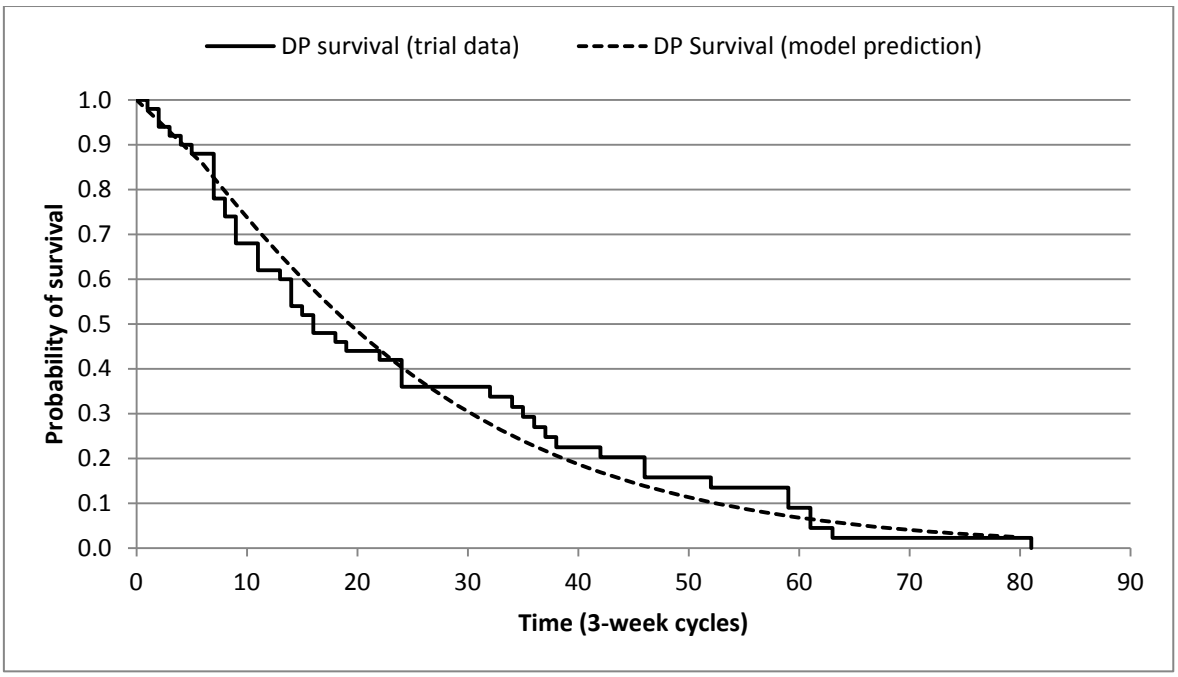

Figure 4.b: Data and model-generated Kaplan-Meier survival curves for DP (HRPC)

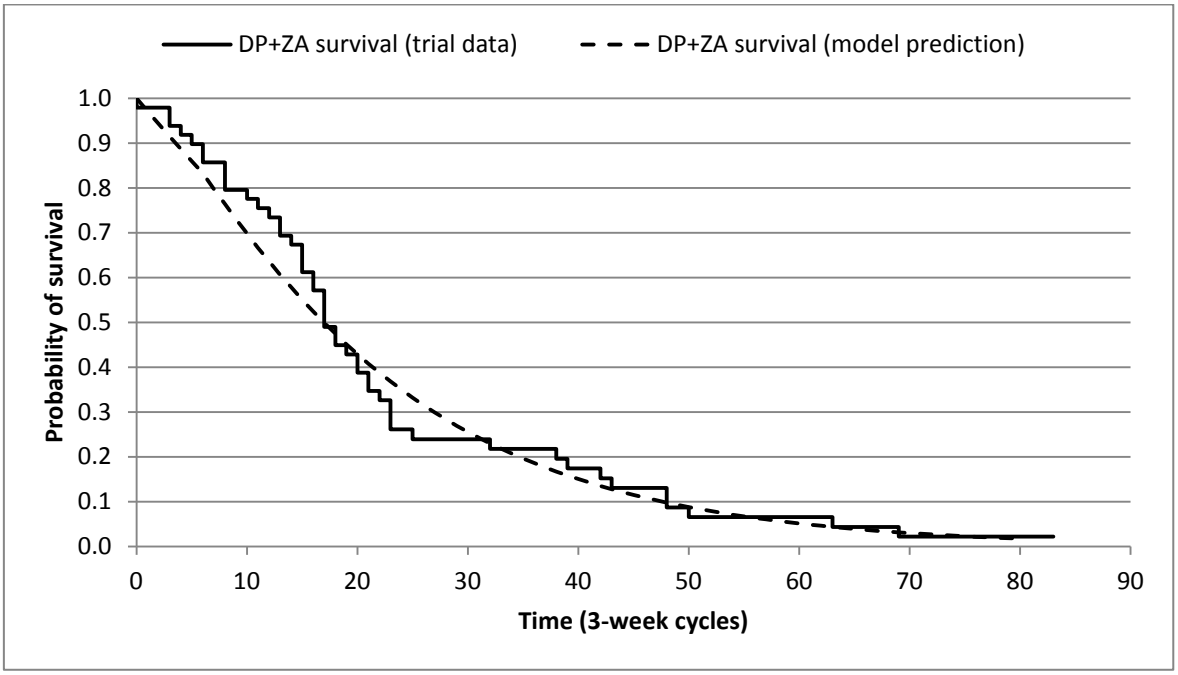

Figure 4.c: Data and model-generated Kaplan-Meier survival curves for DP+ZA (HRPC)

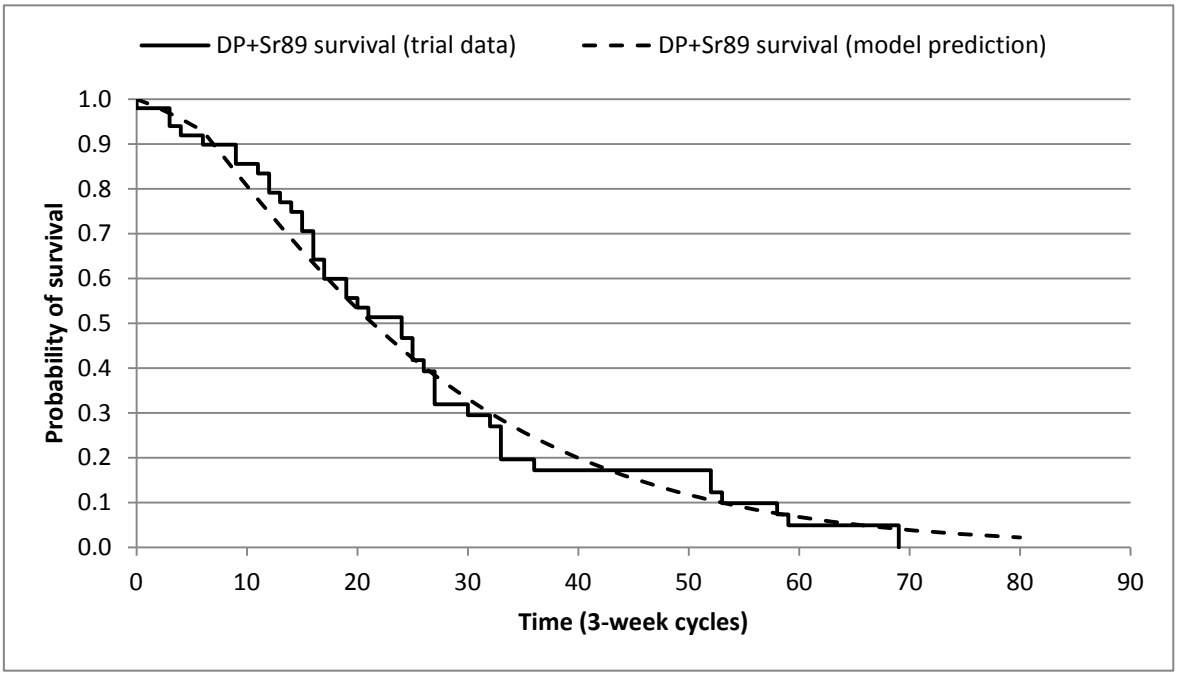

Figure 4.d: Data and model-generated Kaplan-Meier survival curves for DP+Sr89 (HRPC)

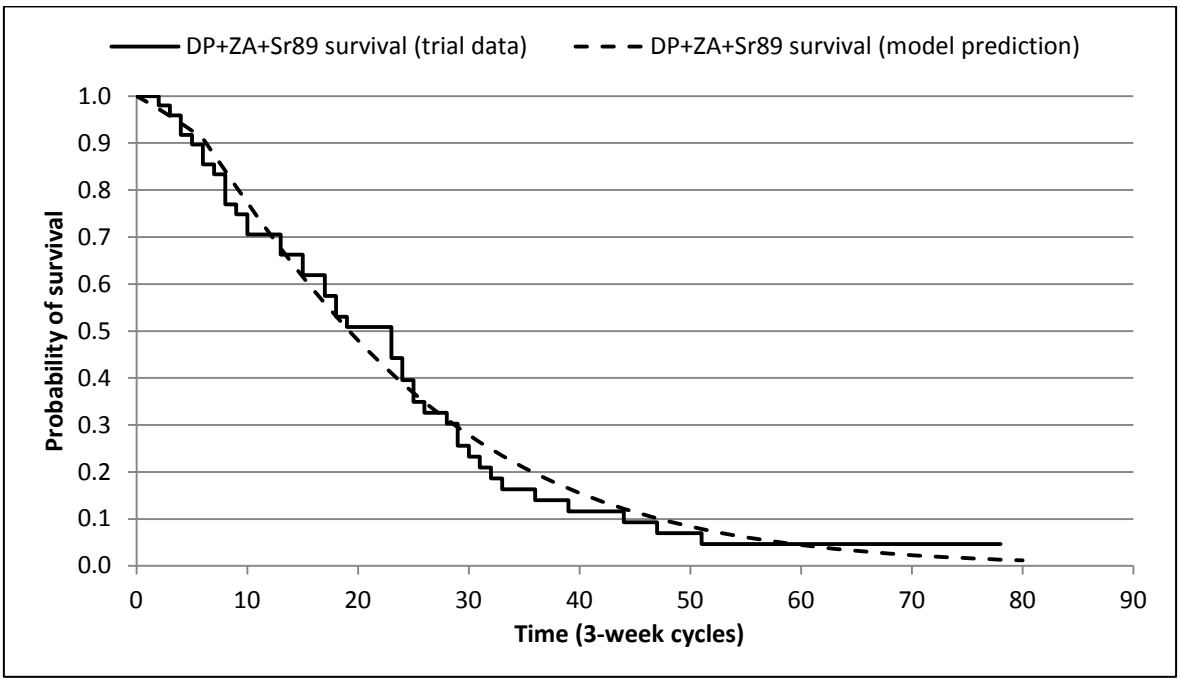

Figure 4.e: Data and model generated Kaplan-Meier survival curves for DP+ZA+Sr89 (HRPC)

