Employing early decision analytic modelling to inform economic evaluation in health care: theory & practice

Kathleen Anne Boyd

MSc Economic & Health Economics, BA (Honours) Economics

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Health Economics and Health Technology Assessment

Institute of Health & Wellbeing

University of Glasgow

October 2012

Abstract

Decision analytic modelling (DAM) is a mathematical technique which is used to structure and synthesise evidence in order to inform decision making, given uncertainty. Decision models are an ideal tool for undertaking economic evaluations as they enable a wide range of data on costs and effects to be synthesised within the model in order to derive cost-effectiveness outcomes. The iterative framework for economic appraisal has been proposed as good practice for undertaking economic evaluations (1), and DAM plays a key role within this framework. In particular there is a role for early stage DAM prior to primary research, to provide an indication of the potential cost-effectiveness of a new health technology (2) given current evidence, and the use of value of information (VOI) techniques to help inform further research priority setting. In practice, support and funding for early stage DAM and full exploitation of VOI techniques is rare.

The aim of this thesis is to examine the role for early decision analytic modelling in informing research priorities and the design of future studies in a health care setting. This thesis explores the feasibility, merits and drawbacks of undertaking early DAM and considers potential reasons as to why it has not been more widely implemented.

This thesis demonstrates the value and importance of early DAM; in both an 'ideal' setting and also in a less desirable, time-constrained setting. Applying early DAM and VOI techniques enables researchers to provide relevant conclusions and recommendations to decision makers, who can make informed decisions as to whether a new intervention should be adopted (or rejected), or whether further information is required to help make the decision; as opposed to making decisions based on subjective reasoning. There is considerable merit with employing early DAM for health care research, such as reduced uncertainty, reduction of costs and efficiency gains, however, some drawbacks exist in terms of whether it is always viable to fully exploit VOI analyses, which may hinder widespread support both inside and out-with the health economics community.

Table of Contents

List of Tables

List of Figures

Contributions

This thesis builds on earlier work which developed and proposed an iterative framework for economic evaluation of health technologies (1;3). The iterative framework utilises a variety of methods such as those for building a decision analytic model (4), undertaking probabilistic sensitivity analysis (5;6) exploring uncertainty surrounding the cost-effectiveness decision (7) and employing value of information techniques (VOI) (8-10). Fenwick and colleagues support the iterative approach for economic evaluation, particularly in the context of 'early' pre-trial modelling followed by VOI analysis to help inform future research priorities(11). The iterative approach was demonstrated for the Health Technology Assessment programme in the UK (10) and for the UK reimbursement decision body, NICE (12); gaining some recognition of its benefits within the health economics community, however since this time there have been few other practical applications in published research. In practice, support and financial funding for early stage decision analytic modelling (and full exploitation of VOI techniques) is rare, and in the health care sector economic evaluations still tend to be funded as a one-off exercise alongside a trial to justify reimbursement. Therefore, this thesis builds a case for employing early decision analytic modelling and the use of value of information (VOI) analyses (9;12;13) for undertaking economic evaluations in health care. This thesis uses applied examples to demonstrate the feasibility, practicality and potential merits of using such techniques within the context of the iterative approach to economic evaluation.

Chapters 3 and 4 use a case study of an economic model which was built as part of a NIHR HTA funded project. Researchers in the wider project team (Dr Fay Crawford & Dr Francesca Chappell) undertook the systematic literature review and meta-analysis mentioned in these chapters, however, all of the economic aspects were undertaken by myself. I developed and undertook the economics literature search, developed and populated the economic models, and undertook all the economics analyses with supervision from Dr Elisabeth Fenwick. Clinical advice was given by Professor Malcolm Dunlop and Dr John Brush, the clinical experts on the research team.

In Chapter 5 the sample size calculation was undertaken to inform a research proposal grant submission. Data and clinical advice were provided by Dr Sarah Stock and Professor John Norrie. I developed the decision analytic model and undertook the sample size calculations with supervision from Professor Andrew Briggs.

The EVSI analysis in Chapter 6 was undertaken for this thesis. Technical support was given from Dr Matt Neilson to re-programme the model in the programming language FORTRAN and re-run the simulations. Advice was also given from Dr Claire McKenna, Professor Karl Claxton & Ms Marta Soares regarding the EVSI process.

Publications and Presentations

The following publications were developed as part of this thesis:

Boyd, K.A., Fenwick, E. & Briggs, A.(2010) 'Using an iterative approach to economic evaluation in the drug development process', *Drug Development Research*, Vol. 71, pp. 470 - 477

Boyd, K.A., Briggs, A.H., Fenwick, E., Norrie, J. & Stock, S. (2011) 'Power and sample size for cost-effectiveness analysis: fFN neonatal screening' *Contemporary Clinical Trials*, Vol.32, No. 6, pp. 893-901

Brush J, **Boyd K**, Chappell F, Crawford F, Dozier M, Fenwick E, Glanville J, McIntosh H, Renehan A, Weller D and Dunlop MG. F. (2011) 'The value of Positron-Emission-Tomography (PET) in pre-operative staging of colorectal cancer' *Health Technology Assessment*, Vol. 15, No. 35. Project ref: 08/65/02

The following presentations were developed as part of this thesis and are being drafted for journal submission:

Conference Presentation and paper, 'A critique of decision analytic modelling in cost-effectiveness analyses of smoking cessation services: what makes a good model?' Health Economics Study Group, January 2008, Norwich, UK

Conference Abstract & poster presentation, 'The value of PET-CT imaging for staging in patients with primary colorectal cancer: a pre-trial economic model for cost-effectiveness analysis', 13th Biennial European Conference of the Society for Medical Decision Making, May 31st 2010, Hall in Tyrol, Austria

Conference Abstract & poster presentation, ' Is PET-CT imaging for patients with recurrent colorectal cancer cost-effective?: a pre-trial economic model', 13th Biennial European Conference of the Society for Medical Decision Making, May 31st 2010, Hall in Tyrol, Austria

Acknowledgements

I am extremely grateful to my supervisors Professor Andrew Briggs and Dr Elisabeth Fenwick for their unwavering support and guidance throughout this process. Thank you both for your patience and your invaluable advice. I would also like to express my thanks to Dr Paula Lorgelly who encouraged me to undertake this PhD and provided me with support in the early stages.

I would like to take this opportunity to thank my various colleagues at Glasgow University who have encouraged and supported me. I am particularly grateful to Dr Matt Neilson for his optimism and for the help he provided in re-programming my model, and also to Dr Nicola Roberts for support and comments on an earlier draft.

I must express my appreciation to my friends and family who helped me maintain an optimistic outlook over the last few years. Finally, I thank my husband Gerard for his love, support, belief in me, and for his patience.

Author's Declaration

Chapter 1   This chapter was produced for thesis submission and has not appeared anywhere else.

Chapter 2   An earlier version of this chapter has been published (Boyd et al. 2010). All drafting was undertaken by Boyd, contributions by collaborating authors were in the form of comments on drafts.

Chapter 3   An earlier version of this chapter has been published (Chapter 10 in Brush, et al. 2011). The analysis and text were revised and expanded for this thesis. Analysis and all drafting were undertaken by Boyd, supervised by Fenwick with advice and comments on earlier drafts. Other collaborating authors provided meta-analysis results for use in the economic models (Chappell & Crawford), clinical advice (Brush & Dunlop) and comments on earlier drafts.

Chapter 4   An earlier version of this chapter has been published (Chapter 10 in Brush, et al. 2011). The analysis was revised and expanded for this thesis. Contributions by collaborating authors were limited to comments on earlier drafts.

Chapter 5   An earlier version of this chapter has been published (Boyd, et al. 2011). Data and clinical advice were provided by Stock & Norrie. All drafting and analysis was undertaken by Boyd with supervision from Briggs. The text and analysis was revised and expanded for this thesis.

Chapter 6   This chapter was produced for thesis submission and has not appeared anywhere else.

Chapter 7   This chapter was produced for this thesis and has not appeared anywhere else.

Abbreviations

AUC	area under the curve
CBA	cost-benefit analysis
CCA	cost-consequence analysis
ceCT	contrast enhanced computerised tomography
cePET/CT	contrast enhanced PET/CT
CEA	cost-effectiveness analysis
CLT	central limit theorem
CRC	colorectal cancer
CT	computerised tomography
CUA	cost-utility analysis
DAM	decision analytic modelling
DDP	drug development process
DTA	diagnostic test accuracy
ENB	expected net benefit
EVPI	expected value of perfect information
EVPPI	expected value of perfect parameter information
EVSI	expected value of sample information
fFN	fetal fibronectin
GLM	generalised linear model
HTA	Health Technology Assessment
ICER	incremental cost-effectiveness ratio
NICE	National Institute of Health and Clinical Excellence
NMB	net monetary benefit
PBS	Pharmaceutical Benefits Scheme
PET/CT	positron emission tomography / computerised tomography
QALY	quality adjusted life year
RCT	randomised controlled trial
VOI	value of information
WTP	willingness to pay

1 Introduction

1.1 Rationale

The UK National Health Service (NHS) spends in excess of £100 billion each year providing health care for UK residents (14). New health technologies are continually emerging and even the large public NHS budget is limited; and therefore, there is a need for decision-making and prioritisation of health care.

Throughout the 1990s economic evaluation became a widely used tool in health care to aid decision making regarding value for money in new and existing health technologies (15). Economic evaluation compares the costs and benefits of health technologies in order to aid reimbursement agencies, such as the National Institute for Health and Clinical Excellence (NICE) in the UK, and the Pharmaceutical Benefits Scheme (PBS) in Australia; in determining which alternatives are cost-effective and can be considered to represent good value for money (16;17). Economic evaluations are undertaken for a range of reasons, however their primary role is to inform decision making given limited health care budgets (15;18;19). They provide a systematic approach to comparing alternative drugs or health technologies in terms of their costs and outcomes.

Many economic evaluations that are commissioned by healthcare funding bodies or pharmaceutical companies are often not considered until the later phases of clinical trials (20), and in the case of publicly funded research, explorative research budgets rarely leave adequate financing to incorporate economic analyses. Therefore, economic evaluations are commonly incorporated at the end of large trials in order to support a case for reimbursement. Sculpher et al. (1) suggest that as national health care decision making agencies are increasingly requiring evidence on cost-effectiveness prior to reimbursement, this has triggered a mind-set amongst the health care sector that the function of health economics is to use clinical trial data for economic evaluations in order to support a case for reimbursement. The authors contend (1) that such trial based evaluations

represent a partial or limited form of economic analysis, due to the 'effectiveness' focus of the trial design and other constraints of using a trial designed without economic evaluation in mind. Such a trial may not reflect real life clinical practice.

An iterative approach to economic appraisal is a framework that has been proposed as good practice (1;3) for evaluating health technologies. The framework proposes that the process of health care evaluation should begin with an explorative modelling approach using indicative studies, and progress to more rigorous assessments, updating the model over time as more data becomes available (1;3). The iterative approach to evaluation provides a structure in which evidence from a range of sources can be synthesised and continually updated in order to answer cost-effectiveness decision problems for a defined context and population. The premise is that rather than using economic evaluation as a restrictive, one-off analysis it should be an iterative process conducted throughout the research process.

Decision analytic modelling (DAM) is a key process within the iterative framework. An 'early' DAM, undertaken prior to primary research, allows explorative evaluation of cost-effectiveness based on existing evidence and can be used to assess any uncertainty surrounding the cost-effectiveness decision (21-23). Developing a DAM and undertaking probabilistic analysis at an early stage also enables the use of value of information analyses (VOI) (9;12;13) which is a set of techniques used to determine the amount society would be willing to pay for information, prior to seeking out that information. VOI can help inform research priorities, which is recommended as part of the iterative framework. If developed in advance of primary research, a decision model can enable full exploitation of VOI techniques and therefore help determine whether further research is potentially worthwhile. VOI can help explore the type of research required to address uncertainty in current evidence, and even help design a trial with regards to an appropriate sample size, allocation to arms, data to collect, etc. Despite these advantages, in practice support and funding for early stage decision analytic modelling (and full exploitation of VOI techniques) is rare, and in the health care sector economic evaluations still

tend to be funded as a one-off exercise alongside a trial to justify reimbursement.

The aim of this thesis is to examine the role of early decision analytic modelling for informing research priorities and the design of future studies in health care within the context of an iterative framework for economic evaluation. The thesis will explore the feasibility, merit and drawbacks of undertaking early decision analytic modelling in practice, and consider potential reasons as to why it has not been more widely implemented.

1.2 Economic evaluation in health care

An economic evaluation requires the comparison of two or more alternative interventions, as well as consideration of both the costs and outcomes of the interventions. Rather than a cost reducing exercise, economic evaluation is concerned with the incremental difference between two or more alternatives, i.e. what additional health benefit can we get for what additional cost?   There are various forms of economic evaluation; most commonly cost-benefit analysis (CBA), cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) (15;18). These three forms of evaluation vary in terms of how they measure health outcomes. CBA measures costs and benefits in the same units (monetary) and therefore the cost can be subtracted from the monetarised health benefit of an intervention to determine a monetary gain or loss. CEA measures health outcomes in natural units, such as reduction in blood pressure, life years gained, or number of cases avoided in preventative interventions. CEA results are presented in terms of the incremental cost per unit of health gain, known as an incremental cost effectiveness ratio (ICER), which is calculated by dividing the difference in cost between two interventions by the difference in effect. CUA is often seen as an extension of CEA, whereby health gains are measured in terms of utility, i.e. quality adjusted life year gains (QALYs) and therefore the ICER outcome would be the incremental cost per QALY gained. QALYs are a generic outcome measure and therefore enable comparison of ICERs across disease areas, which is of particular use to decision making bodies

(24;25). Less common forms of economic type analyses include cost minimisation analysis whereby the outcomes are assumed to be identical and therefore the evaluation is simply focussed on minimising cost, or cost-consequence analysis (CCA) where an array of outcomes are presented in tabular form, but not combined with cost (and is therefore often not considered to be a formal evaluation).

CEA is a popular form of evaluation in health care evaluations because disease specific outcomes can be readily modified to include QALYs and they avoid the difficulty of implicit valuation of benefits in a CBA (15). With regards to the ICER outcome, the larger the value of the ICER the more it costs per unit of effectiveness and therefore the less cost-effective the intervention is in comparison to the alternative. The ICER value must be compared against a monetary threshold of maximum willingness to pay (ceiling ratio) per unit of effectiveness to help aid decisions regarding appropriate resource allocation. I.e. if the ceiling ratio (λ) was £50,000 per QALY gained, then an intervention which cost an additional £9,000 (ΔC) and increased QALYs by 0.2 (Δ QALYs), would have a resultant ICER of £45,000 per QALY and would be considered to be cost-effective in comparison to the alternative, at that ceiling ratio. If the ceiling ratio were £20,000 per QALY, then an ICER of £45,000 per QALY would not be considered cost-effective. Equation 1.1 details this decision rule which determines whether an intervention is cost-effective (if the ICER value is less than or equal to the ceiling ratio)

$$\text{ICER} : \frac{\Delta C}{\Delta E} < \lambda \tag{1.1}.$$

A CEA can also report outcomes in terms of the net monetary benefit (NMB) of an intervention. The incremental NMB is calculated by monetarising the measure of effectiveness[1], i.e. incorporating the ceiling ratio explicitly in the ICER calculation to determine the maximum amount that the measure of effectiveness is worth. This is illustrated in Equation 1.2, which is simply a rearrangement of the ICER algorithm detailed in Equation 1.1. The

---

[1] Presenting CEA results as a net monetary benefit (NMB) is different from a CBA. In a CBA the benefits are measured and valued in monetary terms, where as in CEA the benefits are measured in terms of effectiveness, then transformed into a monetary value using a societal willingness to pay value, per unit of that effect.

intervention is cost-effective if the incremental NMB is greater than zero. The decision rule is now:

$$NMB : \lambda * \Delta E - \Delta C > 0 \qquad (1.2).$$

The incremental NMB portrays any inequalities on the cost scale; i.e. the NMB values represent monetary costs. Alternatively the ICER can be re-arranged to express inequalities on the effect scale (whereby the values represent units of effect) using the Net Health benefit (NHB) approach (4;26) as illustrated in Equation 1.3. The decision rule now is:

$$NHB : \Delta E - \frac{\Delta C}{\lambda} > 0 \qquad (1.3).$$

These two alternative decision rules (Equation 1.2 and 1.3) are equivalent to the decision rule based on the ICER, only they express cost-effectiveness in terms of net monetary benefit or net health benefit. These approaches can be advantageous over the ratio led ICER (4), as the net benefit for each individual intervention can be calculated making comparisons with more than two interventions easier. They also have additional advantages over an ICER presentation when dealing with the outcomes from a probabilistic sensitivity analysis (4) and when calculating cost-effectiveness acceptability curves (CEACs) (7).

Cost-effectiveness analyses can be undertaken using a variety of techniques (27). When economic evaluation was emerging in the area of health care, it developed recognition through two methods: decision modelling and trial based evaluations. Evaluations were initially undertaken with a retrospective approach, using decision models to synthesise existing evidence in order to evaluate health interventions; or alternatively, evaluations were undertaken prospectively alongside randomised controlled trials (RCTs). These two methods were viewed as alternative approaches to economic evaluation, with trial based evaluations growing in popularity and establishing a reputation for being more robust than the modelling approach (28). The main critique of modelling was that the analyses relied heavily upon assumptions (29). In the

mid 1990's however, a different perspective emerged, proposing that the two alternative methods should not be viewed as alternatives, but rather, as complimentary approaches (3). The two methods have different purposes; trials aim to estimate particular parameters or the effects of a health care intervention, whereas, modelling provides a framework based on assumptions in which available evidence can be combined. Therefore they are not alternative methods for economic evaluation in which one can be viewed as 'better' than the other, but rather they should be viewed as complementary approaches (18;22). It is possible, and more beneficial, to use a mixture of the two approaches, with prospective trial information supporting wider secondary evidence to provide a more robust analysis (1).

1.2.1 Modelling

Adopting a retrospective outlook, a cost-effectiveness analysis can be undertaken utilising a variety of published data and literature to provide information on costs, resource use, quality of life and effectiveness information which can then be synthesised in a decision analytic model (30). Decision analytic modelling (DAM) in health economics refers to a mathematical decision making tool which structures evidence on clinical and economic outcomes, to inform health care resource use and clinical practices (30). A decision model is a framework for undertaking an economic evaluation, which can be structured in different ways; however, the role of a model is as a means of synthesising evidence on health outcomes and costs from a variety of sources. Brennan & Akehurst (31) discuss the many roles of economic modelling, and how models can vary substantially with regards to structure and complexity, some being nothing more than extended spreadsheet calculations.

Brennan et al. (32) classify various model structures to indicate the range of modelling approaches and their structural relationships to one another. The authors also provide some guidance on choice of model structure, and highlight that different modelling approaches can produce very different results. The choice of model depends on various factors but is predominantly determined by the decision makers' requirements, the complexity of disease area, and even modellers expertise or preference (32).

In the mid 1990s the discretionary nature of modelling was used to question its credibility (28;29), and since then considerable effort has focussed on promoting consistency in the process of modelling, particularly with regards to developing general principles for identifying and synthesising evidence. Sculpher et al. (33) considered what constitutes good practice in modelling, along with issues around validity and quality in modelling. In 2003 the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) published guidelines for conducting and reporting economic models in response to issues of appropriateness and consistency (30). Philips et al. (34) continued to explore good practices in modelling and commented that despite the prevalence of various guidelines for the assessment of decision models since 1985, there was still no standard definition of what good practice should mean in an economic model. In response to this Philips et al. (34) continued earlier work by Sculpher et al. (33) and attempted to provide such a framework. The authors reviewed and consolidated the range of existing guidelines and developed a standard framework for how 'good practice' should be defined for DAM. The framework was developed around three key areas: structure, data, and consistency; and provides a practical and systematic means for critical assessment of decision analytic models. The authors caution that it is unrealistic to expect all studies to adhere to every point in the framework, but stress that systematic application of the framework will allow the strengths and weaknesses of models under review to be identified.

More recently, Petrou & Gray (35) published a review in the British Medical Journal to highlight issues relating to the design, conduct and analysis of economic evaluations using DAM and to create awareness in the wider medical community. As computing capacity has increased over the last two decades, so have modelling methods (36) and therefore guidelines to promote good modelling methods have also had to adapt and be updated as more complex forms of modelling become ever more popular (37).

1.2.2 Trial based

Throughout the last two decades trial based economic evaluations have become fundamental as health care decision making bodies have specified a

requirement for evidence of cost-effectiveness to support reimbursement decisions. Therefore economic evaluations have been incorporated into randomised controlled trials for pharmacologies and medical therapies (28;38). This type of evaluation is undertaken prospectively, utilising effectiveness data directly from a trial, and therefore can potentially collect resource use, cost and quality of life data directly related to the health technology under evaluation.

Clinical trials are frequently viewed as the 'best vehicle' for economic evaluations as they ensure internal validity, through prospective collection of patient specific data, and they also present the opportunity to collect 'additional' economic data (such as cost and utility information) at a low marginal cost alongside a large scale trial. Therefore, large trials have many potential benefits for conducting economic analyses if the economic component is integrated into the study protocol appropriately, rather than added on as an extra component at a late stage. It is important that the economic requirements contribute to the study design to ensure that the structure of the trial will provide the necessary data for a high quality economic study (39). Incorporation of economic relevant data into the trial dataset enables statistical analyses on the cost, effect and quality of life data, rather than just the effectiveness data. Glick and colleagues (40) set out the methodology for undertaking economic evaluations alongside clinical trials, including details for analysing the data, calculating within trial ICERs and exploring uncertainty in the cost-effectiveness outcomes.

Glick et al. (38;40) specify some 'gold standard' characteristics for conducting an economic evaluation as part of an RCT, detailing six steps for an appropriate analysis. Despite such attempts to promote consistency and improvements in the methods for the design, conduct and analysis of economic data collected alongside clinical trials over the last two decades, there remains a great deal of variation in methodology and reporting of these types of studies (39). In response to this the ISPOR task force developed guidelines for reporting and conduct of economic evaluations alongside trials, in an attempt to improve quality and enhance the credibility of cost-effectiveness analysis (39). More recently Petrou & Gray (41) published a review of the methods for

research and reporting of evaluations alongside RCTS (alongside their review for DAM (35)) to highlight good practice and create awareness in the wider medical community.

Despite the various guidelines, integrating economic evaluations into trial protocols remains challenging (42) and as the primary focus of the trial research is typically on effectiveness (as opposed to cost-effectiveness), it is likely the sample size will be determined considering only the effectiveness outcome. This can potentially result in a trial that is underpowered to detect cost-effectiveness (43;44). Additionally, trials alone rarely provide sufficient evidence to fully inform an economic evaluation (38;39). A single trial may not compare all relevant alternatives, may not collect important quality of life data, or may have a restricted follow-up time period which is inadequate. For example a lifetime horizon follow-up is appropriate for an economic analysis but the cost and feasibility of conducting an RCT in practice restricts follow-up periods, with three to five years being considered a long duration. Such restrictions necessitate extrapolation of trial data over longer time periods, typically using some form of modelling technique. Therefore economic evaluations alongside trials are rarely undertaken solely on the trial data. External information for economic modelling is required, particularly for extrapolation (28) and therefore decision analytic modelling is used to incorporate additional supplementary evidence for the economic evaluation.

Throughout the last two decades the role for cost-effectiveness analysis in health technology assessment has developed substantially (18) and while some may still hold to the view that clinical trials are the 'best vehicle' for economic evaluation, it is recognised within the health economics community that a mixture of trial based and modelling approaches is required for a robust analysis (1;39;41). Regardless of whether a modelling, trial based or mixed approach is used to undertake an economic evaluation, various good practice guidelines and checklists exist (15;30;34;37;39;45) as well as various decision making body guidelines (24;25); which promote a systematic and rigorous approach to data collection, analysis, synthesis of cost and effect data/evidence, reporting of outcomes and exploration of uncertainty and heterogeneity.

1.2.3 DAM in economic evaluation using an iterative approach

In exploring a framework for quality assessment in DAM, Sculpher et al. (33) explain that models and trials have a separate purposes; the purpose of a model is to combine all available information at the time (t) the decision must be made to predict the impact. The purpose of a trial based economic evaluation is to generate new information about one or more parameters at time t+1 that was not available at time t. Therefore, rather than be viewed as alternative methods for economic evaluation, they should be viewed as complementary approaches at different time points. They can both be used to fully inform a decision; i.e. in practice a short-term trial based analysis is often followed by a longer term decision model utilising other evidence to extrapolate the trial findings into a lifetime horizon, providing a more robust analysis. However, rather than merely supporting a mixed methods one-off approach to evaluation, Sculpher and colleagues (3) outline an iterative process for economic evaluation, progressing from early indicative studies, to more rigorous assessments as more data becomes available, reducing uncertainty surrounding the cost-effectiveness estimates over a long time period. Bayesian techniques for economic evaluation developed and promoted throughout the 1990s; such as probabilistic sensitivity analysis for exploring uncertainty in parameter estimates and the model (5;46), and value of information analyses (VOI) (9;13) are ideal tools for facilitating such an iterative process.

Sculpher et al. (3) consider the best way to combine economic and clinical research. They highlight that economic evaluation is perceived as a one-off exercise and attempt to dispel this perception through proposing an iterative process. Claxton promotes an iterative approach through a Bayesian framework for decision making, exploring the uncertainty surrounding a cost-effectiveness decision based on current information (9;13). The authors propose formal VOI assessments to value further research on its potential ability to reduce the expected costs of uncertainty surrounding the decision, rather than deciding on further research and its design through arbitrary means.

Fenwick et al. (21) further support the role of iterative decision analytic modelling, suggesting that as new health technologies emerge, DAM should be undertaken and updated regularly as new information becomes available. The authors demonstrate the application of this process and propose the case for integrating such a framework into the HTA process. The iterative approach recommends a process of 'early' modelling prior to the design of any clinical trials (7;47;48), whereby a decision analytic model is developed based on existing evidence on the costs and effectiveness of an intervention. This provides a preliminary assessment of the level of uncertainty surrounding the costs, effects and cost-effectiveness of the intervention and which can then be used to provide information on whether a trial is required, and if so the appropriate design using VOI methods.

Some researchers have used this approach in practice, however it was not widely adopted and the one-off attitude towards economic evaluation highlighted by Sculpher et al. (3) in 1997, prevailed in economic evaluations undertaken alongside clinical trials (28). The iterative approach was demonstrated for the Health Technology Assessment programme in the UK (10) and for the UK reimbursement decision body, NICE (12); gaining some recognition of its benefits within the health economics community, however since this time there have been few other practical applications in published research.

Fenwick and colleagues continue to support the iterative approach, particularly in the context of 'early' pre-trial modelling followed by VOI analysis to determine whether a trial was worthwhile (11); and in 2006 Sculpher et al. (1) set out a framework for the iterative process, defining it in to five distinct stages. In 2007 Briggs provided further support for the iterative approach, highlighting in the British Medical Journal that VOI must be integrated into the process of commissioning primary research (49). However, despite these latter developments, there is little evidence of funding and support for an iterative approach to economic evaluation in the health care sector. There remains a need for wider acknowledgement and incorporation of the iterative approach in the health care community, particular amongst the range of health care research funding bodies and the pharmaceutical industry.

In practice, support and funding for early stage decision analytic modelling (and full exploitation of VOI techniques) is rare, and in the health care sector economic evaluations still tend to be funded as a one-off exercise alongside a trial to justify reimbursement. Therefore, the aim of this thesis is to examine the role of early decision analytic modelling for informing research priorities and the design of future studies in health care, within the context of an iterative framework for economic evaluation. The thesis will explore the feasibility, merit and drawbacks of undertaking early decision analytic modelling in practice, and consider potential reasons as to why it has not been widely implemented.

1.3 Thesis outline

The thesis is split into five main chapters.   Chapter 2 introduces economic evaluation in the context of publicly and commercially funded health care research. The chapter outlines the methodology used in building a decision analytic model, undertaking probabilistic sensitivity analysis, analysing uncertainty, and using value of information techniques to explore potential future research concerns. These methods are used throughout the remainder of the thesis. The iterative approach to economic appraisal is discussed and the advantages of employing such an approach to funding health care research are highlighted. This chapter supports the case for early stage decision analytic modelling in health care research within the context of an iterative economic framework.

Chapters 3 and 4 demonstrate a case study of good practice in developing an early decision analytic model. The case study details a piece of research that was funded by the National Institute for Health Research Health Technology Assessment programme (50) to build a decision model and undertake an economic evaluation of an emerging technology. This research remit corresponds with the early stages of the iterative economic approach, developing an early decision analytic model to undertake an economic evaluation to determine whether the technology is cost-effective given current

information. Chapter 3 details the development and parameterisation of three economic models, reporting the deterministic results, while Chapter 4 reports the probabilistic results from the models and explores uncertainty and VOI analyses to help inform future research priorities. Chapter 4 demonstrates the importance of exploring uncertainty and how the application of value of information analyses, with early decision analytic models can help to inform the researchers, decision makers and funders on the appropriate next steps to take.

Chapter 5 demonstrates the practicalities of designing clinical trials from an economic perspective. The chapter offers a case study, reporting on an atypical real life case in which a clinical and cost-effectiveness trial was designed and powered to determine the sample size with regards to cost-effectiveness. The desire for a trial had been decided upon without prior economic involvement, i.e. neglecting the iterative economic approach, however, even when the iterative approach has been overlooked, involvement in the conventional design of a trial from an economic perspective is still desirable. This chapter details the development of a simple DAM to inform a sample size calculation for the trial as the research proposal was being prepared. This chapter therefore demonstrates the feasibility of developing decision models without prior funding and how simple models can be used to help design a trial and inform calculations in a real life setting. The chapter simultaneously highlights limitations with undertaking sample size calculations and demonstrates the role for economic considerations to guide non-inferiority margins.