Appendix 4.D. Input parameters for estimation of costs and health-related quality of life for HRPC model

Table 4.d: Occurrence of Grade III/VI adverse events by treatment (HRPC)

Adverse event	DP	DP+ZA	DP+Sr89	DP+ZA+Sr89	Total across treatment arms
Diarrhoea	1	1	1	2	5
Febrile neutropenia	3	3	6	2	14
Haemoglobin	1	1	0	2	4
Infection	5	4	2	2	13
Neutrophils/granulocytes	4	0	2	0	6
Pain	5	3	7	3	18
Urinary retention	0	4	0	1	5
Other	20	13	9	24	66
Total number of adverse events	39	29	27	36	131
Total number of patients in arm	50	49	51	50	200

Table 4.e: Unit costs of adverse events (HRPC)

Adverse event	Unit cost	Source
Diarrhoea	£288	NHS Reference Cost Schedules 2009-10[221]
Febrile neutropenia	£433	NHS Reference Cost Schedules 2009-10[221]
Haemoglobin	£517	NHS Reference Cost Schedules 2009-10[221]
Infection	£369	NHS Reference Cost Schedules 2009-10[221]
Neutrophils/granulocytes	£433	NHS Reference Cost Schedules 2009-10[221]
Pain	£902	NHS Reference Cost Schedules 2009-10[221]
Urinary retention	£414	NHS Reference Cost Schedules 2009-10[221]
Other adverse events	£2025	Curtis *et al*.(2010)[220]

Table 4.f: Counts of patients in receipt of second-line treatment (HRPC)

Second-line treatment	DP	DP+ZA	DP+Sr89	DP+ZA+Sr89	Total across treatment arms
Radiotherapy	20	19	20	17	76
Further chemotherapy	2	6	5	3	16
Radioisotopes	6	3	2	0	11
Total no of patients in arm	50	49	51	50	200

Table 4.g: Unit costs of second-line treatment (HRPC)

Second-line treatment	Administration cycles	Unit cost	Cost per course	Source
Radiotherapy	1	£578	£578	NHS Reference Cost Schedules 2009-10[221]
Further chemotherapy	4	£1160	£4640	Calculated cost of DP
Radioisotope (strontium-89)	1	£1576	£1576	Personal communication, Dr C. Boivin, University Hospital Birmingham

Appendix 4.E. Parameters for probabilistic sensitivity analysis

Table 4.h: Transition parameters and assigned distributions per treatment (HRPC)

Treatment	Parameter	Distribution	Source/comment
Transition probabilities from state 'Progression-free, on treatment' (PGF-OT) to states 'Progression-free, not on treatment' (PGF), 'Progression' (PG) and 'Death' (D)			
DP	Transition probabilities from state PGF-OT to states PGF, PG and D	Dirichlet(204, 5 ,5, 5)	Calculated according to data from the TRAPEZE phase II trial
DP+ZA	Transition probabilities from state PGF-OT to states PGF, PG and D	Dirichlet(203, 4, 4, 6)	
DP+Sr89	Transition probabilities from state PGF-OT to states PGF, PG and D	Dirichlet(204, 5 ,3, 2)	
DP+ZA+Sr89	Transition probabilities from state PGF-OT to states PGF, PG and D	Dirichlet(218, 5 ,2, 3)	
Transition probabilities from state 'Progression-free, not on treatment' (PGF) to 'Progression' (PG) and 'Death' (D)			
DP	Transition probabilities from state PGF to states PG and D	Dirichlet(371, 24, 13)	Calculated according to data from the TRAPEZE phase II trial
DP+ZA	Transition probabilities from state PGF to states PG and D	Dirichlet(248, 26, 11)	
DP+Sr89	Transition probabilities from state PGF to states PG and D	Dirichlet(461, 25, 15)	
DP+ZA+Sr89	Transition probabilities from state PGF to states PG and D	Dirichlet(479, 21, 18)	
Transition probabilities from state 'Progression' (PG) to 'Death' (D)			
DP	Transition probabilities from state PG to state D	Beta(29, 567)	Calculated according to data from the TRAPEZE

Treatment	Parameter	Distribution	Source/comment
DP+ZA	Transition probabilities from state PG to state D	Beta(31, 590)	phase II trial
DP+Sr89	Transition probabilities from state PG to state D	Beta(26, 454)	
DP+ZA+Sr89	Transition probabilities from state PG to state D	Beta(21, 294)	

Table 4.i: Parameters and assigned distributions for resource use and cost (HRPC)

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Comment/Source
Cost of drug acquisition and administration				
DP	Cost of drug acquisition and administration		Gamma(100,11.60)	Gamma distribution fitted by the method of moments on the basis of: Mean value: £1160(from cost analysis) Standard error: £116 (assumption, 10 percent of mean value)
DP+ZA	Cost of drug acquisition and administration		Gamma(100, 13.29)	Gamma distribution fitted by the method of moments on the basis of: Mean value: £1329(from cost analysis) Standard error: £133 (assumption, 10 percent of mean value)
DP+Sr89	Cost of drug acquisition and administration		Gamma(100, 11.60)	Gamma distribution fitted by the method of moments on the basis of: Mean value: £1160(from cost analysis) Standard error: £116 (assumption, 10 percent of mean value)

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Comment/Source
DP+ZA+Sr89	Cost of drug acquisition and administration		Gamma(100, 13.29)	Gamma distribution fitted by the method of moments on the basis of: Mean value: £1329(from cost analysis) Standard error: £133 (assumption, 10 percent of mean value)
Cost of strontium-89 acquisition and administration	Cost of strontium-89 acquisition and administration		Gamma(100, 15.76)	Gamma distribution fitted by the method of moments on the basis of: Mean value: £1576 (expert opinion) Standard error: £133 (assumption, 10 percent of mean value)
Cost of adverse events				
DP	Cost of adverse events	Probability of patients experiencing different adverse events	Diarrhoea~beta(1, 49) Febrile neutropenia~ beta(3, 47) Haemoglobin~beta(1, 49) Infection~beta(5, 45) Neutrophils/granulocytes~beta(4, 46) Pain~beta(5, 45) Urinary retention~beta(0, 50) Other~beta(20, 30)	Based on proportions of patients experiencing adverse events obtained from TRAPEZE phase II trial

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Comment/Source
DP+ZA	Cost of adverse events	Probability of patients experiencing different adverse events	Diarrhoea~beta(1, 48) Febrile neutropenia~beta(3, 46) Haemoglobin: Beta(1, 48) Infection~beta(4, 45) Neutrophils/granulocytes~ beta(0, 49) Pain~beta(3, 46) Urinary retention~beta(4, 45) Other~beta(13, 36)	Based on proportions of patients experiencing adverse events obtained from TRAPEZE phase II trial

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Comment/Source
DP+Sr89	Cost of adverse events	Probability of patients experiencing different adverse events	Diarrhoea~beta(1, 50) Febrile neutropenia~beta(6, 45) Haemoglobin~beta(0, 51) Infection~beta(2, 49) Neutrophils/granulocytes~beta (2, 49) Pain~beta(7, 44) Urinary retention~beta(0, 51) Other~beta(9, 42)	Based on proportions of patients experiencing adverse events obtained from TRAPEZE phase II trial

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Comment/Source
DP+ZA+Sr89	Cost of adverse events	Probability of patients experiencing different adverse events	Diarrhoea~beta(2, 48) Febrile neutropenia~ beta(2, 48) Haemoglobin~beta (2, 48) Infection~beta(2, 48) Neutrophils/granulocytes~ beta(0, 50) Pain~beta(3, 47) Urinary retention~beta(1, 49) Other~beta(24, 26)	Based on proportions of patients experiencing adverse events obtained from TRAPEZE phase II trial
Cost of second-line treatment				
DP	Expected cost of second-line treatment	Probability of patients receiving second-line chemotherapy, radiotherapy or radioisotope treatment	Chemotherapy~beta(20, 30) Radiotherapy ~beta(2, 48) Radioisotopes~beta(6, 50)	Based on proportions of patients who received second-line treatment in TRAPEZE phase II trial
DP+ZA	Expected cost of second-line treatment	Probability of patients receiving second-line chemotherapy, radiotherapy or radioisotope treatment	Chemotherapy~beta(19, 30) Radiotherapy ~beta(6, 43) Radioisotopes~beta(3, 46)	Based on proportions of patients who received second-line treatment in TRAPEZE phase II trial
DP+Sr89	Expected cost of second-line treatment	Probability of patients receiving second-line chemotherapy, radiotherapy or radioisotope treatment	Chemotherapy ~beta(20, 31) Radiotherapy ~beta(5, 46) Radioisotopes~beta(2, 49)	Based on proportions of patients who received second-line treatment in TRAPEZE phase II trial

Treatment	Target parameter	Varied parameter	Distribution/ parameter values	Comment/Source
DP+ZA+Sr89	Expected cost of second-line treatment	Probability of patients receiving second-line chemotherapy, radiotherapy or radioisotope treatment	Chemotherapy~beta(17, 33) Radiotherapy ~beta(3, 47) Radioisotopes~beta(0, 50)	Based on proportions of patients who received second-line treatment in TRAPEZE phase II trial
Cost of terminal care				
DP **DP+ZA** **DP+Sr89** **DP+ZA+Sr89**	Terminal care cost	Terminal care cost	Gamma(16, 101.39)	Gamma distribution fitted on the basis of Mean value:£1532 (from Clegg *et al*.[201]) Standard error: £406 (assumption, 25 percent of the mean value)

Table 4.j: Parameters and assigned distributions for preference-based quality of life scores by treatment (HRPC)

Treatment	Distribution
Utility score for state 'Progression-free, on treatment' (PGF-OT)	
DP	Beta(93.14, 55.99)
DP+ZA	Beta(156.75, 53.4)
DP+Sr89	Beta(109.46, 43.78)
DP+ZA+Sr89	Beta (151.39, 50.52)
Utility score for state 'Progression-free, not on treatment'(PGF)*	
DP	Normal(0.019, 0.062)
DP+ZA	Normal(0.006, 0.044)
DP+Sr89	Normal(0.212, 0.05)
DP+ZA+Sr89	Normal(0.099, 0.059)
Utility score for state 'Progression'(PG)**	
DP	Normal(0.125, 0.087)
DP+ZA	Normal(0.143, 0.072)
DP+Sr89	Normal(0.211, 0.096)
DP+ZA+Sr89	Normal(0.166, 0.085)
*Calculated as score for PGF-OT + (utility increment PGF-OT – PGF) **Calculated as utility for PGF-OT + (utility increment PGF-OT – PG)	

List of References

(1) WHO. The World Health Report 2000. Health Systems: Improving Performance [electronic document] (2000). World Health Organization. Available from: http://www.who.int/whr/2000/en/whr00_en.pdf [Accessed 1 Oct. 2009]

(2) Goodman C. A basic methodology kit. In: Szczepura A, Kankaanpaa J, editors. Assessment of health care technologies. Case studies, key concepts and strategic issues. Chichester: John Wiley and Sons; 1996.

(3) Banta HD. Introduction to the EUR-ASSESS report. *International Journal of Technology Assessment in Health Care* 1997; 13(02):133-143.

(4) Stevens A, Abrams K, Brazier J, Fitzpatrick R, Lilford R. The Advanced Handbook of Methods in Evidence Based Healthcare. London: SAGE Publications; 2001.

(5) Cochrane A. Effectiveness and Efficiency: Random Reflections on the Health Service. Leeds: Nuffield Provincial Hospital Trusts; 1972.

(6) Mossialos E, Le Grand J. Health care and cost containment in the European Union. Aldershot: Ashgate; 1999.

(7) Abel Olsen J. Principles in health economics and policy. Oxford: Oxford University Press; 2009.

(8) Harris, A. Economic evaluation and the reimbursement of pharmaceuticals in Australia [electronic document] (1994). Centre for Health Programme Evaluation. Available from: http://www.buseco.monash.edu.au/centres/che/pubs/wp40.pdf [Accessed 11 Oct. 2009]

(9) Detsky AS. Using cost-effectiveness analysis for formulary decision making: from theory into practice. *Pharmacoeconomics* 1994; 6(4):281-288.

(10) Drummond MF. The use of health economic information by reimbursement authorities. *Rheumatology* 2003; 42 Suppl 3:iii60-iii63.

(11) Langley PC. The November 1995 revised Australian guidelines for the economic evaluation of pharmaceuticals. *Pharmacoeconomics* 1996; 9(4):341-352.

(12) Glennie JL, Torrance GW, Baladi JF, Berka C, Hubbard E, Menon D et al. The revised Canadian Guidelines for the Economic Evaluation of Pharmaceuticals. *Pharmacoeconomics* 1999; 15(5):459-468.

(13) Harris AH, Hill SR, Chin G, Li JJ, Walkom E. The role of value for money in public insurance coverage decisions for drugs in Australia: a retrospective analysis 1994-2004. *Medical Decision Making* 2008; 28(5):713-722.

(14) Graf von der Schulenburg JM, Hoffmann C. Review of European guidelines for economic evaluation of medical technologies and pharmaceuticals. *Health Economics in Prevention and Care* 2000; 1(1):2-8.

(15) UK CRC. Health Research Classification System [electronic document] (2009). UK Clinical Research Collaboration. Available from: http://www.hrcsonline.net/sites/default/files/HRCS_Document.pdf [Accessed 12 Oct. 2009]

(16) Szczepura A, Kankaanpaa J. An Introduction to Health Technology Assessment. In: Szczepura A, Kankaanpaa J, editors. Assessment of Health Care Technologies. Case Studies, Key Concepts and Strategic Issues. Chichester: John Wiley and Sons; 1996.

(17) Cooksey, D. A review of UK health research funding [electronic document] (2006). HMSO. Available from: http://www.hm-treasury.gov.uk/d/pbr06_cooksey_final_report_636.pdf [Accessed 10 Oct. 2009]

(18) Cooper N, Coyle D, Abrams K, Mugford M, Sutton A. Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997. *Journal of Health Services Research and Policy* 2005; 10(4):245-250.

(19) Peto R, Collins R, Gray R. Large-scale randomized evidence: Large, simple trials and overviews of trials. *Journal of Clinical Epidemiology* 1995; 48(1):23-40.

(20) Girling DJ, Parmar MKB, Stenning SP, Stephens RJ, Stewart LA. Clinical trials in cancer: principles and practice. Oxford: Oxford University Press; 2003.

(21) Pocock SJ. Clinical Trials: a practical approach. Chichester: John Wiley & Sons Ltd; 1983.

(22) Fuchs VR, Garber AM. The new technology assessment. *New England Journal of Medicine* 1990; 323(10):673.

(23) NICE. Scientific and social value judgements [electronic document] (2004). National Institute for Health and Clinical Excellence. Available from: http://www.nice.org.uk/niceMedia/Pdf/boardmeeting/brdmay04item6.pdf [Accessed 14 Jan. 2010]

(24) Barton S. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. *British Medical Journal* 2000; 321(7256):255-256.

(25) WMA. Declaration of Helsinki. Ethical principles for medical research involving human subjects. 59th WMA General Assembly, Seoul. [electronic document] (2008). World Medical Association. Available from: http://www.wma.net/en/30publications/10policies/b3/17c.pdf [Accessed 12 Oct. 2009]

(26) Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.

(27) Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *British Medical Journal* 1998; 317(7167):1185-1190.

(28) Friedman LM, Furberg C, DeMets DL. Fundamentals of clinical trials. New York: Springer; 1998.

(29) Le Tourneau C, Lee JJ, Siu LL. Dose Escalation Methods in Phase I Cancer Clinical Trials. *Journal of the National Cancer Institute* 2009; 101(10):708-720.

(30) Rubinstein L, Crowley J, Ivy P, Leblanc M, Sargent D. Randomized phase II designs. *Clinical Cancer Research* 2009; 15(6):1883-1890.

(31) Drummond MF, Sculpher MJ, Torrance GW, O'brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. 3 ed. Oxford: Oxford University Press; 2005.

(32) Bloch DA. Sample size requirements and the cost of a randomized clinical trial with repeated measurements. *Statistics in Medicine* 1986; 5(6):663-667.

(33) Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in clinical trials. *Journal of Chronic Diseases* 1967; 20:637-648.

(34) Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG et al. A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *Journal of Clinical Epidemiology* 2009; 62(5):464-475.

(35) Roland M, Torgerson DJ. What are pragmatic trials? *British Medical Journal* 1998; 316(7127):285.

(36) Tunis SR, Stryer DB, Clancy CM. Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy. *Journal of the American Medical Association* 2003; 290(12):1624-1632.

(37) Woolf SH, Henshall C. Health technology assessment in the United Kingdom. *International Journal of Technology Assessment in Health Care* 2000; 16(2):591-625.

(38) Stevens A, Milne R. Health technology assessment in England and Wales. *International Journal of Technology Assessment in Health Care* 2004; 20(1):11-24.