Chapter 6 follows on from the previous (Chapter 5) case study considering whether the same, or an alternative conclusion would have been drawn had it been possible to adopt an iterative economic approach for the study; i.e. if pre-trial modelling had been undertaken in advance to help set research priorities, enabling the use of expected value of perfect information (EVPI) analysis to explore whether further research was of potential value, and if so using expected value of sample information (EVSI) to estimate an appropriate sample size for the RCT.   The chapter undertakes a re-analysis of the study from Chapter 5, adapting it in line with the iterative framework to compare

the resultant research priorities and sample size requirements under an iterative approach with the outcomes that were determined using the conventional framework from the trial proposal in Chapter 5.

Chapter 7 provides a summary of the thesis. This final chapter also proposes areas for further research and offers the overall conclusions.

2 DAM for economic evaluation in health care

2.1 Introduction

The UK National Health Service (NHS) spends in excess of £100 billion each year providing 'free' health care for UK residents (51). The NHS is the largest publicly funded health care service in the world but is subject to an annual budget and therefore is vulnerable to the two fundamental economic concepts of scarcity and choice. Decisions must be made regarding the choice of which health care products and services to offer, given the limited financial budget. Prioritisation of health care requires decisions or decision-making regarding which illnesses and disease areas should be allocated resources, and within specific disease areas, which treatments, interventions and health services to offer. Therefore, there is a need to decide on the best mix of health technologies, especially as new technologies become available. The term 'health technology' covers a range of methods used to promote health, prevent and treat disease and improve rehabilitation (50). It is often used to represent all health care goods and services, such as pharmacologies, surgical techniques, intervention packages, or technologies. Economic evaluation is a means of assessing alternative health technologies by comparing their costs and health benefits, and is increasingly being used as a major input for decision making bodies and reimbursement agencies throughout various countries for determining which interventions and health technologies represent good value for money (16;17).

Economic evaluations are undertaken for a range of reasons, however their primary role is to inform decision making given limited health care budgets (15;19), through providing a systematic approach to comparing alternative drugs or health technologies in terms of their costs and consequences. Economic evaluations are not used for informing decisions regarding individual patients, but rather, they are undertaken to help guide drug formularies, disease-specific guidelines, the development of patient services, justification for existing or new services and also drug reimbursement decisions (19). In the UK, decision making bodies for health technology reimbursement, such as the

National Institute for Health and Clinical Excellence (NICE) and the Scottish Medicines Consortium (SMC), require economic evaluation as part of their decision making process (24;25).

There is a formal requirement for economic evaluation to be an integral component of NHS research and development funded projects (52). However, health care research from other publicly funded research bodies and within the commercial sector is not subject to such requirements and therefore, may or may not include economic components within their research.

The National Institute for Health Research (NIHR) is an NHS funded research body who support various research programmes (52), such as the Public Health Research (PHR) programme and the Health Technology Assessment (HTA) programme (50). The HTA produces independent research about the effectiveness, cost-effectiveness and broader impacts of health care treatments and technologies. The research generated is intended for an NHS audience, particularly those who plan and provide care in the NHS. They commission both primary research and evidence synthesis, and the research topics are identified and prioritised to meet the needs of the NHS. The HTA identify evidence gaps or decision problems, for which they commission research, through various means such as consultation with key NHS stakeholders, extracting research recommendations from various resources, direct recommendations and communication with researchers (50).

The Medical Research Council (MRC) is another publicly funded health care research body, and is one of seven research councils in the UK who are responsible for investing money in research in the UK (53). The MRC aims to support medical research in order to 'increase the pace of transition to better health'. They set their research priorities in order to support scientists and deliver improved health outcomes across society. The MRC funds research through a range of grants, commissioned calls, and fellowships and have a range of board panels and groups to decide on what research should be funded or not (54). The Chief Scientist Office (CSO) is a Scottish government funded research body who support and promote research aimed at improving the quality and cost-effectiveness of health services and health care offered by

NHS Scotland (55). The CSO funds research through a variety of research grants and fellowships. It is commonly the case that health care funding bodies such as the HTA, MRC and CSO define an overall research question and commission research (normally through a call to researchers) on specific health care topics to meet their organisations priorities; however, some funding bodies also issue open calls with an overall theme or disease area, which encourage researchers to identify specific decision problems within that broad area which need to be addressed.

In the pharmaceutical industry, research decisions are undertaken internally, and drug development follows a well known and formally recognised process consisting of four distinct phases of clinical assessment (51;56-58). Within each phase rigorous testing is undertaken in order to reduce uncertainty, and ensure that only the safest and most marketable drug will proceed to market.

This chapter introduces economic evaluation in the context of public and commercial health care research. An overview of economic evaluation in the health care sector is provided followed by an outline of the methodologies used in decision analytic modelling for an economic evaluation in health care. The methodologies include building a decision analytic model, undertaking probabilistic sensitivity analysis, analysing decision uncertainty, and using value of information (VOI) techniques to explore potential future research concerns. Following the decision modelling methodology, section 2.9 introduces the iterative approach to economic appraisal as an appropriate framework for undertaking research in the health care sector. The advantages of employing such an approach to funding health care research (both private and public) are highlighted, in terms of reducing uncertainty, reduction of costs and efficiency gains. This chapter supports the case for early stage decision modelling in health care research within the context of an iterative economic approach.

2.2 Economic evaluation in health care

Economic evaluations are undertaken to help inform decision making, for example to help determine which drugs or health technologies to fund given limited health care budgets (15). As discussed in Chapter 1, cost-effectiveness analyses can be undertaken using a variety of techniques (27), most notably using decision analytic modelling, trial based evaluations, or a combination of the two. It is this latter approach which utilises prospective trial information to support wider secondary evidence to promote a robust economic analysis (1).

Reimbursement agencies across the world, such as the Australian Pharmaceutical Benefits Scheme (PBS) and the National Institute of Health and Clinical Excellence (NICE) in the UK, require evidence on cost-effectiveness for their reimbursement decisions. In response to such requirements, the pharmaceutical industry has added a 'fourth hurdle' (59) to their drug development and licensing process; while publicly funded health care research has begun to incorporate economic components in large scale (Phase III) studies and occasionally in Phase II studies. In addition to providing evidence of quality, safety and effectiveness, value for money must also be demonstrated.

Numerous countries have introduced reimbursement policies over the last two decades, requiring some form of economic analysis to support the introduction of new drugs, (and some also have the same requirements in support of new health care technologies or interventions (17)), however, the stringency and rigour of guidelines for undertaking the analyses vary from country to country (16;17). Many economic evaluations that are undertaken or commissioned by pharmaceutical companies are often not considered until the latter phases of clinical trials (20) and in the case of publicly funded research, Phase II and explorative research budgets rarely leave adequate financing to incorporate economic analyses. Therefore, economic evaluations are commonly incorporated at the end of a larger scale phase III or phase IV trials utilising effectiveness, quality of life and cost data from the trial in order to support a

case for reimbursement. Large scale Phase III trials are designed to show evidence surrounding the clinical effectiveness of a new intervention in comparison to a control to support the case for regulatory submission, and the economic component is often 'piggy-backed' on to the end of the trial (28).

Sculpher et al. (1) suggest that reimbursement agencies' requirements may have triggered a commercial mind-set that the function of health economics is to use RCT data for economic evaluations in order to support a business case for reimbursement; however, some agencies such as NICE in the UK require a more comprehensive approach using decision analytic modelling to synthesise trial information with other evidence (24). An approach in which decisions are based upon economic evaluations which only incorporate RCT data can be considered short-sighted with regards to the purpose of and approach taken towards economic evaluation. Sculpher et al. (1) argue that such trial based evaluations represent a partial or limited form of economic analysis, due to the 'effectiveness' focus of the trial design and other constraints of using a single trial designed without economic evaluation in mind. Such a trial may not reflect real life clinical practice, costs may not be the same, it may not consider all relevant alternatives, may have a limited time horizon, lack relevance to the decision context, or in general may not adhere to the gold standard characteristics for economic evaluation within a trial (40).

A more comprehensive approach to economic evaluation involves using decision analytic modelling to synthesise evidence from a variety of information sources followed by an exploration of uncertainty, including uncertainty surrounding the cost-effectiveness decision.

2.3 Decision Analytic Modelling

Decision analysis represents a technique for structuring decision problems. It is an approach that has been used in a range of disciplines such as Engineering, Law, Business (60), and more recently environmental remediation. It has also been adopted in health care as a framework for making decisions under uncertainty. A succinct definition is provided by Snider et al. (page 27) (61):

> "Decision analysis is an explicit, quantitative and systematic approach to decision making under conditions of uncertainty"

Within health care, decision analysis has been adopted as a framework for conducting health care evaluations and economic evaluations (4;61). An economic evaluation can be designed using a decision model, enabling a wide range of data on costs and effects to be synthesised within the model in order to derive cost-effectiveness outcomes. Therefore, in the context of economic evaluation, a decision analytic model (DAM) is a mathematical tool used to structure a problem regarding two or more interventions for a specific patient group. The model uses mathematical relationships to synthesise input information (such as clinical and cost inputs) from numerous sources and defines a series of possible consequences depending on the options being compared (4).

The basic steps for building a decision analytic model involve specifying the decision problem, defining the boundaries, specifying the model structure, identifying and synthesising evidence and dealing with uncertainty. Assessing the value of additional research is also a key component of the modelling process (4), given that the point in decision analysis is to inform decision making given uncertainty. These key components are now discussed.

2.3.1 Specifying the problem

This stage involves considering the objective of the evaluation, so as to clearly define the relevant aspects of the question to be addressed by the analysis. The alternative interventions or health technologies being evaluated need to

be specified, which is likely to be the new intervention compared to a control or current practice, however, the evaluation may involve more than two alternatives. The outcome measures should be defined, i.e. what will be the primary measure of effectiveness. Additionally the relevant disease area and patient population should be specified. Other aspects to be included are the setting (i.e. primary care) and the perspective of the analysis. The perspective will tend to be determined by who is funding the research (i.e. an NHS funded research project may request a NHS perspective as opposed to a societal perspective which could incorporate 'indirect' costs incurred by patients and carers).

2.3.2 Defining boundaries

This stage involves considering what is relevant and not relevant to be included in the analysis, i.e. what time horizon is relevant to the model. Should the outcomes be modelled over the patients' lifetime or is a shorter duration more appropriate to answer the research question? This will relate to the outcome of interest, for example, is it life expectancy or quality adjusted life years (QALYs) that is the effectiveness endpoint of interest, or is it an outcome which requires a shorter duration, i.e. cases of morbidity post-surgery? Defining boundaries for the model also relates to the potential impacts of the interventions under evaluation, and whether they need to be included in the model. For example is it relevant to incorporate negative side-effects from the interventions? If the interventions under evaluation involve treatment with drugs, are there any possible side-effects and if so, are they relevant to incorporate in the analysis? Researchers should ask, 'do they impact on the costs or patient quality of life within the time horizon relevant to the model?' In some cases side-effects may be negligible and not expected to have a large impact on the patient population; or alternatively may be longer-term impacts out-with the relevant model timeframe. In such cases it may be appropriate to consider such impacts to be out-with the boundaries of the model.

Defining boundaries ensures that the model is representative, and an appropriate (yet simplified) reflection of real life. It is important that the

model adheres to quality assurance in that the boundaries are not so restrictive that they bias the analysis by excluding important factors.

2.3.3 Structuring the problem

An appropriate structure for the model should be determined based on the problem specification and model boundaries already identified. A model can be as simple or as complex as required, as long as it is fit for purpose.

Decision Trees are a common structure used for simple analyses or those with short time horizons. A decision tree is a model which maps the patient pathway, assigning cost and outcomes to alternative pathways or 'branches' throughout the tree (36). Decision trees are popular due to their simple structure, and the transparent nature of the tree which makes them useful for short term analyses; however, they are less valuable for modelling complicated disease conditions or over long time durations as they can become 'bushy' very quickly if there are numerous decision options (decision nodes) at different stages in the model, and they lack an explicit time variable (35).

Markov models are more useful for analyses over a longer duration, involving transitions between various health states and outcomes over time (62). The main limitation with Markov models is that they do not account for the history of progression in the model. This is known as the Markovian assumption (63) and means that transition probabilities between health states are independent of the history of the patient and depend only on the current health state. Despite this limitation, Markov models remain a common structure for modelling lifetime outcomes, however, in recent years more complex forms of modelling have become ever more popular (37).

Over the last decade Discrete Event Simulations (32;36;64) have grown in popularity, simulating the progress of individuals through a health care system. The nature of DES allows complex modelling, memory of patient history in the model and interaction between patients over long time durations, thereby overcoming some of the restrictions of Markov modelling; however they require specialist software and programming skills to develop and run (36). Dynamic

modelling is an alternative method of complex modelling which allows for impacts to affect the behaviour of an entire population or an entire health system, accounting for evolving factors and effects over long time durations. These traits have made dynamic models popular for modelling infectious diseases. Although these more complex modelling methods require programming skills, they have considerable value to researchers and therefore encourage improvements in researchers' computational skills (35;36).

The choice of model structure depends on various factors but they are predominantly determined by the decision makers' requirements, the complexity of disease area, and even modellers expertise or preference (32). The type of model used is subjective to the modeller, but as different modelling approaches can produce very different results it is important for researchers to be explicit in the reasoning for choice of model structure.

2.3.4 Synthesising evidence

This stage involves the systematic combination of evidence from a range of sources in order to populate the decision model. There needs to be a systematic approach taken to identifying all the relevant evidence for the model. The model input evidence is classified into parameters which are used to reflect the data on clinical events, effectiveness, utility information, resource use, and unit cost information.

Clinical parameters tend to be incorporated as probabilities, i.e. the probability of a clinical event occurring (where a probability is a number between zero and 1 which reflects the likelihood of a specific event occurring) such as disease prevalence to define whether the population has a specific disease or not. An example of parameters represented by probabilities could be for patients with a specific disease that has three alternative treatment options: A, B or C. The probability of receiving one of the three treatments (A, B or C) is a number between zero and 1 which reflects the likelihood that a patient will receive one of these treatments rather than the other two. I.e. there is a probability of 0.3 that a patient will receive treatment A, a 0.6

probability of receiving treatment B and a probability of 0.1 that treatment C will be given.