(39) DOH. Best research for best health. A new national health research strategy [electronic document] (2006). Department of Health. Available from: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4127152.pdf [Accessed 12 Sept. 2009]

(40) NIHR HTA. Annual report 2008 [electronic document] (2008). NIHR Health Technology Assessment programme. Available from: http://www.hta.ac.uk/publicationspdfs/Annualreports/HTAannualreport08web.pdf [Accessed 14 Nov. 2009]

(41) NIHR HTA. Identifying and prioritising HTA research [electronic document] (2006). NIHR Health Technology Assessment programme. Available from: http://www.hta.ac.uk/publicationspdfs/Infoleaflets/identificationleaflet.pdf [Accessed 13 Nov. 2009]

(42) NIHR HTA. HTA Clinical Evaluation and Trials process [online]. Available from: http://www.hta.ac.uk/funding/clinicaltrials/index.shtml [Accessed 10 Nov. 2010]

(43) UK CRC. From Donation to Innovation [electronic document] (2007). UK Clinical Research Collaboration. Available from: http://www.ukcrc.org/index.aspx?o=1471 [Accessed 4 Oct. 2009]

(44) EME. About EME [online]. Available from: http://www.eme.ac.uk/ [Accessed 12 Nov. 2011]

(45) EME. Commissioned remit [online]. Available from: http://www.eme.ac.uk/about/commissioned_remit.asp [Accessed 12 Aug. 2011]

(46) EME. Researcher-led remit [online]. Available from: http://www.eme.ac.uk/about/commissioned_remit.asp [Accessed 12 Aug. 2011]

(47) CRUK. What we do [online]. Available from: http://aboutus.cancerresearchuk.org/what-we-do/ [Accessed 11 Jan. 2012]

(48) CRUK. Phase III Clinical Trial Grants [online]. Available from: http://science.cancerresearchuk.org/funding/find-grant/all-funding-schemes/phase-III-clinical-trial-grants/ [Accessed 4 Jan. 2010]

(49) CRUK. Clinical Trials Awards and Advisory Committee [online]. Available from: http://science.cancerresearchuk.org/funding/funding-committees/clinical-trials-advisory/ [Accessed 5 Jan. 2010]

(50) ABPI. Annual Report 2008/2009 [electronic document] (2009). Association of the British Pharmaceutical Industry. Available from: http://www.abpi.org.uk/publications/pdfs/AnnualReport08.pdf [Accessed 12 Dec. 2009]

(51) Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.

(52) ABPI. ABPI guidance notes on joint working between pharmaceutical companies and the NHS and others for the benefit of patients [electronic document] (2009). http://www.abpi.org.uk/_layouts/download.aspx?SourceUrl=http://www.abpi.org.uk/our-work/library/guidelines/Documents/ABPI_Code_Guidance_Notes.pdf [Accessed 12 Dec. 2009]

(53) Higgins, J. P. T. and Green, S. Cochrane handbook for systematic reviews of interventions version 5.0. 2 [updated September 2009] [electronic document] (2009). The Cochrane Collaboration. Available from: http://www.cochrane-handbook.org/ [Accessed 13 Nov. 2009]

(54) Oxman AD. Systematic reviews: checklists for review articles. *British Medical Journal* 1994; 309(6955):648.

(55) Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P. The role of modelling in prioritising and planning clinical trials. *Health Technology Assessment* 2003; 7(23):iii-125.

(56) Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York: University of York; 2009.

(57) Lin J, Wilbur WJ. PubMed related articles: a probabilistic topic-based model for content similarity. *BMC bioinformatics* 2007; 8(1):423.

(58) Glasziou P, Bain C, Colditz G. Systematic reviews in health care: a practical guide. Cambridge: Cambridge University Press; 2001.

(59) Longworth, L., Bojke, L., Tosh, J., and Sculpher, M. MRC-NICE scoping project: identifying the National Institute for Health and Clinical Excellence's methodological research priorities and an initial set of priorities [electronic document] (2009). Centre for Health Economics, University of York. Available from: http://www.york.ac.uk/media/che/documents/papers/researchpapers/rp51_identifying_NICE%27s_methodological_research_priorities.pdf [Accessed 15 Dec. 2009]

(60) Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials* 1995; 16(1):62-73.

(61) Moher D, Klassen TP, Jones AL, Pham BA, Cook DJ, Jadad AR et al. Assessing the quality of reports of randomised trials included in meta-analyses: attutudes, practice, evidence and guides. In: Stevens A, Abrams K, Brazier J, Fitzpatrick R, Lilford R, editors. The Advanced Book of Methods in Evidence Based Healthcare. London: SAGE Publications; 2001.

(62) Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. *Journal of Health Services Research and Policy* 1997; 2(1):26-30.

(63) Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *International Journal of Technology Assessment in Health Care* 2005; 21(02):240-245.

(64) Donaldson M, Sox H. Setting Priorities for Health Technology Assessment: A Model Process. Washington, D.C.: National Academies Press; 1992.

(65) Henshall C, Oortwijn W, Stevens A, Granados A, Banta D. Priority setting for health technology assessment - Theoretical considerations and practical approaches - A paper produced by the priority setting subgroup of the EUR ASSESS project. *International Journal of Technology Assessment in Health Care* 1997; 13(2):144-185.

(66) Lara M, Goodman C. National priorities for the assessment of clinical conditions and medical technologies: report of a pilot study. Washington, D.C.: National Academies Press; 1990.

(67) Townsend J, Buxton M, Harper G. Prioritisation of health technology assessment. The PATHS model: methods and case studies. *Health Technology Assessment* 2003; 7(20):iii, 1-iii,82.

(68) Black N. A national strategy for research and development: lessons from England. *Annual Review of Public Health* 1997; 18:485-505.

(69) Harper G, Townsend J, Buxton M. The preliminary economic evaluation of health technologies for the prioritization of health technology assessments: A discussion. *International Journal of Technology Assessment in Health Care* 1998; 14(4):1998.

(70) Oortwijn WJ, Vondeling H, van BT, van VC, Bouter LM. Priority setting for health technology assessment in The Netherlands: principles and practice. *Health Policy* 2002; 62(3):227-242.

(71) Sassi F. Setting priorities for the evaluation of health interventions: when theory does not meet practice. *Health Policy* 2003; 63(2):141-154.

(72) EME. Annual Review 2009-2010 [electronic document] (2010). Efficacy and Mechanism Evaluation programme. Available from: http://www.eme.ac.uk/about/annualreview.asp [Accessed 11 May 2010]

(73) Douw K, Vondeling H. Selection of new health technologies for assessment aimed at informing decision making: A survey among horizon scanning systems. *International Journal of Technology Assessment in Health Care* 2006; 22(02):177-183.

(74) Oortwijn W, Banta D, Vondeling H, Bouter L. Identification and priority setting for health technology assessment in The Netherlands: actors and activities. *Health Policy* 1999; 47(3):241-253.

(75) Robert G, Stevens A, Gabbay J. Identifying new health technologies. In: Stevens A, Abrams K, Brazier J, Fitzpatrick R, Lilford R, editors. The Advanced Book of Methods in Evidence Based Healthcare. London: SAGE Publications; 2001.

(76) CRUK. Guidelines for Preparing Research Grant Applications within eGMS [electronic document] (2011). Cancer Research UK. Available from: http://science.cancerresearchuk.org/prod_consump/groups/cr_common/@fre/@fun/documents/generalcontent/clinical-trials-app-guidelines.pdf [Accessed 18 Oct. 2011]

(77) EME. Annual Review 2011-2012 [electronic document] (2012). Efficacy and Mechanism Evaluation programme. Available from: http://www.eme.ac.uk/about/pdfs/EME_AnnualReview_2011_12FINAL.pdf [Accessed 22 Oct. 2011]

(78) NIHR HTA. FAQ: What criteria will be used to assess my proposal? [online]. Available from: http://www.hta.ac.uk/funding/troubleshooting/clinicaltrials.html#hta115 [Accessed 10 Mar. 2012]

(79) Lomas J, Fulop N, Gagnon D, Allen P. On being a good listener: Setting priorities for applied health services research. *Milbank Quarterly* 2003; 81(3):363-384.

(80) Fleurence R, Torgerson D. Setting priorities for research. *Health Policy* 2004; 69(1):Jul.

(81) NIHR HTA. Submitting a successful application for HTA Clinical Evaluation and Trials funding [electronic document] (2012). http://www.hta.ac.uk/funding/clinicaltrials/Submitting_application.pdf [Accessed 1 Mar. 2012]

(82) Noorani HZ, Husereau DR, Boudreau R, Skidmore B. Priority setting for health technology assessments: a systematic review of current practical approaches. *International Journal of Technology Assessment in Health Care* 2007; 23(3):310-315.

(83) Menon D, Stafinski T. Engaging the public in priority-setting for health technology assessment: findings from a citizens' jury. *Health Expectations* 2008; 11(3):282-293.

(84) Chalmers I. What do I want from health research and researchers when I am a patient? *British Medical Journal* 1995; 310(6990):1315-1318.

(85) Entwistle V, Calnan M, Dieppe P. Consumer involvement in setting the health services research agenda: Persistent questions of value. *Journal of Health Services Research and Policy* 2008; 13(SUPPL. 3):October.

(86) Jones R, Lamont T, Haines A. Setting priorities for research and development in the NHS: A case study on the interface between primary and secondary care. *British Medical Journal* 1995; 311(7012):21.

(87) Oortwijn WJ, Vondeling H, Bouter L. The use of societal criteria in priority setting for health technology assessment in The Netherlands. Initial experiences and future challenges. *International Journal of Technology Assessment in Health Care* 1998; 14(2):226-236.

(88) Lansang MA. Priority setting for health research: Lessons from developing countries. *Health Policy and Planning* 2000; 15(2):2000.

(89) EME. EME application process [online]. Available from: http://www.eme.ac.uk/funding/application_process.asp [Accessed 10 Mar. 2011]

(90) Eddy DM. Selecting technologies for assessment. *International Journal of Technology Assessment in Health Care* 1989; 5(4):485-501.

(91) Detsky AS. Using cost-effectiveness analysis to improve the efficiency of allocating funds to clinical trials. *Statistics in Medicine* 1990; 9(1-2):173-184.

(92) Townsend J, Buxton M. Cost effectiveness scenario analysis for a proposed trial of hormone replacement therapy. *Health Policy* 1997; 39(3):181-194.

(93) Hanney S, Buxton M, Green C, Coulson D, Raftery J. An assessment of the impact of the NHS Health Technology Assessment Programme. *Health Technology Assessment* 2007; 11(53):iii-xi, 1.

(94) Weinstein MC. Cost-effective priorities for cancer prevention. *Science* 1983; 221(4605):17-23.

(95) Phelps CE, Parente ST. Priority setting in medical technology and medical practice assessment. *Medical Care* 1990; 28(8):703-723.

(96) Drummond MF, Crump BJ, Little VA. Funding research and development in the NHS. *Lancet* 1992; 339(8787):230-231.

(97) Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Economics* 1996; 5(6):513-524.

(98) Davies L, Drummond M, Papanikolaou P. Prioritizing investments in health technology assessment. Can we assess potential value for money? *International Journal of Technology Assessment in Health Care* 2000; 16(1):73-91.

(99) Detsky AS. Using economic analysis to determine the resource consequences of choices made in planning clinical trials. *Journal of Chronic Diseases* 1985; 38(9):753-765.

(100) Drummond MF, Davies LM, Ferris FL, III. Assessing the costs and benefits of medical research: the diabetic retinopathy study. *Social Science and Medicine* 1992; 34(9):973-981.

(101) Detsky AS. Are clinical trials a cost-effective investment? *Journal of the American Medical Association* 1989; 262(13):1795-1800.

(102) Coyle D, Grunfeld E, Wells G. The assessment of the economic return from controlled clinical trials. A framework applied to clinical trials of colorectal cancer follow-up. *European Journal of Health Economics* 2003; 4(1):6-11.

(103) Phelps CE, Mooney C. Correction and update on 'priority setting in medical technology assessment'. *Medical Care* 1992; 30(8):744-751.

(104) Buxton M, Elliot R, Hanney S, Henkel M, Keen J, Sculpher M et al. Assessing Payback from Department of Health Research and Development: Preliminary Report. Vol 2: Eight Case Studies. 19 ed. 1994.

(105) Buxton M, Hanney S. Assessing Payback from Department of Health Research and Development: Preliminary Report. Vol 1: The Main Report'. 19 ed. 1994.

(106) Raiffa H. Decision Analysis: Introductory Lectures on Choice under Uncertainty. Reading: Addison-Wesley; 1968.

(107) Pratt JW, Raiffa H, Schlaifer R. Introduction to statistical decision theory. Cambridge, US: The MIT Press; 1995.

(108) Lindley DV. Making decisions. 2 ed. Chichester: John Wiley & Sons; 2001.

(109) Meltzer D. Addressing uncertainty in medical cost-effectiveness analysis: Implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. *Journal of Health Economics* 2001; 20(1):109-129.

(110) 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 Technology Assessment* 2004; 8(31):iii-60.

(111) Bojke L, Claxton K, Sculpher MJ, Palmer S. Identifying research priorities: The value of information associated with repeat screening for age-related macular degeneration. *Medical Decision Making* 2008; 28(1):33-43.

(112) Ginnelly L, Claxton K, Sculpher MJ, Golder S. Using value of information analysis to inform publicly funded research priorities. *Applied Health Economics and Health Policy* 2005; 4(1):37-46.

(113) Philips Z, Claxton KP, Palmer S, Bojke L, Sculpher MJ. Priority setting for research in health care: An application of value of information analysis to glycoprotein IIb/IIIa antagonists in non-ST elevation acute coronary syndrome. *International Journal of Technology Assessment in Health Care* 2006; 22(3):379-387.

(114) NICE. Guide to the methods of technology appraisal [electronic document] (2008). National Institute for Health and Clinical Excellence. Available from: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf [Accessed 7 Nov. 2009]

(115) Claxton K. The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics* 1999; 18(3):341-364.

(116) Fleurence R. Setting priorities for research: A practical application of 'payback' and expected value of information. *Health Economics* 2007; 16(12):Dec.

(117) Drummond MF, Davies LM, Ferris FL, III. Assessing the costs and benefits of medical research: the diabetic retinopathy study. *Social Science & Medicine* 1992; 34(9):973-981.

(118) Davies L, Drummond M, Papanikolaou P. Prioritising investment in Health Technology Assessment: can we assess the potential value for money? 1999. York, University of York. Discussion Paper 170.

(119) Fleurence R. Setting priorities for research: A practical application of 'payback' and expected value of information. *Health Economics* 2007; 16(12):Dec.

(120) 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 Technology Assessment* 2004; 8(31):iii-60.

(121) Claxton K, Sculpher M. Using value of information analysis to prioritise health research: Some lessons from recent UK experience. *Pharmacoeconomics* 2006; 24(11):1055-1068.

(122) McKenna C, Claxton K. Addressing adoption and research design decisions simultaneously: the role of value of sample information analysis. *Medical Decision Making* 2011; 31(6):853-865.

(123) Eckermann S, Karnon J, Willan AR. The value of value of information: best informing research design and prioritization using current methods. *Pharmacoeconomics* 2010; 28(9):699-709.

(124) Fenwick E, Claxton K, Sculpher M. The value of implementation and the value of information: Combined and uneven development. *Medical Decision Making* 2008; 28(1):21-32.

(125) Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S. Applied methods of cost-effectiveness analysis in health care. Oxford: Oxford University Press; 2011.

(126) McPake B, Normand C. Health Economics: An International Perspective. 2 ed. London: Routledge; 2008.

(127) Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada [3rd edition] [electronic document] (2006). Canadian Agency for Drugs and Technologies in Health. Available from: http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf [Accessed 11 Oct. 2009]

(128) Johannesson M. Theory and methods of economic evaluation of health care. Dordrecht: Kluwer Academic Publishers; 1996.

(129) Luce BR, Elixhauser A. Estimating costs in the economic evaluation of medical technologies. *International Journal of Technology Assessment in Health Care* 1990; 6(1):57-75.

(130) Brazier J, Ratcliffe J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press; 2007.

(131) Pauly MV. Valuing health care benefits in money terms. In: Sloan FA, editor. Valuing health care. Cambridge: Cambridge University Press; 1995.

(132) Sugden R, Williams AH. The principles of practical cost-benefit analysis. Oxford: Oxford University Press; 1978.

(133) Brouwer WBF, Koopmanschap MA. On the economic foundations of CEA. Ladies and gentlemen, take your positions! *Journal of Health Economics* 2000; 19(4):439-459.

(134) Phelps CE, Mushlin AI. On the (near) equivalence of cost-effectiveness and cost-benefit analyses. *International Journal of Technology Assessment in Health Care* 1991; 7(1):12.

(135) Donaldson C. The (Near) Equivalence of Cost-Effectiveness and Cost-Benefit Analyses: Fact or Fallacy? *Pharmacoeconomics* 1998; 13(4):389-396.

(136) Mishan EJ. Cost Benefit Analysis. London: Macmillan Books; 1971.

(137) McIntosh E, Clarke P, Frew E, Louvier J. Applied Methods of Cost-Benefit Analysis in Health Care. Oxford: Oxford University Press; 2010.

(138) Ryan M, Watson V, Amaya-Amaya M. Methodological issues in the monetary valuation of benefits in healthcare. *Expert Reviews in Pharmacoeconomics and Outcomes Research* 2003; 3(6):717-727.

(139) Morris S, Devlin N, Parkin D. Economic analysis in health care. Chichester: John Wiley & Sons; 2007.

(140) Torrance GW. Preferences for health outcomes and cost-utility analysis. *American Journal of Managed Care* 1997; 3 Suppl:S8-20.

(141) Williams A. Economics of coronary artery bypass grafting. *British Medial Journal* 1985; 291(6491):326-329.

(142) Haycox, A. and Walker, A. What is cost-minimisation analysis? [electronic document] (2009). Oxford Bandolier. Available from: http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/What_is_cost-min.pdf [Accessed 14 Feb. 2010]

(143) Briggs AH, O'brien BJ. The death of cost-minimization analysis? *Health Economics* 2001; 10(2):179-184.

(144) Coast J. Is economic evaluation in touch with society's health values? *British Medical Journal* 2004; 329(7476):1233-1236.

(145) Weinstein MC, Zeckhauser R. Critical ratios and efficient allocation. *Journal of Public Economics* 1973; 2(2):147-157.

(146) Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics* 2004; 13(5):437-452.

(147) McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics* 2008; 26(9):733-744.

(148) 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. *Journal of Health Services Research and Policy* 2007; 12(1):56-58.

(149) Tambour M, Zethraeus N. Bootstrap confidence intervals for cost-effectiveness ratios: some simulation results. *Health Economics* 1998; 7(2):143-147.

(150) O'Brien B, Briggs AH. Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. *Statistical Methods in Medical Research* 2002; 11(6):455-468.

(151) Stinnett AA, Mullahy J. Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* 1998; 18(2):S68-S80.

(152) Stinnett AA, Mullahy J. The negative side of cost-effectiveness analysis. *Journal of the American Medical Association* 1997; 277(24):1931-1932.

(153) Briggs A. Economics notes: handling uncertainty in economic evaluation. *British Medical Journal* 1999; 319(7202):120.

(154) Kuntz K, Weinstein M. Modelling in economic evaluation. In: Drummond M, McGuire A, editors. Economic evaluation in health care. Oxford: Oxford University Press; 2001.

(155) Glick HA, Jalpa DA, Sonnad SS, Polsky D. Economic evaluation in clinical trials. Oxford: Oxford University Press; 2007.

(156) Meltzer D, Smith P. Theoretical issues relevant to the economic evaluation of health care technologies. In: Pauly MV, McGuire A, Barros P, editors. Handbook of Health Economics. Oxford: Elsevier; 2012.

(157) Hunink M, Glasziou P, Siegel J, Weeks J, Pliskin J, Elstein A et al. Decision Making in Health and Medicine. Cambridge: Cambridge University Press; 2001.

(158) Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Economics* 2006; 15(7):677-687.

(159) Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T et al. Modelling in economic evaluation: an unavoidable fact of life. *Health Economics* 1997; 6(3):217-227.