Some parameters may be reported as rates rather than probabilities and therefore it is important to note the distinction between the two. A rate is the instantaneous potential for the occurrence of an event, expressed per number of patients at risk (4), whereas a probability is the likelihood of an event occurring over a specific time period. Government statistics tend to report population level mortality rates which can be useful in economic models, however they may require to be converted into transition probabilities. It is possible to convert a rate to a probability over a specified time period, if the rate can be assumed to be constant over that time period: (4). Equation 2.1 details how the probability (p) can be calculated given the instantaneous rate (r) and time period (t) of interest for which the rate is assumed to be constant:

$$p = 1 - \exp\{-rt\} \hspace{4cm} (2.1).$$

The probabilities and other parameters (effectiveness, survival, utility, resource use) in the model are assigned from various sources such as medical literature, expert opinion, and statistical analysis. For example, utility for the patient population may be derived from a survey or from an observational study; the probability of different treatment options for a specific disease may be informed by a combination of published disease guidelines and clinical expert opinion, while effectiveness data may be derived from one or more clinical trials.

The synthesis of data from various sources gives rise to potential issues, such as no direct comparisons from RCTs for the interventions of interest, different follow-up times from different studies and heterogeneity (where there is a difference in patient characteristics between trials). Such issues are dealt with through a variety of methods including indirect and mixed treatment comparisons and meta-regression (65;66).

Once the model has been developed and populated with evidence relating to the mean value for each of the model parameters, the economic evaluation is

undertaken to calculate the outcomes, which in a cost-effectiveness analysis would be the incremental cost-effectiveness ratio or the incremental net benefit of the new intervention(s) in comparison to the alternative(s). This is a basic 'deterministic' analysis which uses point estimates to represent the evidence for the parameters in the model. The next step involves handling uncertainty in the model, and in particular the method used to deal with uncertainty in the model parameters.

2.3.5 Handling Uncertainty

Regardless of whether an economic evaluation is undertaken based on a decision analytic model, or informed solely from a single clinical trial, it will be subject to uncertainty. Han et al. (67) highlight that there is a wide variety of uncertainty referred to in health care research, spanning different types, sources and manifestations of uncertainty. The authors propose a conceptual taxonomy of different types with the intention of encouraging an organised approach to dealing with uncertainty (67). With regards to decision modelling, uncertainty can pertain to sampling variation, heterogeneity, methodological uncertainty, structural uncertainty, parameter uncertainty and decision uncertainty; each of which requires to be dealt with differently(4-6).

In clinical studies, variability in individual patients is normally addressed through randomisation and analysis of baseline statistics. Sensitivity analysis can be undertaken with access to patient level data, but in a model where the data has been derived from published evidence, this may or may not have been addressed by those reporting the evidence, and it cannot be addressed through collection of more data (4).

Assessing heterogeneity requires consideration of various study population subgroups to assess whether other factors inherent to these groups influence the study outcomes by confounding or overriding the actual treatment effects, i.e. different age groups, differences in gender or disease specification. The National Institute for Health and Clinical Excellence (NICE) formally recognised the importance of heterogeneity and subgroup analysis in their 2004 Health Technology Appraisal guidance. More recently this guidance was revised and

now specifically recommends subgroup analysis in all technology appraisals submitted to NICE for consideration (24).

Methodological uncertainty refers to uncertainty regarding whether the methods used were the most appropriate. Methodological uncertainties can be dealt with through the use of a benchmark approach or reference case for appropriate methodology, i.e. by following good practice guidelines for undertaking modelling (34;37). The generalisability of model results can be explored by altering parameters in the model that may have been specific to a particular setting. Uncertainty regarding the structure of the model (68) can be dealt with by one-way sensitivity analyses and scenario analyses, modifying one or more structural aspects or assumptions of the model and determining the impact on outcomes. More recently it has been suggested that a formal framework is required to addressed structural uncertainty, whereby a global model could be developed including parameters which encompass all possible structural choices so that they can be addressed using probabilistic methods(69).

Parameter uncertainty refers to uncertainty in the point estimates used to reflect the specific parameters in the model, i.e. uncertainty in the mean utility value assigned to a specific disease group, or uncertainty in the probability of an event. Parameter uncertainty can be dealt with deterministically through univariate and multivariate sensitivity analysis, however, probabilistic sensitivity analysis (PSA) can also be undertaken to adequately address this, which is discussed in the next section. Finally, decision uncertainty should also be explored. Decision uncertainty refers to the level of uncertainty in the cost-effectiveness outcome from the model, i.e. uncertainty in the estimated cost-effectiveness of the treatment in comparison to the alternative(s) (70;71). Parameter uncertainty is now discussed in more detail followed by decision uncertainty.

2.4 Assessing parameter uncertainty

Parameters in a model are represented by point estimates, which reflect the mean value for each parameter and cost-effectiveness is determined based on the mean value for each of the parameter inputs. This is known as a deterministic analysis. In a probabilistic sensitivity analysis (PSA), uncertainty in the mean parameter estimates is accounted for by assigning a distribution to each parameter and drawing a random estimate from that distribution to represent the point estimate. By drawing randomly from the parameter distributions simultaneously for all parameters in the model and repeating this random draw numerous times (in a Monte Carlo simulation) uncertainty in the parameter estimates is accounted for and the resultant cost, effect and cost-effectiveness outcomes from the model can be calculated for each simultaneous random draw (iteration) (4-6). The average cost, effect and cost-effectiveness across all the iterations in the Monte Carlo simulation are taken to represent the probabilistic outcomes which account for uncertainty in the input parameters.

Using any number of iterations greater than 1000 is generally considered to be acceptable (4) to reflect uncertainty in the model parameters, however it is wise to test the stability of results over a varying number of iterations, i.e. 1000, 2000, 5000, and 10,000 iterations. If there are negligible or minimal changes to the incremental cost and incremental effectiveness outcomes between variations in the number of iterations then the outcomes can be considered stable; however, if there is variation in outcomes between the different number of iterations (i.e. between 1000, 2000 and 5000 iterations), then a greater number of iterations may be required, such as 10,000, and 20,000 before the results can be considered stable. The time taken to run the Monte Carlo simulation will depend on the complexity of the model, the computing software used and also on the number of iterations specified for the Monte Carlo simulation. As the number of iterations increases, the time taken to run the simulation will also increase, so it is important to find a balance between checking the stability of PSA results and needlessly running very time

consuming simulations; for example checking results at 100,000 iterations is unnecessary if they were found to be stable at 10,000 and 20,000 iterations.

In making parameters probabilistic, there are a variety of distributional forms to choose from. Briggs and colleagues (4) explain that the type of parameter and its method of estimation provide an indication of the appropriate distributional form to use. Let us first consider the different distributional forms available, followed by a discussion of fitting parameters to the forms.

2.4.1 Distributional forms

A distribution shows how the total probability of a random variable (i.e. the parameter of interest) is distributed. This will vary depending on the family or form of distribution (72), each of which will reflect different points on a scale. Continuous parameters (such as cost) require a continuous scale and therefore a distributional form that is appropriate to capture that scale. Likewise non-continuous parameters (such as a proportion, which will fall between zero and 1) will require a distributional form that lies on a scale bound between zero and 1.

Normal (Gaussian) distribution

The most commonly used continuous distribution is the normal distribution, often referred to as the Gaussian distribution (72). The Standard Normal Distribution has an expected mean value of zero, and a variance of one $N(0,1)$, however, a random variable (parameter) from the normal distribution is capable of assuming any value between negative infinity and infinity ($-\infty$, ∞). Specifying the Normal distribution requires a mean value of the parameter of interest and the standard deviation (73).

Many distributions can be approximated to the normal distribution (73), under an assumption based on the central limit theorem (CLT) which assumes that as a sample size gets very large (tending towards infinity) the sampling distribution of the mean will be normally distributed regardless of the underlying distribution of the data (4). Therefore, in a decision analytic

model, the normal distribution can be considered a candidate for any of the parameters in the model; however, this is only so long as the data informing the parameter is of a large enough sample size to justify a normal assumption.

Lognormal

The Lognormal distribution is a continuous probability distribution of a random variable (parameter) whose logarithm is normally distributed. The lognormal scale is confined from zero to infinity (0, ∞), and therefore variables on this scale cannot have negative values. The lognormal distribution is represented by the parameter mean (ų) and standard deviation (σ). Lognormal distributions are appropriate for parameters that are non-negative, highly skewed or have a multiplicative format, i.e. ratios, and as they are on the log scale, they enable transformation to and from the normal scale.

Gamma distribution

The Gamma distribution is another continuous distribution, which is on a scale from zero to infinity (0, ∞), and therefore allows for data that has a positive right skew. Gamma distributions are defined by two parameters: a shape parameter (α) and a scale parameter (β). The Gamma distribution is represented by Gamma (α, β).

Beta (Binomial) distribution

The Beta distribution is characterised by two parameters: alpha (α) and beta (β) such that the Beta distribution is represented by Beta (α,β). Alpha is considered the lower bound (location parameter) and beta the upper bound (scale parameter). The standard Beta distribution is constrained on the interval zero to one (0, 1) (where α =0 and β=1), and is therefore appropriate for binomial data, such as parameters represented by proportions. However, the Beta distribution is not necessarily binomial; the distribution depends on the value of the scale (β) parameter (74).

Dirichlet distribution

A Dirichlet distribution can be considered as the multinomial equivalent to the Beta distribution (75). It is used to represent multinomial data where there are numerous categories, each represented by a proportion, of which the proportions sum to 1. Therefore, the Dircihlet distribution is characterised by a vector of positive real values (α), such that the distribution is represented by Dir(α).

2.4.2 Fitting parameters to distributions

Briggs et al.(4) explain that the type of parameter and its method of estimation provide an indication of the appropriate distributional form for each parameter in a decision model. A Bayesian approach is recommended (6;76) whereby the choice of distribution is based on prior beliefs about the parameter. Therefore the characteristics of the parameter of interest should be considered when deciding upon an appropriate distribution. For example, with cost parameters, the minimal value a cost will be is zero (as it is not possible to have a negative cost) but the cost could potentially range up to infinity, in which case it would be appropriate to consider either gamma or a lognormal distribution for cost variables, based on the prior beliefs about the cost parameter characteristics. Likewise where the parameter of interest is represented by a probability (and the parameter is binomial in nature), such as a probability of having a disease, the probability is bound on an interval between zero and 1, and therefore, based on this prior knowledge of the parameter characteristics, it is appropriate to assume a Beta distribution, which is also bounded on the 0-1 interval. Where parameter estimates are derived using multivariate logistic regression, these can be represented by a lognormal distribution. By following the standard distributional assumptions and considering the characteristics of the specific parameters in the model, there are only a few possible candidate distributions for most parameters in a model.

Table 2-1 details a list of common parameters used in decision analytic models (DAM) and the distributional forms that are commonly used to represent them.

Table 2-1: Typical parameters and their distributional forms in a DAM

Common Parameters	Possible Distribution	Distribution scale
Costs	Gamma	(0, ∞)
	Lognormal	(0, ∞)
	Normal	(-∞, ∞)
Utilities	Beta	(0, 1)
Disutilities	Gamma on disutility	(0, ∞)
	Lognormal on disutility	(0, ∞)
Prevalence	Beta	(0, 1)
	Dirichlet	(0, 1)
Treatment Probabilities	Beta	(0, 1)
	Dirichlet	(0, 1)
Mortality Probabilities	Beta	(0,1)
Relative Risks	Lognormal	(0, ∞)

Examples of parameters which are commonly represented by Beta distributions are prevalence parameters (to represent the proportion of patients who have or do not have a disease) and diagnostic test accuracy (DTA). Utility parameters are commonly represented by the beta distribution; however this distribution can only be used if it is appropriate to assume utility is bound on the interval zero to one. Utility values are actually bound on the scale negative infinity to one (-∞, 1), and therefore in cases where very low or negative utilities are possible (such as in cancer or other severe and life threatening disease areas) the beta distribution is not appropriate. In such situations the distribution is better represented by a gamma or lognormal distribution, bound on disutility. I.e. where disutility (D) is equivalent to 1 minus the utility value (D = 1-utility), thereby bounding the distribution on the scale zero to infinity (0, ∞) (4).

In fitting a Beta distribution: Beta (α, β), the alpha parameter (α) is represented by the number of events of interest, while Beta (β) is equivalent

to the sample size minus the events. If the sample size and number of events of interest are reported in a publication (or other source of evidence), then these are used to represent the alpha and calculate the beta for the Beta distribution. Alternatively if the published evidence only reports the mean and standard error for a parameter point estimate then the Method of Moments can be used to derive the alpha and beta estimates from the mean and standard error and are then fitted to the Beta distribution (4).

The Dirichlet distribution is used to represent multinomial data where there are numerous categories, each represented by a proportion, of which the proportions sum to 1. If the overall sample size and the number of events of interest for each category are reported, then these can be used to calculate the proportion for each category and fit the Dirichlet distribution. The Dirichlet distribution enables the proportion within each category to vary, but in relation to the other categories so that they all sum to 1. Briggs and colleagues (4;75) provide further details on the Dirichlet distribution and a step by step approach for fitting the distribution.

In fitting a Gamma distribution: Gamma (α, β), a similar approach is adopted to that of the Beta, whereby the sample mean (parameter point estimate) and variance reported in published data can be used to determine the shape (α) and scale (β) parameters using the gamma Method of Moments (4).

Costs tend to be represented by gamma or lognormal distributions, however, where it is unlikely that costs will be highly skewed and the data used to derive cost estimates is symmetric around the mean (and sufficiently large), the central limit theorem (CLT) can be assumed to apply and a normal distribution fitted, using the mean value and standard deviation for the cost estimate. For the normal and lognormal distributions the mean and standard deviation are adequate to fit the distribution.

In cases where only partial evidence has been reported, i.e. a mean value with no standard error, assumptions can be used to determine an appropriate standard error (which is large enough to reflect a wide range of uncertainty). Additionally if 95% confidence intervals have been reported rather than a

specification of the standard error, then the standard error can be calculated using the 95% confidence limits, however, it is important to note that these calculations will differ for each distribution, based on the standard distributional assumptions (72).

2.4.3 Illustrating parameter uncertainty

Once distributions have been applied to each of the appropriate parameters in the decision model, the probabilistic sensitivity analysis can be performed. Monte Carlo simulation is used with a minimum of 1000 iterations to sample random values from each parameter distribution simultaneously to provide different cost, effect and cost-effectiveness outcomes for each vector of input parameters. The average cost, effect and cost-effectiveness outcomes across all 1000 iterations represent the probabilistic outcomes. The incremental costs and effect outcomes for each of the 1000 iterations can be plotted on a cost-effectiveness plane to illustrate the uncertainty (5). 95% confidence intervals (uncertainty intervals) can be represented using the lower and upper percentiles from the simulation results using the percentile method, i.e. taking the 0.025 and 0.975 percentiles (4). Figure 2-1 illustrates a cost-effectiveness plane, which has been used to plot the range of incremental cost and incremental effect outcomes from a Monte Carlo simulation.

Intervention versus Control
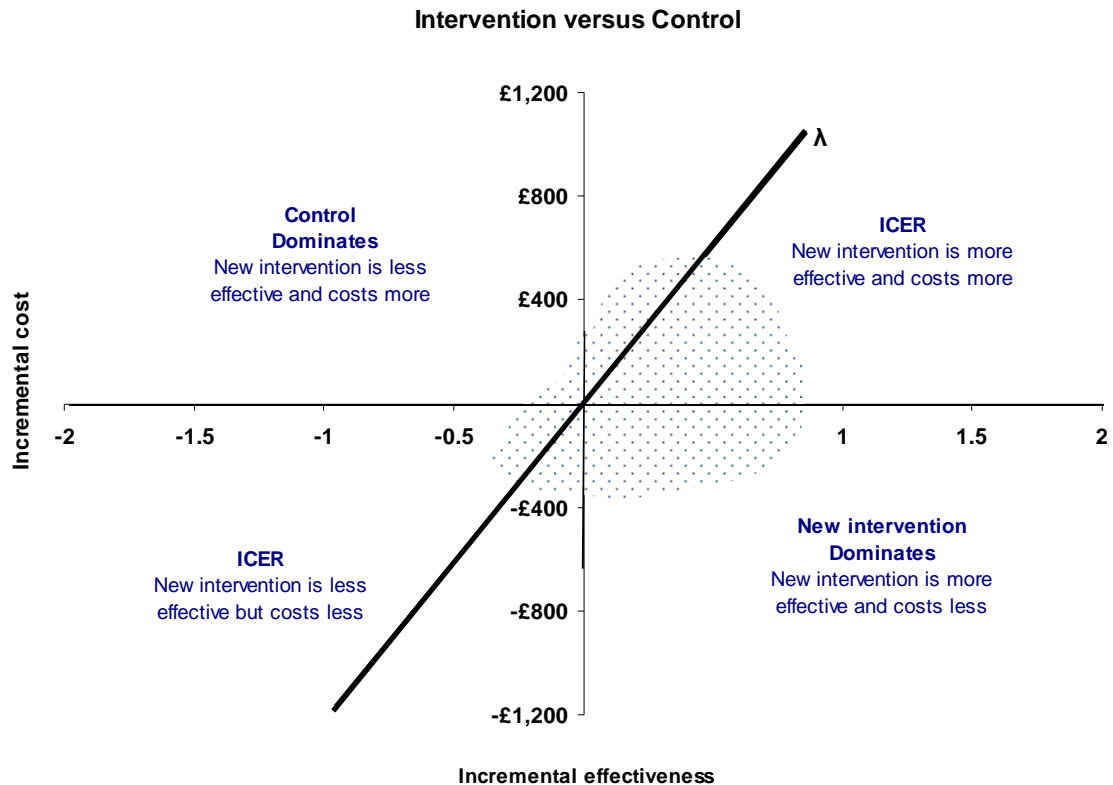
Figure 2-1: The cost-effectiveness plane

The cost-effectiveness plane illustrates the difference in effectiveness (Intervention minus Control) per patient on the horizontal axis against the difference in cost per patient (Intervention minus Control) on the vertical axis.

If the cost-effectiveness estimates fall into the North West (NW) quadrant, this demonstrates that the new intervention is less effective than the control, but costs more, and is therefore dominated by the control.   Alternatively if the outcomes fall into the South East (SE) quadrant, this represents improved effectiveness with the new intervention, yet it is also cost saving, and therefore the new intervention is the cost-effective strategy and it dominates the Control.

If the outcomes fall into either the North East (NE) or South West (SW) quadrants this represents a situation where a trade-off is required, i.e. improved effectiveness but at a greater cost in the NE quadrant, or a cost saving but at the expense of reduced effectiveness in the SW quadrant. In the

NE and SW quadrants a value judgement regarding societies willingness to pay (or willingness to accept savings) per unit of effectiveness gained (lost) is required in order to determine cost-effectiveness. The maximum acceptable cost-effectiveness ratio or ceiling ratio (λ) represents societies willingness to pay, and is illustrated in Figure 2-1 by the thick black line that crosses through the origin of the cost-effectiveness plane splitting the NE and SW quadrants. While the value of the ceiling ratio can be considered subjective or a value judgement, such as the UK recommended acceptable range of £20,000-£30,000 per QALY (24) which has no firm basis in evidence or theory (77); it can also be considered to be a reflection of opportunity cost. I.e. the opportunity cost is the technologies that are displaced at the margin, by those technologies which are recommended by NICE. In principle the ceiling ratio can be estimated using willingness to accept or willingness to pay, for example society's willingness to pay for improved health care through increased taxation (15). Indeed, numerous attempts have been made to measure consumers' and society's monetary valuation of improved health or valuation of a quality adjusted life year (QALY) (78;79).

The ceiling ratio is considered to be symmetrical for both the NE and SW quadrants, so that incremental values below the ceiling ratio (values which fall to the right of the ceiling ratio (λ) in Figure 2-1) will be considered cost-effective, and those that are greater than the ceiling ratio (values that fall to the left of the ceiling ratio (λ) in Figure 2-1) will not be considered cost-effective.   The UK decision making body NICE, recommends a monetary threshold of approximately £30,000 per QALY gained (24), and therefore interventions with an ICER of less than £30,000 are considered to be cost-effective in the UK.

The joint distribution of the costs and effects from the Monte Carlo simulation are plotted on the cost-effectiveness plane to demonstrate the impact of uncertainty in the model parameters on uncertainty in the model outcomes (expected incremental cost and effects). Uncertainty in incremental outcomes is demonstrated when the results spread across the y-axis. Likewise a spread through the origin passing through the horizontal axis represents uncertainty in the incremental cost of the intervention. The 95 percentile intervals for the

incremental costs and incremental effects from the probabilistic sensitivity analyses can be calculated; however it is not straightforward to calculate 95% intervals for the joint incremental cost-effectiveness ratio when the joint distribution crosses into more than one quadrant.

In situations where the range of outcomes spread from either the NE or SW quadrants across the origin into the NW or SE quadrants (dominant quadrants), some of the ICER values will be negative, and this will cause problems for establishing confidence intervals around the ICER (4;40). Therefore it is recommended to represent uncertainty in the cost-effectiveness estimate using a cost-effectiveness acceptability curve (CEAC) (4;7;71).

2.5 Decision uncertainty: CEAC

The cost-effectiveness plane was useful in terms of illustrating the impact of individual parameter uncertainty (within the model) on the model outcomes, i.e. uncertainty as to the existence and extent of any difference in costs and uncertainty as to the existence and extent of any difference in effect between the Intervention and Control. However, any uncertainty in these may or may not impact on decision uncertainty.   Decision uncertainty refers to uncertainty in whether the intervention can be considered cost-effective in comparison to the alternative(s), i.e. uncertainty in answer to the question: is the intervention cost-effective? When probabilistic outcomes spread into more than one quadrant of the cost-effectiveness plane, it can be difficult to determine whether the intervention is cost-effective or not and therefore CEACs can be used to illustrate the level of uncertainty in the estimated cost-effectiveness of the treatment in comparison to the alternative(s) (7;70;71).

Having undertaken an economic analysis and presented cost-effectiveness, it is important to address the initial research question, which was: 'is the new intervention cost-effective in comparison to the control'? To answer this question and give useful outcomes to decision-makers we need to consider two underlying questions: Given the current evidence and decision uncertainty,

should the technology be adopted (or not)? and, Is further research required to help support this decision? (4) If these underlying questions are not addressed, decision makers may struggle to interpret the cost-effectiveness results, particularly regarding how to make the decision to adopt or reject the intervention given the uncertainty. In answering these questions the researchers can help inform funders and decision makers on how to interpret the model results and the appropriate next steps to take.

A cost-effectiveness acceptability curve (CEAC) is an ideal tool for illustrating decision uncertainty (uncertainty around the question of whether or not the intervention is cost-effective) across a wide range of alternative ceiling ratios (7;11). As demonstrated in Figure 2-1 a monetary threshold of willingness to pay (the ceiling ratio) is required to determine cost-effectiveness and ICER values that fall below the ceiling ratio (values which fall to the right of the ceiling ratio (λ) in Figure 2-1) will be considered cost-effective, and those that are greater than the ceiling ratio (values that fall to the left of the ceiling ratio (λ) in Figure 2-1) will not be considered cost-effective. Therefore the ceiling ratio can be varied across a range of different values to illustrate how the evidence in favour of cost-effectiveness of the intervention varies at different monetary thresholds (4)

The effect, cost and joint cost-effectiveness results from the probabilistic analysis are used to derive a cost effectiveness acceptability curve (CEAC) which represents the probability that the intervention is cost-effective at different monetary thresholds of willingness to pay (7;11). The process of generating a CEAC involves calculating the proportion of iterations (from the PSA) where the intervention is cost-effective, given a specified ceiling ratio, and then re-calculating the proportion at a variety of different ceiling ratios. The CEAC plots the proportion of cost-effective iterations against the ceiling ratio. The proportion of cost-effective iterations can be calculated for the intervention and the alternative(s), so that they can both (all) be plotted graphically, as illustrated in Figure 2-2.

For example, given two interventions (Treatment and Control), the net benefit for treatment (NB_T) and control (NB_C) can be calculated as well as the

incremental net benefit $(\text{INMB} = \lambda * \Delta E - \Delta C)$; and the ceiling ratio (λ) can be varied across a range of values, i.e. from £0 to £100,000 in increments of £500. In this way the NMB from the net benefit calculation varies with the ceiling ratio (λ). For each of the alternative ceiling ratios the net benefit under the treatment (NB_T) and control (NB_C) interventions are calculated across all the 1000 Monte Carlo simulations and the average is taken (ENB_T) (ENB_C) to determine which has the highest expected net benefit. At each ceiling ratio, the intervention with the greatest expected net benefit $\left[\max(\text{ENB}_T : \text{ENB}_C)\right]$ is the most cost-effective and is therefore the optimal choice at that specific ceiling ratio. The proportion of cost-effective iterations for the Treatment and Control at each ceiling ratio can be plotted to illustrate the CEACs for each intervention, as illustrated in Figure 2-2.
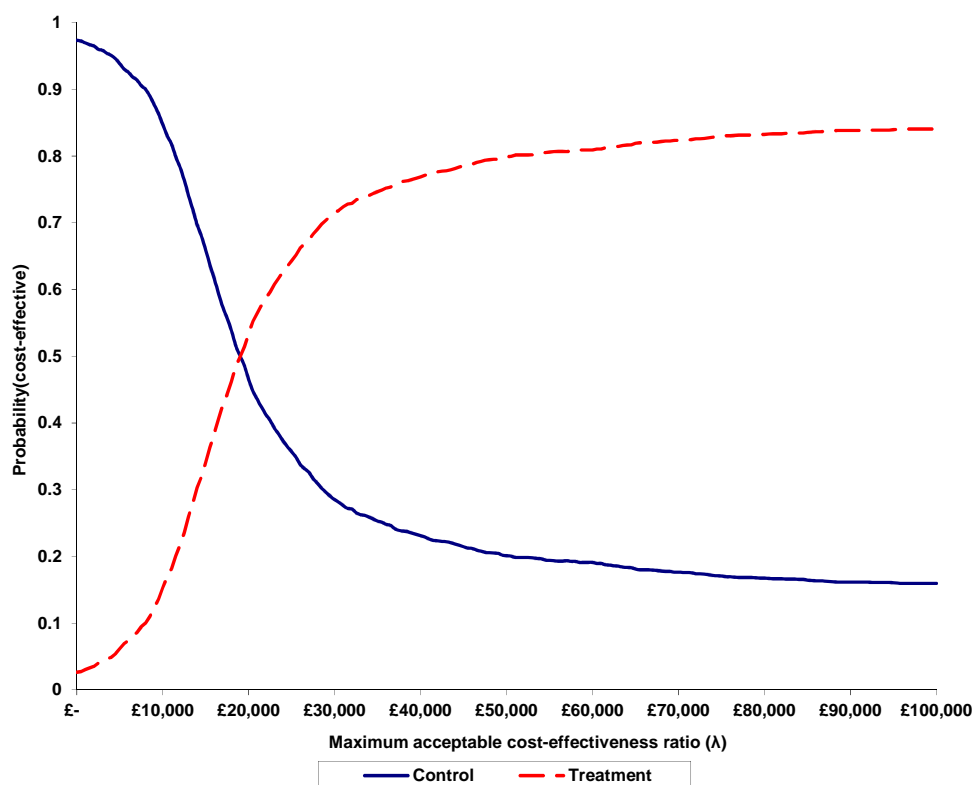
Figure 2-2: CEAC for Treatment versus Control

Figure 2-2 illustrates a hypothetical CEAC, showing the probability of each intervention being cost-effective at different ceiling ratios. At a ceiling ratio of £10,000 there is an 85% probability that the Control is cost-effective and

only a 15% probability that Treatment is cost-effective. The Control would be considered the optimal choice at this ceiling ratio (as it has the greatest expected net benefit and is therefore the most cost-effective at this ceiling ratio), and the CEAC characterises the uncertainty in this optimal choice. At a ceiling ratio of £50,000 Treatment is the cost-effective strategy, but the CEAC shows that this is only with an 80% probability; there remains a 20% probability that Treatment is the wrong choice, which is the uncertainty in the decision. The aim of calculating a CEAC is to summarise and illustrate graphically the probability that a particular intervention is the optimal choice over a wide range of values for the ceiling ratio (λ) (7). Fenwick and colleagues (11) demonstrate this process with a clinical example and present the CEAC.

The CEAC is straightforward to calculate, construct and interpret (11) and is therefore an ideal technique for presenting uncertainty in the cost-effectiveness outcome from a PSA to decision makers who have to make the choice of whether to adopt or reject a new intervention, based on the current evidence. Uncertainty over the results of an analysis leads to the possibility of incorrect decision making, which has a cost in terms of benefits from the correct decision forgone. Decision makers want to avoid incorrect decisions and therefore a CEAC can be a useful means of summarising the uncertainty surrounding the cost-effectiveness decision.

In order to adequately address the initial research question ('is the new intervention cost-effective in comparison to the control?') we need to consider two underlying questions: Given the current evidence and decision uncertainty, should the technology be adopted (or not)?; and secondly, is further research required to help support this decision? (4). Summarising decision uncertainty with a CEAC can be considered to be a first step in addressing the second important question; however, further techniques are required to answer the question in full (71).

2.6 Value of Information

Having developed a decision analytic model, undertaken probabilistic sensitivity analysis and considered decision uncertainty; the PSA results can also be used to undertake Value of Information analyses (VOI). VOI is based on the rationale that decisions based on existing information will be uncertain and given this uncertainty, there is a chance that the wrong decision will be made which will have a cost in terms of health implications for patients receiving suboptimal care and inefficient use of health care resources (10;12). VOI analyses value further research on its potential ability to reduce the expected costs of uncertainty surrounding the cost-effectiveness decision, rather than deciding on further research and its design through arbitrary means (9;12;13). Claxton and colleagues (22) promote VOI within a Bayesian framework for economic evaluation and decision making, in order to address uncertainty surrounding a cost-effectiveness decision and address whether the intervention should be adopted based on current evidence, or whether more evidence is required to support this decision.