(160) NICE. Guide to the single technology appraisal process [electronic document] (2008). National Institute for Health and Clinical Excellence. Available from: http://www.nice.org.uk/media/913/06/Guide_to_the_STA-proof_6-26-10-09.pdf [Accessed 14 June 2010]

(161) Bell DE, Raiffa H, Tversky A. Decision making: Descriptive, normative, and prescriptive interactions. Cambridge: Cambridge University Press; 1988.

(162) Yassine AA, Chelst KR, Falkenburg DR. A decision analytic framework for evaluating concurrent engineering. *IEEE Transactions on Engineering Management* 1999; 46(2):144-157.

(163) Schlaifer R. Probability and statistics for business decisions. An introduction to managerial economics under uncertainty. New York: McGraw-Hill; 1959.

(164) Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Economics* 2006; 15(12):1295-1310.

(165) NICE. Guide to the multiple technology appraisal process [electronic document] (2009). National Institute for Health and Clinical Excellence. Available from: http://www.nice.org.uk/media/916/6B/Guide_to_the_MTA-proof_8-26-10-09.pdf [Accessed 15 Nov. 2009]

(166) Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *Journal of Health Services Research and Policy* 2004; 9(2):110-118.

(167) Keeler E. Decision trees and Markov models in cost-effectiveness research. In: Sloan FA, editor. Valuing health care. Cambridge: Cambridge University Press; 1995.

(168) Roberts T, Robinson S, Barton P, Bryan S, McCarthy A, Macleod J et al. The correct approach to modelling and evaluating chlamydia screening. *Sexually Transmitted Infections* 2004; 80(4):324-325.

(169) Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics* 2000; 17(5):479-500.

(170) Andronis L, Barton P, Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. *Health Technology Assessment* 2009; 13(29):iii, ix-61.

(171) Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. *Medical Decision Making* 1985; 5(2):157-177.

(172) Willan A, Briggs AH. Statistical analysis of cost-effectiveness data. Chichester: John Wiley & Sons; 2006.

(173) Black WC. The CE Plane: a Graphic Representation of Cost-Effectiveness. *Medical Decision Making* 1990; 10(3):212-214.

(174) Van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. *Health Economics* 1994; 3(5):309-319.

(175) Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics* 2001; 10(8):779-787.

(176) Barton GR, Briggs AH, Fenwick EA. Optimal cost-effectiveness decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI). *Value in Health* 2008; 11(5):886-897.

(177) Buxton M, Hanney S. How can payback from health services research be assessed? *Journal of Health Services Research & Policy* 1996; 1(1):35-43.

(178) Coyle RG. Decision analysis. London: Thomas Nelson and Sons; 1972.

(179) Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet* 2002; 360(9334):711-715.

(180) Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. *Medical Decision Making* 1998; 18(1):95-109.

(181) Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M et al. Probabilistic sensitivity analysis for NICE technology assessment: Not an optional extra. *Health Economics* 2005; 14(4):339-347.

(182) Petitti DB. Meta-analysis, decision analysis, and cost-effectiveness analysis: methods for quantitative synthesis in medicine. New York: Oxford University Press; 2000.

(183) Philips Z, Claxton K, Palmer S. The half-life of truth: what are appropriate time horizons for research decisions? *Medical Decision Making* 2008; 28(3):287-299.

(184) Welton NJ, Ades AE, Caldwell DM, Peters TJ. Research prioritisation based on expected value of perfect partial information: a case-study on interventions to

increase uptake of breast cancer screening. *Journal of the Royal Statistical Society* 2008; 171(4):807-841.

(185) Oostenbrink JB, Al MJ, Oppe M, Rutten-Van Molken MPMH. Expected value of perfect information: An empirical example of reducing decision uncertainty by conducting additional research. *Value in Health* 2008; 11(7):1070-1080.

(186) Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *Journal of the American Medical Association* 1999; 282(15):1458-1465.

(187) Fenwick E, Claxton K, Sculpher M. The value of implementation and the value of information: Combined and uneven development. *Medical Decision Making* 2008; 28(1):21-32.

(188) Hoomans T, Fenwick EAL, 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 in Health* 2009; 12(2):315-324.

(189) Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester: John Wiley & Sons; 2004.

(190) Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. *Medical Decision Making* 2004; 24(2):207-227.

(191) Yokota F, Thompson KM. Value of information analysis in environmental health risk management decisions: past, present, and future. *Risk Analysis* 2004; 24(3):635-650.

(192) Koerkamp BG, Nikken JJ, Oei EH, Stijnen T, Ginai AZ, Hunink MGM. Value of information analysis used to determine the necessity of additional research: MR imaging in acute knee trauma as an example. *Radiology* 2008; 246(2):420-425.

(193) Boyle P, Ferlay J. Cancer incidence and mortality in Europe, 2004. *Annals of Oncology* 2005; 16(3):481-488.

(194) CRUK. Lung cancer- UK mortality statistics [online]. Available from: http://info.cancerresearchuk.org/cancerstats/types/lung/mortality/ [Accessed 9 Feb. 2010]

(195) Ferlay J, Parkin DM, Steliarova-Foucher E. Estimates of cancer incidence and mortality in Europe in 2008. *European Journal of Cancer* 2010; 46(4):765-781.

(196) Parkin DM. Global cancer statistics in the year 2000. *Lancet Oncology* 2001; 2(9):533-543.

(197) WHO. Cancer Key Facts [online]. Available from: http://www.who.int/mediacentre/factsheets/fs297/en/ [Accessed 14 Nov. 2010]

(198) NCCAC. The diagnosis and treatment of lung cancer. Methods, evidence and guidance [electronic document] (2005). National Collaborating Centre for Acute Care for the National Institute for Health and Clinical Excellence. Available from: http://guidance.nice.org.uk/CG24/Guidance/pdf/English [Accessed 12 June 2010]

(199) Rivera M, Detterbeck F, Loomis S. Epidemiology and Classification of Lung Cancer. In: Detterbeck F, Rivera M, Socinski M, Rosenman J, editors. Diagnosis and treatment of lung cancer: An evidence-based guide for the practicing clinician. Philadelphia: W.B. Saunders Company; 2001.

(200) Visbal AL, Leighl NB, Feld R, Shepherd FA. Adjuvant Chemotherapy for Early-Stage Non-small Cell Lung Cancer. *Chest* 2005; 128(4):2933-2943.

(201) Clegg A, Scott DA, Sidhu M, Hewitson P, Waugh N. A rapid and systematic review of the clinical effectiveness and cost-effectiveness of paclitaxel, docetaxel, gemcitabine and vinorelbine in non-small-cell lung cancer. *Health Technology Assessment* 2001; 5(32):1-195.

(202) Non-small Cell Lung Cancer Collaborative Group. Chemotherapy for non-small cell lung cancer. *Cochrane Database of Systematic Reviews* 2000;(2):CD002139.

(203) Longo D, Fauci A, Kasper D, Hauser S, Jameson J, Loscalzo K. Harrison's principles of internal medicine. 18 ed. New York: McGraw-Hill; 2011.

(204) Numico G, Russi E, Merlano M. Best supportive care in non-small cell lung cancer: is there a role for radiotherapy and chemotherapy? *Lung Cancer* 2001; 32(3):213-226.

(205) Thatcher N, Hopwood P, Anderson H. Improving quality of life in patients with non-small cell lung cancer: research experience with gemcitabine. *European Journal of Cancer* 1997; 33 Suppl 1:S8-13.

(206) Smit EF, van Meerbeeck JP, Lianes P, Debruyne C, Legrand C, Schramel F et al. Three-arm randomized study of two cisplatin-based regimens and paclitaxel plus gemcitabine in advanced non-small-cell lung cancer: a phase III trial of the European Organization for Research and Treatment of Cancer Lung Cancer Group--EORTC 08975. *Journal of Clinical Oncology* 2003; 21(21):3909-3917.

(207) Non-small Cell Lung Cancer Collaborative Group. Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomised clinical trials. Non-small Cell Lung Cancer Collaborative Group. *British Medical Journal* 1995; 311(7010):899-909.

(208) Zatloukal P, Petruzelka L, Zemanova M, Kolek V, Skrickova J, Pesek M et al. Gemcitabine plus cisplatin vs. gemcitabine plus carboplatin in stage IIIb and IV non-small cell lung cancer: a phase III randomized trial. *Lung Cancer* 2003; 41(3):321-331.

(209) Alberola V, Camps C, Provencio M, Isla D, Rosell R, Vadell C et al. Cisplatin plus gemcitabine versus a cisplatin-based triplet versus nonplatinum sequential doublets in advanced non-small-cell lung cancer: a Spanish Lung Cancer Group phase III randomized trial. *Journal of Clinical Oncology* 2003; 21(17):3207-3213.

(210) Khan ZM, Rascati KL, Koeller JM. Economic analysis of carboplatin versus cisplatin in lung and ovarian cancer. *Pharmacoeconomics* 1999; 16(1):43-57.

(211) Sacristan JA, Kennedy-Martin T, Rosell R, Cardenal F, Anton A, Lomas M et al. Economic evaluation in a randomized phase III clinical trial comparing gemcitabine/cisplatin and etoposide/cisplatin in non-small cell lung cancer. *Lung Cancer* 2000; 28(2):97-107.

(212) Schiller J, Tilden D, Aristides M, Lees M, Kielhorn A, Maniadakis N et al. Retrospective cost analysis of gemcitabine in combination with cisplatin in non-small cell lung cancer compared to other combination therapies in Europe. *Lung Cancer* 2004; 43(1):101-112.

(213) Lees M, Aristides M, Maniadakis N, McKendrick J, Botwood N, Stephenson D. Economic evaluation of gemcitabine alone and in combination with cisplatin in the treatment of nonsmall cell lung cancer. *Pharmacoeconomics* 2002; 20(5):325-337.