VOI is based on the concept of the opportunity cost of making a wrong decision, given uncertainty. The Expected Value of Perfect Information (EVPI) is a methodological approach which uses the uncertainty surrounding the cost-effectiveness decision based on current evidence, the patient population, and technology lifespan and societies willingness to pay; to place a monetary value on the worth of further research. VOI techniques can be used in combination with evidence from a CEAC to make informed decisions as to whether a new intervention should be adopted (or rejected) based on current evidence, or whether further information is required to help make the decision, as opposed to making research decisions subjectively. The expected cost of uncertainty is determined jointly by the probability that a decision based on existing information will be wrong and the consequences of a wrong decision.

Decisions based on existing information will always have some uncertainty (as represented by the inverse of the CEAC) and therefore there will be a chance that the wrong decision will be made. In choosing whether or not to adopt an

intervention (Treatment) that had, for example, an 80% probability of being cost effective at society's willingness to pay threshold (λ), such as that illustrated on the CEAC in Figure 2-2, there is uncertainty. At a ceiling ratio of £50,000 the Treatment intervention would be considered the optimal choice (as it has the greatest expected net benefit $[ENB_T > ENB_C]$ and is therefore the most cost-effective at this ceiling ratio), however, it has only an 80% probability of being cost-effective in comparison to the alternative (Control). While Treatment is the optimal choice, 20% of the time it would have been the wrong decision, and this represents uncertainty in the decision to adopt the Treatment strategy. Uncertainty over the results of an analysis implies the possibility of incorrect decision making which imposes a cost in terms of the benefits forgone (9). Therefore, there may be value in obtaining more information, if it will reduce this uncertainty. Claxton and colleagues (4;9;21) provide a thorough description and worked examples of the VOI process.

2.6.1 EVPI per decision/patient

In a perfect world where there is perfect information, the optimal (most cost-effective) intervention would be chosen every time, however, in an imperfect world there is uncertainty and in choosing an intervention which on average is optimal, there is a probability that it will be the sub-optimal choice. The expected value of perfect information (EVPI) calculates the opportunity cost of making the wrong decision.

The EVPI process within the context of an iterative (Bayesian) framework for economic evaluation has been promoted throughout the last fifteen years (1;8;13;80), however it is not utilised in practice very often. The application of the value of information approach was demonstrated for the Health Technology Assessment programme in the UK (10) and for the UK reimbursement decision body, the National Institute for Health and Clinical Excellence (NICE) (12); and has gained recognition, however, practical applications of EVPI in published literature, while increasing in number in recent years, remain limited.

The process of calculating the EVPI follows on from calculating the CEAC using the results from a probabilistic sensitivity analysis. As discussed in section 2.5, in determining the CEAC from the probabilistic sensitivity analysis results, the net benefit can be calculated for each intervention (Treatment (NB_T) and Control (NB_C)) for all 1000 iterations of the Monte Carlo simulation, and the average of these is taken to calculate the expected net benefit for each, in order to determine the optimal strategy. Given current evidence the optimal intervention is the intervention which has the highest expected net benefit across all 1000 Monte Carlo iterations $[\max(ENB_T : ENB_C)]$. The ENB for the optimal strategy is the value of the decision given current information. Under perfect information, the optimal strategy would be chosen each time, and this can be accounted for by choosing the strategy which maximises net benefit for each iteration of the Monte Carlo simulation, from iteration 1 through to 1000: $[\max(NB_{T1} : NB_{C1})], [\max(NB_{T2} : NB_{C2})]..... [\max(NB_{T1000} : NB_{C1000})]$. The mean of the 1000 optimal NB choices is the expected value of the decision given perfect information $[E[\max(NB_T : NB_C)]]$. The expected value of perfect information is the difference between the value of the decision with and without perfect information, as detailed in Equation 2.2 (4;9)

$$EVPI = E[\max(NB_T : NB_C)] - \max(ENB_T : ENB_C) \qquad\qquad (2.2).$$

This gives the expected value of perfect information per decision, i.e. the EVPI surrounding the decision as a whole for each time this decision is made for an individual patient or patient episode. The EVPI is re-calculated for a wide range of alternative ceiling ratios (λ), and can be represented graphically to illustrate how the value of perfect information varies at different ceiling ratios.

2.6.2 EVPI population level

Having calculated the EVPI per decision/patient, it is important to account for what this represents in terms of the relevant patient population who would benefit from the additional information. The 'effective population' must be determined. This is the population of patients over the expected lifetime of

the technology who would benefit from further information about the technology (4). The EVPI per decision/patient can be multiplied by the 'effective population' to generate the population level EVPI.

As a first step, the annual population of patients should be determined. Published data on the annual incidence (I) for the specific disease of interest is an ideal way of estimating the annual patient population. The expected lifetime of the intervention or technology can then be estimated in years (t). While this is likely to be an assumption, it is possible to make a realistic assessment of the intervention or technology lifetime by considering the period over which the information about the decision will be useful (t) (4). For example, a technology that will be undergoing continual development and upgrading is likely to have a short lifetime as it will be replaced every few years by newer more advanced models. Therefore, an appropriate timeframe would be one where the estimates of effectiveness for the technology used in the model are relevant and unlikely to change. Given an annual disease incidence (I) and a technology or intervention lifetime of T years (t), the effective population can be calculated, applying a discount rate (r) for patients in future years, and summing the population across the years (t). The effective population is multiplied with the EVPI per decision to give the population level EVPI (EVPI$_{pop}$). Equation 2.3 details the population level EVPI calculation (4;9).

$$EVPI_{pop} = EVPI * \sum_{t=1,2,\ldots T} \frac{I_t}{(1+r)^t}$$ (2.3)

The population level EVPI can then be plotted on a graph, as with the EVPI per decision/patient, to represent how population EVPI varies with alternative ceiling ratios (λ). At societies' maximum ceiling ratio, the population level EVPI can be interpreted as the maximum amount the health care system should be willing to pay for additional evidence to inform this decision in the future. This can be considered to be an upper bound on the value of conducting further research (4;9), a necessary but not specific condition for determining the worth of future research.

The EVPI can be used to indicate whether further research is likely to be worthwhile; however, it does not take into account the cost or type of further research (81). This may be construed as a drawback to EVPI, however it is still useful in providing an indication of whether further research is worthwhile or not and is an important step in the stages of VOI analyses. For example, very low population level EVPI values can be used to suggest whether further research is worthwhile or not. A population EVPI of £8000 indicates that research is very unlikely to be worthwhile, and the optimal intervention should be adopted based on current evidence. In contrast a population EVPI of £30 million indicates that further research is very likely to be worthwhile. However, for a mid range population level EVPI of say, £800,000, it is more difficult to decide whether research would be worthwhile. Depending on the disease area and the type and scale of research, it may cost less or more than the £800,000 EVPI, which is interpreted as the maximum amount the health care system should be willing to pay for additional evidence to inform this decision in the future.

An intuitive explanation of EVPI is that it puts a monetary value on the worth of further research dependent on how much that research will reduce current uncertainty by. A systematic (or non systematic) literature review can tell you there is a lack of evidence from randomised controlled trials (RCTs), but it does not tell you the worth or value of undertaking a further RCT to address the lack of current evidence. The added value of EVPI (compared to not doing EVPI) is that it explicitly values potential research in terms of addressing current uncertainty, and can then be used to compare with the potential cost of further research. If EVPI is not undertaken then further research is decided upon arbitrarily.

Nosyk, et al. (82) provide a recent example of employing EVPI techniques in a study on dosing strategies for an influenza vaccine for individuals with human immunodeficiency virus (HIV). This publication highlights some of the advantages of undertaking an EVPI, as the authors found that the likely cost of further research would have been much greater than the expected value of any further research. The authors model the cost-effectiveness of three alternative dosing strategies for an influenza vaccine. Strategy C was found to

be the most effective, while strategy A was also potentially cost-effective in a patient sub-group. The authors report substantial uncertainty surrounding the potential cost-effectiveness of strategy A compared to strategy C, in one of the patient sub-groups. Rather than recommending another trial to address this outright, the authors undertook an EVPI to establish a maximum acceptable cost of further research to reduce uncertainty surrounding the implementation of strategy A versus C. The population EVPI was estimated to be $418,000 US dollars, considerably less than the cost of a large scale trial for the patient sub-group. The authors highlight that the cost of further research is likely to be much greater than the value of that research, and therefore recommend adopting strategy C based on current evidence.

This case study of a practical application of EVPI highlights the merit and importance of undertaking EVPI analyses. Even though EVPI does not explicitly account for the cost of further research, and can therefore only indicate whether further research is potentially worthwhile; when the cost of a trial is likely to be substantially higher than the maximum willingness to pay for further research (EVPI), then the optimal intervention can be adopted based on current evidence. EVPI helps ensure research resources are not wasted on further trials that have little value. EVPI provides a justification for spending large amounts of (scarce) research money on one trial rather than another. Without EVPI, it is only assumed that further research will add value, it is not explicitly checked.

2.6.3 EVPI for parameters (EVPPI)

In circumstances where the EVPI analysis indicates further research is potentially worthwhile, the next step involves identifying what type of research. 'Further research' does not necessarily mean that a large scale, randomised controlled trial is required. To define the type, and possibly the scale, of further research (to reduce uncertainty in the cost-effectiveness decision), researchers need to consider what information it is that is required? The requirement for further information is driven by uncertainty; so in considering what further research is potentially of value, the parameters that are driving uncertainty in the cost-effectiveness model need to be identified.

It is these parameters that would add the most value through further information. Expected value of perfect parameter information (EVPPI) is used to identify parameters for which more precise estimates would potentially be most valuable (4;21;83).

The value of reducing uncertainty surrounding particular parameters in the model (EVPPI) can be determined using an approach similar to the EVPI. In fact the EVPPI is the difference in expected value of the decision with perfect information for some parameters and the expected value of the decision with current information about the parameters. Equation 2.4 details the EVPPI algorithm.

$$\text{EVPPI} = E_{\theta}\left[\max_{t}\left(E_{\theta}NB(t,\theta)\right)\right] - \max\left[E_{\theta}NB(t,\theta)\right] \qquad\qquad (2.4)$$

The various steps in the EVPPI process are now outlined:

1. The first step involves choosing a parameter of interest for which perfect information is required, drawing a random value from its probabilistic distribution and then holding this value constant, to represent 'perfect' information for the parameter of interest.

2. The Monte Carlo simulation is re-run holding the 'perfect' parameter constant but allowing the probabilistic draws from all other parameters.

3. The average NB under the treatment and control (average NB_T) (average NB_C) from the 1000 iteration Monte Carlo simulation is then recorded along with the intervention identity that gives the maximum expected net benefit $\left[\max(NB_T : NB_C)\right]$.

4. Following the Monte Carlo simulation a second random draw is undertaken for the 'perfect' parameter of interest, to represent a new 'perfect' value which can be held constant. Steps 1 through to 3 are repeated 1000 times, each time holding a different value for the perfect parameter estimate constant while the other parameters in the Monte Carlo simulation vary. For each Monte Carlo simulation the mean net benefit for Treatment and

Control is recorded along with the intervention identity that gives the maximum net benefit.

5. Once the process has been completed, the 1000 stored mean NBs for Treatment and Control and maximum intervention identities are used to calculate the expected net benefit (ENB) for each intervention $[ENB_T, ENB_C]$ and the expected maximum net benefit $[E[\max(NB_T : NB_C)]]$ across the 1000 Monte Carlo simulations outcomes.

6. The intervention with the greatest ENB $[\max(ENB_T : ENB_C)]$ is the expected value of a decision based on current information, i.e. the intervention which has the greatest ENB and would therefore be the optimal (cost-effective) choice.

7. The expected maximum net benefit $[E[\max(NB_T : NB_C)]]$ is the average of the 1000 maximum net benefit interventions from each of the Monte Carlo simulations. This is the expected net benefit with perfect parameter information.

8. The final stage in the EVPPI process is to subtract the ENB of the decision under current information from the ENB of the decision with perfect parameter information to give the expected value of perfect parameter information (4).

Equation 2.4 detailed the EVPPI algorithm with the parameters of the model represented by (θ) including the perfect parameter of interest and the other parameters which keep their distributions from the PSA. NB(t,θ) is the net benefit of an intervention (t) if the parameters of the model take the value θ (4).

The EVPPI can be run for specific parameters in the model, but also for groups of parameters, where a specified group are held constant rather than a single parameter. It is useful to begin EVPPI with groups of parameters, for example some parameters may have been estimated from the same data source and it

may not make sense to consider them individually, such as parameters from a survival analysis. If parameters are correlated it is also important to group them together, so as to preserve the correlation structure (4). If a parameter is included in a group, but its correlated parameters are not included in that group, the correlation with those parameters may be broken and the EVPPI could potentially be over or under estimated depending on the type and direction of the correlation relationship (4). Additionally, it is useful to consider small groups of parameters in order to scope out what types of parameters are driving uncertainty, rather than undertake the process for every individual parameter in the model, which would be exceptionally time consuming (4). Parameter groups for EVPPI should be grouped logically, i.e. running all the utility parameters together and all the mortality parameters together. Claxton and colleagues recommend that the groups should be chosen to match the type of research that would be conducted (4;9), e.g. clinical parameters. It must be noted that the EVPPI for individual parameters do not sum to the overall EVPI, and likewise the EVPPI for groups of parameters is not equal to the sum of the individual parameter EVPPIs, because when considered in isolation individual parameters do not impact on the difference in net benefits in the same way (4).

The EVPPI calculated is the value surrounding the decision as a whole for each time this decision is made for an individual patient or patient episode (4), and therefore (following the same approach used to determine EVPI) it is important to account for what this represents in terms of the relevant patient population who would benefit from the additional parameter information.   The 'effective population' can be determined in the same way as described for EVPI in Equation 2.3. If undertaken directly following the EVPI calculation it is likely that the same effective population is relevant for the EVPPI however, it may well be that the relevant patient population who would benefit from the additional parameter information is different from that used for the EVPI. For example if the parameter or group of parameters of interest are specific to one sub-group of patients. The relevant population of patients should be considered who would benefit from further information about the technology, over the expected lifetime of the technology (4).

The EVPPI between the different parameters or parameter groups is compared, and the research planned (and designed if worthwhile) based around the parameters of interest that will provide the most value in seeking further research, although it must be noted that this is still only potentially worthwhile. The EVPI and EVPPI place an upper bound on the potential value of further research, which provides a necessary but not sufficient condition for acquiring further information (4;9). To establish a sufficient condition, i.e. to determine if research will be worthwhile and identify an efficient research design, the marginal benefit and marginal cost of further research needs to be accounted for (9).

2.7 EVSI

Having determined that further research is potentially required, and having used EVPPI to specify the most likely type of information and type of study (i.e. an RCT to gain clinical information on diagnostic test accuracy, or a survey used to determine patient quality of life) that is potentially required, VOI techniques can also be used to help design a trial with regards to identifying an optimal design and sample size. The expected value of sampling (EVSI) (9) is the technique used to quantify the expected value to the decision maker of obtaining sample information before making a decision (84). The EVSI is the difference between the expected value of a decision after the proposed research (with sample information) and the expected value of the decision given current information (83). Calculating the EVSI follows a similar approach to that used in calculating the EVPPI.

Ades et al. (83) describe the framework for the expected value of sample information (EVSI) approach, detailing the various stages involved in a step by step outline, with the accompanying algorithms. These are briefly summarised here. The EVSI process is based around specific parameters of interest, which are the uncertain parameters (θ_i) that require further information from the proposed piece of research.

1. For an assumed new sample size (N), the initial step of the EVSI involves drawing a sample of the parameters of interest (θ_i) from the prior distribution, i.e. from the probabilistic Monte Carlo simulation undertaken for the pre-trial model; draw the parameters of interest from the first iteration. These mean estimates from the first iteration represent the prior estimates for the parameters of interest.

2. Secondly, draw a random sample (for example from a Beta binomial distribution) to simulate the true events (x) given the sample size (N) and the probability estimate of the parameter of interest (θ_i). I.e. for the specified sample size and the mean estimates from the first Monte Carlo iteration; generate one draw from the binomial likelihood to represent the number of events (x). Using this draw and the prior mean for the parameter of interest, calculate a posterior estimate (θ_{ip}).

3. The third step is to put the posterior probabilities for the parameters of interest (θ_{ip}) back into the model (replacing the prior estimates for the parameters of interest) and recalculate the net benefit (NB) for each intervention (t), NB(t,θ_{ip}) using the posterior probabilities for the parameters of interest and the prior probabilities of the other parameters in the model. If the model is linear, then re-calculating the NB for each intervention requires only the posterior probability point estimate for the parameter(s) of interest, while a non-linear model requires the use of the posterior distribution to re-run the Monte Carlo simulation, making the process considerably more time consuming for non-linear models.

4. The net benefit for each intervention is calculated and stored, identifying the intervention (t) which maximises the net benefit.

5. This process is repeated again (steps 1-4), using the second iteration of prior means from the Monte Carlo simulation, and continually repeated for all of the prior estimates, storing the net benefit for each intervention and the intervention identity that gives the maximum expected net benefit for each.

6. Once the process has been repeated for all of the prior iterations in the Monte Carlo simulation the stored NBs and maximum intervention identities are used to calculate the expected net benefit (ENB) for each intervention.

7. The intervention with the greatest ENB is the expected value of a decision based on current information, i.e. the intervention which has the greatest ENB and would therefore be the optimal choice.

8. The expected maximum net benefit is calculated (i.e. the average of the maximum NBs from step 4); this is the expected value of a decision based on sample information for the selected sample size N.

9. The ENB of a decision under current information is subtracted from the ENB of a decision based on sample information to give the expected value of sample information.

Equation 2.5 below details this EVSI algorithm with the parameters of the model represented by (θ) including parameters of interest for which we are uncertain and complimentary parameters for which we do not seek further information. NB(t,θ) is the net benefit of an intervention (t) if the parameters of the model take the value θ:

$$EVSI = E_{\theta}\left[\max_{t}\left(E_{\theta}NB(t,\theta)\right)\right] - \max\left[E_{\theta}NB(t,\theta)\right] \tag{2.5}$$

This EVSI calculation can be repeated for different sample sizes, so as to generate different EVSI values for a range of sample sizes, i.e. N = 10, 50, 100, 250, 500, 1000, 1500, 2000. The EVSI can be plotted on a graph to demonstrate how it varies with increasing sample size. As sample size becomes very large, EVSI should tend towards the EVPI and level off, but may not necessarily reach the same value as EVPI (9;85).

The EVSI calculated is the value per decision, and therefore (following the same approach used to determine EVPI and EVPPI) it is important to account for what this represents in terms of the relevant patient population who would benefit from the additional parameter information. The 'effective

population' can be determined in the same way as described in Equation 2.3, to estimate the population of patients over the expected lifetime of the technology who would benefit from further information (4;9). However, when calculating the population EVSI, Claxton cautions that consideration should to be given to patients in the effective population who would be in the trial, and therefore have already benefitted from the treatment, so can potentially be subtracted from the effective future population (9).

2.8 ENBS

The greater the sample size, the less uncertainty around the variables of interest, however, as sample size increases, so does the cost of the study. Therefore the optimal sample size of a study can only be determined once the cost of the research is accounted for. The difference between the expected benefits of sample information (EVSI) and the cost of acquiring the sample information is the expected net benefit of sampling (ENBS) (4;9). The ENBS is concerned with finding an appropriate balance between reducing uncertainty and an escalating sample requirement. The optimal sample size for the proposed trial is the value of n (sample size) that generates the maximum ENBS.

The traditional approach to designing and selecting an optimal trial sample size (72) gives no consideration to the marginal costs and benefits of sample information (13), however, ENBS does. Claxton et al. (86) explain that calculating the expected value of sample information and then considering the ENBS produces a technically efficient research design, by considering the willingness to pay for the proposed research. The ENBS is the difference between the EVSI and the cost of acquiring the sample information (9); with the optimal sample being the value of N that generates the maximum expected net benefit of sampling. The ENBS of any sample size (n) can be calculated given the population EVSI for that sample size and the cost of research at that sample size (C_n), as described in Equation 2.6:

$$ENBS_n = popEVSI_n - C_n \tag{2.6}.$$

Having undertaken the population level EVSI calculation, the effective population will be defined (but may be modified as the sample size increases to exclude members of the effective population who take part in the trial). The research costs will entail a variety of costs such as research staff time, administration, etc. that can be classified as fixed, however there will inevitably be some variable costs such as treatment costs, medical materials, and potentially a small administrative cost per additional person recruited into the sample. Therefore, the cost of the study is likely to increase marginally as the sample size rises. The ENBS should be calculated across a wide range of sample sizes, and where ENBS reaches its maximum, that is the optimal sample size. The cost of the research, ENBS and EVSI can be plotted on a graph for a wide range of sample sizes to demonstrate how these three components interact and illustrate the optimal sample size.

Essentially, the EVSI approach is used to measure the value of a range of sample sizes for a proposed study, and then compare those values against the likely cost of the study at the different sample sizes, to help identify an optimal (efficient) sample size. EVSI and ENBS techniques are used to establish the efficiency of the proposed research design, ensuring a cost-effective use of research funding. Typically economic evaluation is concerned with the cost-effectiveness of health technologies to ensure efficient use of health care budgets, but in using VOI techniques, EVSI and ENBS in particular, we are concerned with the efficient use of research funds for funding potential studies and trials.

An example of a recent practical application of EVSI is provided by Stevenson et al. (87) in the disease area of osteoporosis. The authors employ VOI and EVSI techniques to establish whether an RCT of two alternative treatments for bone fracture prevention in postmenopausal women with previous fractures would be cost-effective. The authors' model the cost-effectiveness of Vitamin K compared to Alendronate based on existing information, and found that while Vitamin K dominated Alendronate, this cost-effectiveness outcome was subject to considerable uncertainty. Using VOI techniques, the authors established that

further research was potentially worthwhile to reduce uncertainty in the cost-effectiveness decision. The authors then undertook an EVSI calculation to establish whether a head to head RCT would be an efficient (cost-effective) use of research funding. After calculating the EVSI across a range of sample sizes, the cost of the research design is incorporated to calculate the ENBS across the range of sample sizes. The authors conclude that a sample size between 2000 and 5000 per arm would be sufficient to answer the cost-effectiveness question, and is efficient in terms of a valuable use of research money. The authors highlight that their case study allows the evidence base for decision making to be strengthened while prohibiting research that is likely to provide no additional value.

2.9 An Iterative Approach to Economic Evaluation

Having provided an overview of the methodologies used in decision analytic modelling for economic evaluations, this chapter now reconsiders the research context within which economic evaluations are undertaken and supports the iterative approach to economic appraisal as an appropriate framework for undertaking research in the health care sector.

Health care decision making bodies and reimbursement agencies in different countries have varying requirements and different degrees of stringency for determining whether health technologies represent good value for money (16;17). In the UK there is a formal requirement for economic evaluation to be an integral component of NHS research and development funded projects (52), however, health care research from other publicly funded research bodies and within the commercial sector is not subject to such requirements. Therefore research undertaken in the health care sector may or may not include economic components, in the various phases of the research process. Many economic evaluations that are commissioned by healthcare funding bodies or pharmaceutical companies are often not considered until the latter phases of clinical trials (20;28). Even in publicly funded research, explorative and Phase II research budgets rarely leave adequate financing to incorporate economic

components. Therefore, economic evaluations are commonly only incorporated in the latter phases of the research process, using information from larger scale definitive trials in order to support a case for reimbursement (1). Theoretically this one-off approach to economic evaluation is inefficient, as it only enables identification of new health technologies/services as unlikely to be cost-effective at a late phase of the research process. Sculpher et al. (1) caution that such a viewpoint also potentially limits the quality of an economic evaluation, leading to incomplete economic analyses base primarily on data from one large scale RCT. A robust economic analysis requires modelling and data synthesis of all available data, for example with prospective trial information supporting a wider secondary evidence base (1;39).

An iterative approach to economic appraisal is a framework that has been proposed as best practice (1;3) for evaluating health care technologies. The framework proposes that the process of health care evaluation should begin with an explorative modelling approach using indicative studies to assess cost-effectiveness based on existing information, and progress to more rigorous assessments, updating a decision model over time as more data becomes available (1;3). The iterative approach to evaluation provides a structure in which evidence from a range of sources can be synthesised and continually updated in order to answer cost-effectiveness decision problems for a defined context and population. The premise is that rather than using economic evaluation as a restrictive, one-off analysis it should be an iterative process conducted alongside all stages of healthcare research.

The iterative framework has five main stages and utilises the key methodologies for decision analytic modelling which were outlined in sections 2.3 to 2.8 in this chapter. The iterative economic framework utilises decision analytic modelling as a key tool for evidence synthesis along with Bayesian updating in order to answer cost-effectiveness decision problems, as illustrated in Figure 2.3.

The iterative approach to economic appraisal

STAGE 1	STAGE 2	STAGE 3	STAGE 4	STAGE 5
Identify decision problems Explore and form the research question	*Synthesis and modelling given available evidence* Systematic reviews Meta-analyses Expert opinion Develop an early probabilistic decision model based on the available evidence	*Setting research priorities* Use early decision model Sensitivity analysis Value of Information analysis Is further research required? Focus of research Optimal research design	*Primary Research* Undertake research e.g. RCT using appropriate design and power	*Synthesis & modelling* Primary trial outcomes Newly available evidence Bayesian updating of the probabilistic decision model

Identify decision problem	Early probabilistic model incorporating current evidence and opinion	VOI analysis	Primary research	Bayesian updating	Probabilistic model incorporating updated evidence

Figure 2-3: The iterative approach to economic appraisal
Adapted from Sculpher et al. (2006)

Figure 2.3 is based upon the five stage iterative approach initially illustrated by Sculpher et al. (1) and adapted to present the economic tools and decision modelling processes undertaken at each stage. The five stages of the iterative framework presented by Sculpher et al. are now discussed, providing some further detail regarding how the economic tools and methodologies for decision analytic modelling fit in with each stage.

2.9.1 Stages of the iterative approach

Stage 1: Identify decision problems

This first stage of the iterative framework is an explorative stage, focussing on identifying potentially important decision problems for different patient populations and sub-populations within various healthcare areas. Similar to the explorative nature of pre-clinical research and phase I in the drug development process, stage 1 of the iterative economic approach is used to explore the decision area and formulate an appropriate (economic) question.

This stage of the iterative process is also similar to the first two steps in building a decision analytic model: 'specifying the problem' whereby the objective of the evaluation is specified in order to 'define boundaries' whereby relevant aspects of the analysis are clearly defined such as comparators, outcome measures, disease area, patient population and perspective of the analysis. An economic analysis informing decisions needs to be clear about which patient populations are being considered, the treatments, therapies and technologies currently available to the specified patient population, and the role for any new treatment. The iterative process therefore begins by exploring the literature and existing information to identify the decision problem and specify the details.

This stage of economic appraisal is often, but not necessarily, pre-determined by public funding bodies and agencies who have identified the decision problem through various means, such as consultation with key NHS stakeholders, extracting research recommendations from various resources and communication with researchers (52). They may or may not have already defined the relevant interventions of interest and identified the relevant patient groups.

Stage 2: Synthesis and modelling given available evidence

Once a specific decision problem has been identified, stage 2 will explore any existing and available information, including expert opinion, and then undertake evidence synthesis and early economic modelling based on this evidence. Undertaking systematic reviews and meta-analyses on the effectiveness of the current standard of care and the new drug or health technology can generate evidence for use in a probabilistic decision model (5). Meta-analyses are a useful means of synthesising evidence on a primary effectiveness outcome, however, decision analytic modelling in itself is a mechanism for evidence synthesis and therefore a wide range of relevant information (such as costs, outcomes and longer term impacts) from a range of sources can be gathered at this stage and synthesised in the early decision model.

With regards to the decision modelling methodology, stage 2 of the iterative economic process uses many of the tools, such as synthesising information from a variety of sources to building a deterministic decision model, fitting distributions to the model parameters and undertaking probabilistic sensitivity analysis. In building an early DAM, expert opinion may be required where evidence is limited and this can be appropriately reflected in the use of wide confidence intervals around the base case estimate in the probabilistic sensitivity analysis.

Developing an early decision model can provide an indication of whether the new drug or technology is expected to be cost-effective (2) and the associated uncertainty.

Stage 3: Setting of research priorities

This stage involves a formal process of research prioritisation determined by the current evidence and early decision model outcomes from stage 2; i.e. whether we require further information to reduce the uncertainty surrounding cost-effectiveness, and if so, to identify the focus of further research and an appropriate research design (2).

Considering the decision modelling methodology outlined earlier in this chapter, value of information analysis (9;10;12) is an ideal tool for addressing such questions and setting research priorities. VOI is based on the rationale that decisions based on existing information will be uncertain and given this uncertainty, there is a chance that the wrong decision will be made which will have a cost in terms of health implications for patients receiving suboptimal care and inefficient use of health care resources (10;12). The expected cost of uncertainty is therefore based on the probability that a decision based on current information will be wrong, and the consequences of the wrong decision. This can be interpreted as the expected value of perfect information (EVPI), because perfect information can eliminate the possibility of making the wrong decision (10;12). If the EVPI exceeds the expected costs of additional research then it is potentially worthwhile undertaking further research to

gather more information. EVPI can only inform on whether further research is potentially worthwhile, because perfect information doesn't exist.