(214) Palmer AJ, Brandt A. The cost-effectiveness of four cisplatin-containing chemotherapy regimens in the treatment of stages III B and IV non-small cell lung cancer: an Italian perspective. *Monaldi Archives for Chest Disease* 1996; 51(4):279-288.

(215) Minshall ME, Liepa AM. Review of the pharmacoeconomic research on gemcitabine in the treatment of advanced non-small cell lung cancer. *American Journal of Managed Care* 1998; 4(11):1591-1602.

(216) Szczepura A. Healthcare outcomes: gemcitabine cost-effectiveness in the treatment of non-small cell lung cancer. *Lung Cancer* 2002; 38 Suppl 2:S21-S28.

(217) Schiller JH, Harrington D, Belani CP, Langer C, Sandler A, Krook J et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *New England Journal of Medicine* 2002; 346(2):92-98.

(218) Parmar MKB, Machin D. Survival Analysis: A Practical Approach. Chichester: John Wiley & Sons; 1995.

(219) Collett D. Modelling survival data in medical research. London: Chapman & Hall/CRC; 2003.

(220) Curtis, L. Unit Cost of Health and Social Care 2010 [electronic document] (2010). University of Kent. Available from: http://www.pssru.ac.uk/pdf/uc/uc2010/uc2010.pdf [Accessed 12 Jan. 2011]

(221) DOH. National Schedule of Reference Costs 2009-2010 [electronic document] (2011). Department of Health. Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_123459 [Accessed 5 Sept. 2010]

(222) British Medical Association and Royal Pharmaceutical Society of Great Britain. British National Formulary [online]. Available from: http://bnf.org/bnf/bnf/current/ [Accessed 14 Apr. 2010]

(223) NCCC. The diagnosis and treatment of lung cancer (update) [electronic document] (2011). National Collaborating Centre for Cancer for the National Institute for Health and Clinical Excellence. Available from: http://www.nice.org.uk/nicemedia/live/13465/54202/54202.pdf [Accessed 16 May 2010]

(224) Novello S, Kielhorn A, Stynes G, Selvaggi G, De MF, Maestri A et al. Cost-minimisation analysis comparing gemcitabine/cisplatin, paclitaxel/carboplatin and vinorelbine/cisplatin in the treatment of advanced non-small cell lung cancer in Italy. *Lung Cancer* 2005; 48(3):379-387.

(225) Maslove L, Gower N, Spiro S, Rudd R, Stephens R, West P. Estimation of the additional costs of chemotherapy for patients with advanced non-small cell lung cancer. *Thorax* 2005; 60(7):564-569.

(226) Cardenal F, Lopez-Cabrerizo MP, Anton A, Alberola V, Massuti B, Carrato A et al. Randomized phase III study of gemcitabine-cisplatin versus etoposide-cisplatin in the treatment of locally advanced or metastatic non-small-cell lung cancer. *Journal of Clinical Oncology* 1999; 17(1):12-18.

(227) Crino L, Scagliotti GV, Ricci S, De MF, Rinaldi M, Gridelli C et al. Gemcitabine and cisplatin versus mitomycin, ifosfamide, and cisplatin in advanced non-small-cell lung cancer: A randomized phase III study of the Italian Lung Cancer Project. *Journal of Clinical Oncology* 1999; 17(11):3522-3530.

(228) Danson S, Middleton MR, O'Byrne KJ, Clemons M, Ranson M, Hassan J et al. Phase III trial of gemcitabine and carboplatin versus mitomycin, ifosfamide, and cisplatin or mitomycin, vinblastine, and cisplatin in patients with advanced nonsmall cell lung carcinoma. *Cancer* 2003; 98(3):542-553.

(229) Gridelli C, Gallo C, Shepherd FA, Illiano A, Piantedosi F, Robbiati SF et al. Gemcitabine plus vinorelbine compared with cisplatin plus vinorelbine or cisplatin plus gemcitabine for advanced non-small-cell lung cancer: a phase III trial of the Italian GEMVIN Investigators and the National Cancer Institute of Canada Clinical Trials Group. *Journal of Clinical Oncology* 2003; 21(16):3025-3034.

(230) Rosell R, Gatzemeier U, Betticher DC, Keppler U, Macha HN, Pirker R et al. Phase III randomised trial comparing paclitaxel/carboplatin with paclitaxel/cisplatin in patients with advanced non-small-cell lung cancer: a cooperative multinational trial. *Annals of Oncology* 2002; 13(10):1539-1549.

(231) Scagliotti GV, De MF, Rinaldi M, Crino L, Gridelli C, Ricci S et al. Phase III randomized trial comparing three platinum-based doublets in advanced non-small-cell lung cancer. *Journal of Clinical Oncology* 2002; 20(21):4285-4291.

(232) Tappenden P, Chilcott J, Ward S, Eggington S, Hind D, Hummel S. Methodological issues in the economic analysis of cancer treatments. *Eur J Cancer* 2006; 42(17):2867-2875.

(233) Fleurence R, Hollenbeak CS. Rates and probabilities in economic modelling: transformation, translation and appropriate application. *Pharmacoeconomics* 2007; 25(1):3-6.

(234) CRUK. Prostate cancer incidence statistics [online]. Available from: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/prostate/incidence/uk-prostate-cancer-incidence-statistics [Accessed 7 Oct. 2011]

(235) Jani AB. Management strategies for locally advanced prostate cancer. *Drugs Aging* 2006; 23(2):119-129.

(236) Saad F. Treatment of bone complications in advanced prostate cancer: rationale for bisphosphonate use and results of a phase III trial with zoledronic acid. *Seminars in Oncology* 2002; 29(6 Suppl 21):19-27.

(237) Selley S, Donovan J, Faulkner A, Coast J, Gillatt D. Diagnosis, management and screening of early localised prostate cancer. *Health Technology Assessment* 1997; 1(2):i, 1-i,96.

(238) Eisenberger MA, Blumenstein BA, Crawford ED, Miller G, McLeod DG, Loehrer PJ et al. Bilateral orchiectomy with or without flutamide for metastatic prostate cancer. *New England Journal of Medicine* 1998; 339(15):1036-1042.

(239) Muthuramalingan SRP. Management of patients with hormone refractory prostate cancer. *Clinical Oncology* 2004; 16(8):505-516.

(240) James ND, Bloomfield D, Luscombe C. The changing pattern of management for hormone-refractory, metastatic prostate cancer. *Prostate Cancer and Prostatic Diseases* 2006; 9(3):221-229.

(241) NCCC. Prostate Cancer: diagnosis and treatment. Evidence review [electronic document] (2008). National Collaborative Centre for Cancer for the National Institute for Health and Clinical Excellence. Available from: http://guidance.nice.org.uk/CG58/Guidance/pdf/English [Accessed 16 June 2011]

(242) NICE. Docetaxel for the treatment of hormone-refractory metastatic prostate cancer [electronic document] (2006). National Institute for Health and Clinical Excellence. Available from: http://www.nice.org.uk/nicemedia/live/11578/33348/33348.pdf [Accessed 17 Mar. 2011]

(243) Tannock IF, de WR, Berry WR, Horti J, Pluzanska A, Chi KN et al. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. *New England Journal of Medicine* 2004; 351(15):1502-1512.

(244) Collins, R., Fenwick, E., Trowman, R., Perard, R., Norman, G., Light, K., and Birtle, A. A systematic review and economic model of the effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer. [electronic document] (2005). http://www.nice.org.uk/guidance/index.jsp?action=download&o=33344 [Accessed 1 June 2010]

(245) Saad F, Gleason DM, Murray R, Tchekmedyian S, Venner P, Lacombe L et al. A randomized, placebo-controlled trial of zoledronic acid in patients with hormone-refractory metastatic prostate carcinoma. *Journal of the National Cancer Institute* 2002; 94(19):1458-1468.

(246) Dhillon S, Lyseng-Williamson KA. Zoledronic acid : a review of its use in the management of bone metastases of malignancy. *Drugs* 2008; 68(4):507-534.

(247) Saad F, Karakiewicz P, Perrotte P. The role of bisphosphonates in hormone-refractory prostate cancer. *World Journal of Urology* 2005; 23(1):14-18.

(248) Neville-Webbe HL, Gnant M, Coleman RE. Potential anticancer properties of bisphosphonates. *Seminars in Oncology* 2010; 37 Suppl 1:S53-S65.

(249) Santini D, Vespasiani GU, Vincenzi B, Picardi A, Vasaturo F, La CA et al. The antineoplastic role of bisphosphonates: from basic research to clinical evidence. *Annals of Oncology* 2003; 14(10):1468-1476.

(250) Quilty PM, Kirk D, Bolger JJ, Dearnaley DP, Lewington VJ, Mason MD et al. A comparison of the palliative effects of strontium-89 and external beam radiotherapy in metastatic prostate cancer. *Radiotherapy and Oncology* 1994; 31(1):33-40.

(251) Saad F, Al DA, Perrotte P, McCormack M, Benard F, Valiquette L et al. Therapeutic approach to hormone-refractory prostate cancer. *Canadian Journal of Urology* 2006; 13 Suppl 2:52-56.

(252) Collins R, Fenwick E, Trowman R, Perard R, Norman G, Light K et al. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer. *Health Technology Assessment* 2007; 11(2):iii-xviii, 1.

(253) Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Medical Decision Making* 1993; 13(4):322-338.

(254) Naimark D, Krahn MD, Naglie G, Redelmeier DA, Detsky AS. Primer on medical decision analysis: Part 5--Working with Markov processes. *Medical Decision Making* 1997; 17(2):152-159.

(255) Dolan, P., Gudex, C., Kind, P., and Williams, A. A social tariff for EuroQol: Results from a UK genera population survey [electronic document] (1995). University of York. Available from: http://www.york.ac.uk/media/che/documents/papers/discussionpapers/CHE%20Discussion%20Paper%20138.pdf [Accessed 10 Feb. 2010]

(256) Gudex C, Dolan P, Kind P, Williams A. Health state valuations from the general public using the visual analogue scale. *Quality of Life Research* 1996; 5(6):521-531.

(257) Garber AM, Solomon NA. Cost-effectiveness of alternative test strategies for the diagnosis of coronary artery disease. *Annals of Internal Medicine* 1999; 130(9):719-728.

(258) Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics* 1998; 13(4):397-409.

(259) CRUK. Lung cancer incidence statistics [electronic document] (2011). Cancer Research UK. Available from: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/lung/incidence/ [Accessed 16 Nov. 2011]

(260) Shah, K., Cookson, R., Culyer, A. J., and Littlejohns, P. NICE's Social Value Judgements about Equity in Health and Health Care [electronic document] (2011). Centre for Health Economics, University of York. Available from:

http://www.york.ac.uk/media/che/documents/papers/researchpapers/CHERP70_Nice%27s_social_value_judgements_about_equity_in_health.pdf [Accessed 10 Dec. 2011]

(261) Gelman A, Carlin J, Stenr H, Rubin D. Bayesian data analysis. London: Chapman & Hall; 1995.

(262) Claxton K, Thompson KM. A dynamic programming approach to the efficient design of clinical trials. *Journal of Health Economics* 2001; 20(5):797-822.

(263) Brennan A, Kharroubi S, O'Hagan A, Chilcott J. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Medical Decision Making* 2007; 27(4):448-470.

(264) Coyle D, Oakley J. Estimating the expected value of partial perfect information: A review of methods. *European Journal of Health Economics* 2008; 9(3):251-259.

(265) Oakley, J. Simulation sample sizes for Monte Carlo partial EVPI calculations [electronic document] (2009). http://www.jeremy-oakley.staff.shef.ac.uk/techrep568_06.pdf [Accessed 19 June 2011]

(266) Oakley JEB. Simulation sample sizes for Monte Carlo partial EVPI calculations. *Journal of Health Economics* 2010; 29(3):468-477.

(267) Griffin S, Welton NJ, Claxton K. Exploring the Research Decision Space: The Expected Value of Information for Sequential Research Designs. *Medical Decision Making* 2010; 30(2):155-162.

(268) Hoomans T, Ament AJ, Evers SM, Severens JL. Implementing guidelines into clinical practice: what is the value? *Journal of Evaluation in Clinical Practice* 2011; 17(4):606-614.

(269) Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.

(270) Brennan A, Kharroubi SA. Expected value of sample information for Weibull survival data. *Health Economics* 2007; 16(11):1205-1225.

(271) Brennan A, Kharroubi SA. Efficient computation of partial expected value of sample information using Bayesian approximation. *Journal of Health Economics* 2007; 26(1):122-148.

(272) Karnon J. Planning the efficient allocation of research funds: An adapted application of a non-parametric Bayesian value of information analysis. *Health Policy* 2002; 61(3):329-347.

(273) Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Medical Decision Making* 2012; 32(5):733-743.

(274) Ubel PA, DeKay ML, Baron J, Asch DA. Cost-effectiveness analysis in a setting of budget constraints--is it equitable? *New England Journal of Medicine* 1996; 334(18):1174-1177.

(275) Dolan P, Cookson R. A qualitative study of the extent to which health gain matters when choosing between groups of patients. *Health Policy* 2000; 51(1):19-30.

(276) Sassi F, Archard L, Le GJ. Equity and the economic evaluation of healthcare. *Health Technology Assessment* 2001; 5(3):1-138.

(277) Kerr, D., Shemanski, L, and McBride, R. Data Monitoring Committees in practice [online]. Available from: http://www.appliedclinicaltrialsonline.com/appliedclinicaltrials/article/articleDetail.jsp?id=87247 [Accessed 13 Mar. 2013]

(278) NIHR HTA. HTA Clinical Evaluation and Trials: an open call specification document [electronic document] (2009). NIHR Health Technology Assessment programme. Available from: http://www.hta.ac.uk/funding/clinicaltrials/CETSpecDocument.pdf [Accessed 15 Feb. 2010]

(279) NIHR HTA. Identifying research topics [online]. Available from: http://www.hta.ac.uk/suggest/identifyingtopics.shtml [Accessed 4 Mar. 2010]

(280) NIHR HTA. Commissioned calls [online]. Available from: http://www.hta.ac.uk/funding/standardcalls/index.shtml [Accessed 1 Sept. 2012]

(281) NIHR HTA. Deciding research priorities: Guidance for service users when commenting on vignettes [electronic document] (2010). NIHR Health Technology Assessment programme. Available from: http://www.hta.ac.uk/public/decidingresearchpriorities.pdf [Accessed 11 Oct. 2010]

(282) Johansson PO. An introduction to modern welfare economics. Cambridge: Cambridge University Press; 1991.

(283) Williams I, McIver S, Moore D, Bryan S. The use of economic evaluations in NHS decision-making: a review and empirical investigation. *Health Technology Assessment* 2008; 12(7):iii, ix-iii,175.

(284) Schwartz D, Flamant R, Lellouch J. Clinical Trials. London: Academic Press; 1980.

(285) Armitage P. Inference and decision in clinical trials. *Journal of Clinical Epidemiology* 1989; 42(4):293-299.

(286) DfT. Guidance document TAG Unit 3.5.4: Cost Benefit Analysis [electronic document] (2012). Department for Transport. Available from: http://www.dft.gov.uk/webtag/documents/expert/pdf/u3_5_4-cost-benefit-analysis-020723.pdf [Accessed 15 Apr. 2013]

(287) Willan AR. Sample size determination for cost-effectiveness trials. *Pharmacoeconomics* 2011; 29(11):933-949.

(288) Eckermann S, Willan AR. The option value of delay in health technology assessment. *Medical Decision Making* 2008; 28(3):300-305.

(289) 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):1-323.

(290) NICE. Technology appraisal recommendation categories [online]. Available from: http://www.nice.org.uk/newsroom/nicestatistics/TADecisionsRecommendationCategories.jsp [Accessed 3 Feb. 2013]

(291) Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. *British Medical Journal* 2004; 329(7459):224-227.

(292) Williams I, McIver S, Moore D, Bryan S. The use of economic evaluations in NHS decision-making: a review and empirical investigation. *Health Technology Assessment* 2008; 12(7):iii, ix-iii,175.

(293) NIHR HTA. Working with policy customers [electronic document] (2013). National Institute for Health Research Health Technology Assessment programme. Available from: http://www.hta.ac.uk/publicationspdfs/Infoleaflets/policy_customers_leaflet.pdf [Accessed 5 Feb. 2013]

(294) HM Treasury. Budget 2013 [electronic document] (2013). HM Treasury. Available from: http://cdn.hm-treasury.gov.uk/budget2013_complete.pdf [Accessed 15 Apr. 2013]

(295) DOH. Equity and excellence: Liberating the NHS [electronic document] (2010). The Stationary Office. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/135875/dh_117794.pdf.pdf [Accessed 21 Apr. 2013]

(296) Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value in Health* 2003; 6(1):9-17.

(297) Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment* 2004; 8(36):iii-xi, 1.

(298) Weatherly H, Drummond M, Claxton K, Cookson R, Ferguson B, Godfrey C et al. Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. *Health Policy* 2009; 93(2-3):85-92.

(299) Yokota F, Thompson KM. Value of information literature analysis: a review of applications in health risk management. *Medical Decision Making* 2004; 24(3):287-298.

(300) Achilladelis B, Antonakis N. The dynamics of technological innovation: the case of the pharmaceutical industry. *Research Policy* 2001; 30:535-589.

(301) Cooper NJ, Sutton AJ, Ades AE, Paisley S, Jones DR. Use of evidence in economic decision models: practical issues and methodological challenges. *Health Economics* 2007; 16(12):1277-1286.

(302) Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics* 2008; 26(9):753-767.

(303) Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008; 17(3):279-301.

(304) Claxton K, Neumann PJ, Araki S, Weinstein MC. Bayesian value-of-information analysis. An application to a policy model of Alzheimer's disease. *Int J Technol Assess Health Care* 2001; 17(1):38-55.

(305) Chaloner K, Rhame FS. Quantifying and documenting prior beliefs in clinical trials. *Statistics in Medicine* 2001; 20(4):581-600.

(306) Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Methods to elicit beliefs for Bayesian priors: a systematic review. *Journal of Clinical Epidemiology* 2010; 63(4):355-369.

(307) Sandler AB, Nemunaitis J, Denham C, von PJ, Cormier Y, Gatzemeier U et al. Phase III trial of gemcitabine plus cisplatin versus cisplatin alone in patients with locally advanced or metastatic non-small-cell lung cancer. *Journal of Clinical Oncology* 2000; 18(1):122-130.

(308) Giaccone G. Early results of a randomized phase III trial of platinum-containing doublets versus a nonplatinum doublet in the treatment of advanced non-small cell lung cancer: European Organization for Research and Treatment of Cancer 08975. *Seminars in Oncology* 2002; 29(3 Suppl 9):47-49.

(309) Grigorescu AC, Draghici IN, Nitipir C, Gutulescu N, Corlan E. Gemcitabine (GEM) and carboplatin (CBDCA) versus cisplatin (CDDP) and vinblastine (VLB) in advanced non-small-cell lung cancer (NSCLC) stages III and IV: a phase III randomised trial. *Lung Cancer* 2002; 37(1):9-14.