This approach can also be applied to specific parameters of interest, to help determine the focus of further research. Expected value of perfect information for parameters (EVPPI) is used to identify which parameters (or groups of parameters) in the model have the greatest uncertainty, and impact on decision uncertainty in the model outcome (2). If there is little impact on decision uncertainty from the specific model parameters, then it is unlikely to be worthwhile undertaking further research, however, some parameters uncertainty may translate into considerable decision uncertainty for the cost-effectiveness decision, indicating a potential need for further research in these specific parameters. Therefore undertaking early modelling (at stage 2 of the iterative process) which incorporates sensitivity analysis or formal value of information analysis, provides a means to identify such parameters through assessing the EVPPI (2), and enables appropriate priority setting at stage 3.

Once EVPI and EVPPI have been undertaken and an indication of the potential value of further research in a specific parameter(s) has been identified (e.g. potential further exploration of the main treatment effect of a new drug compared to the current standard drug) the optimal study design must be determined. For example, if the treatment parameter is likely to be subject to selection bias then an RCT design may be appropriate. Any trial or study deemed to be a research priority should be powered and designed appropriately. For example, if primary research is commissioned based on insufficient data surrounding the treatment effect parameter, then the study should be powered to detect a difference in effectiveness. However, if insufficient information surrounding cost-effectiveness was highlighted from stages 2 and 3, then the research should be powered to detect an economic difference between the alternatives under comparison, i.e. a difference in cost-effectiveness (44).

Having used EVPPI to indicate the type of information and type of study of potential value (i.e. an RCT to gain clinical information on diagnostic test accuracy, or a survey used to determine patient quality of life), VOI techniques

can again be used to aid optimal study design. The expected value of sample information (EVSI) (9;83) and the expected net benefit of sampling (ENBS) can be calculated to determine an appropriate sample size.

Stage 4: Primary research (e.g. RCTs)

Trials in this stage should have been designed and powered appropriately to detect a difference in the key parameter driving the primary research, based on the stage 2 early model outcomes and the research priorities established in stage 3, using VOI analyses.

If an economic evaluation is being undertaken alongside the primary research, attempts should be made to adhere to the gold standard characteristics for economic evaluations within clinical trials as suggested by Glick et al. (38;40). The gold standard economic evaluation within a clinical trial will use an appropriate comparator, such as current practice or a commonly used cost-effective alternative. The trial will be adequately powered to assess homogeneity of economic results across a range of clinical settings, will provide detailed cost information and will allow adequate follow-up time to assess the full impact of the therapy. This will strengthen the design of the research and improve the quality of the economic evaluation.

Stage 5: Synthesis and modelling with updated evidence

In this stage new evidence is incorporated into the information set used within the model, i.e. the early decision model from stage 2 can be updated with the data generated in the primary trials undertaken in stage 4, and any other evidence published during the interim. This is based on a Bayesian concept (8;88) whereby evidence or knowledge about each parameter in the model can be updated by new information as it becomes available.   Having synthesised the primary research outcomes with any other relevant data in stage 5, the iterative process then loops back to stage 2 again.   It is important to consider whether the decision problem specified in stage 1 was answered adequately, and if not, does it need re-defined given any changes to the technology or disease field?

2.9.2 Benefits of the Iterative Approach

The iterative economic framework is gradually becoming recognised by some funding and decision making bodies in the UK, such as the Medical Research Council in the case of complex interventions (89). However, funding for economic evaluation of new health care technologies by both the pharmaceutical industry and publicly funded research bodies remains predominantly a one-off exercise. Vallejo-Torres et al. (20) support the iterative economic approach and explain that it can offer advantages in three specific ways: (i) by allowing the estimation of potential cost-effectiveness to be part of the investment/research decision process (which will improve internal efficiency and potentially save money, avoiding investment in health care technologies or interventions which could never be cost-effective), (ii) by supporting health care research bodies to prioritise between several competing and possibly cost-effective alternatives, and (iii) by enabling identification of the parameters that have the largest impact on the likely cost-effectiveness of a specific health technology or intervention from an early stage, it promotes efficient allocation of scarce public and private resources.

The key advantages of employing an iterative economic approach to funding research in the health care sector are now discussed, in terms of reducing uncertainty, reduction of costs and efficiency gains.

Improve decision making / reduce uncertainty

The research process is based on a step wise approach, each stage feeding information into the next in order to reduce uncertainty and aid decision making throughout. For example, in order to introduce one new drug a pharmaceutical company will begin with thousands of new chemical entities and will eliminate many through pre-clinical research. In the drug development process four phases are used to improve the decision making process further regarding the most effective, safest, highest quality and most marketable drug for the patient population. Likewise research funders such as the HTA identify evidence gaps or decision problems for which they commission research through various means, (52) in order to identify the most important

research questions to fund. Economic appraisal also requires a step wise approach. There should be various stages to improve information regarding the important parameters and comparators in the model and the uncertainty surrounding the cost-effectiveness of the drug or health technology. Given the iterative approach for economic appraisal, the use of a single one-off trial is inadequate as a sole input for economic appraisal.

An early stage probabilistic decision model can help set research priorities and inform whether further primary research is required. If undertaken, further research such as an RCT will provide information on effectiveness, cost and other important parameters which can then be used directly to update the original model. Following, this a second iteration of stages 3 to 5 of the iterative approach would follow, whereby the probabilistic economic model is updated and VOI is undertaken to explore whether any further information is needed to reduce uncertainty. If there is a potential need for further research a decision model can again help inform the design of the further studies, for example; if quality of life, cost or subgroup populations information is required then the analysis may highlight that an RCT is not an optimal study design at this stage, given strong evidence from an earlier RCT. It may be more appropriate to undertake an observational study or evidence on other parameters, negating the demand for another large scale, expensive, RCT. Thus, adopting an iterative economic approach in the health care research process can improve decision making.

Reduction of costs

The drug development process involves a four phased step wise approach to reduce uncertainty and improve decision making regarding which drugs to proceed with, however, it is also time consuming, costly and of high risk to pharmaceutical companies (58). In pre-clinical research only a few potentially suitable drugs will be identified from amongst thousands, incurring huge sunk cost for the investment in unsuccessful medicines. The drugs that progress to clinical assessment undergo rigorous testing throughout the four phases, and as the certainty of success in one particular drug increases, the number of participants required for trials increases, as does the trial length and the

coinciding expenditure. It is the rigorous nature of this process which requires major time and cost investment in the development of a new drug. DiMasi et al. (90) estimated the total cost of research and development for a new drug to be approximately $802million, for the price year 2000.

Considering this huge research expenditure incurred mainly by the pharmaceutical industry it may appear reasonable to limit economic evaluation to a one-off trial based exercise, as routinely adding in stages 2, 3 and 5 of the iterative approach into this research process will have further cost implications (1); however, earlier modelling at stage 2 (an early DAM) could reduce the subsequent evaluation and research costs further down the process. For example, by eliminating the need for a large scale RCT further down the line or establishing cost-effectiveness earlier. Sculpher et al. (1) note that any cost implications of incorporating the iterative economic framework are likely to be a small proportion of the cost of any subsequent primary RCT, and a sensible investment to ensure appropriate primary studies are funded. Likewise in publicly funded health care research, incorporating an economic evaluation and modelling exercise (in advance of) and alongside an explorative Phase II research study could reduce (or possibly eliminate if found to be not worthwhile) the cost of any subsequent larger scale Phase III trials.

The main purpose of the iterative economic approach is to identify priorities for research at the later, more expensive stages of evaluation. The framework therefore helps avoid undertaking costly time consuming studies; i.e. avoiding trials for an intervention or technology that has little chance of being cost-effective; or indeed avoiding a study which is not necessary given current evidence and therefore is unlikely to be a cost-effective use of resources. Investment in early stage modelling and value of information analysis can lower costs later on for public and private health care research bodies and any further economic evaluations.

Efficiency gain with an iterative approach

Earlier funding for economic analysis improves the economic evaluation process, benefitting the NHS and reimbursement decision making bodies

through early assessment of potential cost-effectiveness and helping to identify and set research priorities. This process can help public initiatives to speed up approval decisions on important healthcare technologies or sought after drugs and make faster termination decisions for cost-ineffective drugs. Likewise, in the case of publicly funded health care research, an iterative approach will encourage considerations of cost-effectiveness earlier in the research process and interventions unlikely to be cost-effective can be set aside for those with more promise. Involving decision modelling in early stage explorative research will encourage efficient allocation of publicly funded research finances.

In the pharmaceutical industry decisions on the success of a drug are made at each phase based on a mix of objective and subjective information regarding safety, effectiveness, appropriateness for patient population, cost and potential future benefits of the drug (20). The focus of these decisions is on potential for success in the regulatory submission, however, decisions to terminate development on investigational drugs primarily for economic reasons (with the reimbursement submission in mind) has historically occurred very late in the clinical testing process (3;20). Through using the iterative framework for economic evaluation these decisions can be addressed much earlier and the health care sector can benefit by identifying which interventions to progress through to further research, based on potential cost-effectiveness as well as effectiveness. The impact on research and development costs can be substantial while the impact on efficiency could result in more innovation and new therapies reaching patients sooner (91).

Hill & Freemantle (92) note that some health technology companies are beginning to use economic evaluation at earlier stages to help decision making regarding clinical trial parameters, and informing termination decisions on products in development. Value of information analysis is an ideal tool for these kinds of decisions. As illustrated in Figure 2.3, VOI can be used in stage 3 of the iterative economic approach as a tool for setting research priorities and estimating the value of additional research to society; however, publicly funded research bodies and the pharmaceutical industry could adopt this approach within their internal decision making process for whether or not they should fund additional research and also for prioritising health care

technologies with the greatest potential for cost-effectiveness. This could reduce the cost of investment and reduce uncertainty regarding the final reimbursement decision. In the case of private health care research organisations, VOI analysis can be adapted to a companies' perspective to assess the value of undertaking further research in a commercial context (20), and the value of further information to society could be considered in the spending of public research monies.

Currently, research funding bodies may implicitly determine the value of proposed research projects, based on a variety of internal decision making criteria and subjective expert opinions; however, the value is not explicitly determined. If VOI techniques were applied to all project proposals submitted to funding bodies, then alternative research projects could be compared incrementally in terms of their value or net benefit to society. Such an approach would encourage efficient use of research money.

In the commercial context, Hill & Freemantle (92) suggest interim reimbursement decision making for pharmaceuticals in order to improve efficiency. The authors propose this would shorten time to purchase through a two-stage process of economic modelling and reimbursement decision making prior to a randomised controlled trial, and post-trial to confirm or revoke the interim decision. This approach could benefit the pharmaceutical industry, but may have detrimental effects for poor decision making in the interim based on uncertain cost-effectiveness estimates. However, as Vallejo-Torres (20) suggest, an iterative approach to economic evaluation throughout the lifecycle of a drug or health technology accounts for all new evidence as it becomes available. This is the same regardless of whether the research is regarding drug development or wider sector health care technologies and services.

Incorporating the iterative economic approach in both a commercial and public funding context can further improve efficiency and decision making. With regards to the current state of play in the publicly funded research arena, on paper funders do 'encourage' incorporation of an economic component in early stage and explorative research; however explorative research budgets (particularly for Phase I and phase II studies) rarely include adequate financing

to incorporate economic analyses, which as a result inevitably end up excluded at this stage.   With regards to the current state of play in a commercial context there is little evidence of incorporation of the iterative approach in the drug development industry, however some authors note pharmaceuticals are beginning to consider economic considerations at earlier stages, such as including the selection of compounds to develop, the choice of clinical trial parameters, and 'go/no-go' decisions on products in development (92). To encourage support of an iterative approach in the commercial context, the UK reimbursement agency NICE set up a scientific advice consultancy service in 2008, specifically to provide advice to pharmaceutical companies who have products in development that may be referred for technical appraisal (93). This promotes efficiency and saves time for both pharmaceutical companies and decision making bodies such as NICE. For example, through seeking advice from NICE early in the drug development process, for example at phase II, pharmaceutical companies can ensure that any clinical and cost-effective studies undertaken at phase III onwards meet the requirements for the NICE technology appraisal.

By adopting an iterative approach to economic evaluation in health care research, not only will economic evaluation be more useful and of better quality, it will also be advantageous to the research funders whether they be private or public. These various advantages support the case for an iterative economic framework for undertaking research in the health care sector. Early decision analytic modelling, in which the DAM is built in advance of primary research, is a key process of the iterative approach. The framework allows a decision model to be developed on existing evidence and fully exploited in terms of addressing uncertainty given current information. VOI techniques can also be applied to help set research priorities and inform on the potential value and design of future studies. Applying early stage DAM and VOI techniques enables meaningful recommendations to decision makers, who can then make informed decisions as to whether a new intervention should be adopted (or rejected) based on current evidence, or whether further information is required to help make the decision, as opposed to making research funding decisions based on subjective reasoning.

Despite the advantages of employing early DAM, in practice, support and financial funding for early stage modelling and full exploitation of VOI techniques is rare in the health care sector. Given this disparity; the remainder of this thesis examines the role of early DAM for informing research priorities in the health care sector and considered potential reasons as to why it is not more widely implemented. The thesis uses practical applications of building early decision analytic models, in which the various DAM tools described in sections 2.3 to 2.8 are applied, to demonstrate the feasibility and importance of early DAM and the design of future studies in the health care sector.

2.10 Summary

This chapter introduced economic evaluation in the context of public and commercial health care research and discussed the methods for undertaking decision analytic modelling for economic evaluation. Following a review of the methodology, the iterative approach to economic appraisal was put forward as an appropriate framework for undertaking research in the health care sector and the benefits of such an approach were highlighted in terms of reduced uncertainty, reduction of costs and efficiency gains. Incorporating an iterative economic framework into both private and publicly funded research in the health care sector can improve efficiency, research design and decision making. Early decision analytic modelling, prior to primary research, is a key process within the iterative framework, and enables exploration of uncertainty and the use of VOI techniques to inform future research priorities. However, in practice early DAM has been less well supported and funding for early DAM is rare. This thesis now explores the feasibility, merit and drawbacks of undertaking early decision analytic modelling through practical applications and considers potential reasons as to why it is not more widely implemented.

3 Early DAM: a case of good practice

3.1 Introduction

This chapter demonstrates a case of good practice in utilising early decision analytic modelling to help inform the design of future research. The case study details a piece of research that was funded by the National Institute for Health Research Health Technology Assessment programme (50) to undertake a systematic literature review and economic evaluation of an emerging technology which showed potential benefit as an addition to current technology, rather than as a replacement. This research remit corresponds with stages 2 and 3 of the iterative economic approach, developing an early decision analytic model based on existing evidence to undertake an economic evaluation to determine whether the technology is cost-effective given current information and whether further information is needed to make a more informed decision.   In this case, the decision problem had been identified and pre-specified by the funding body, that is, stage 1 of the iterative approach. As discussed in Chapter 2 this is often, but not necessarily always, the case.

3.1.1 The decision problem

Worldwide, colorectal cancer (CRC) accounts for more than one million cancers per year or 9% of all new cancer cases.   In the UK, CRC is the third most common malignancy after lung and breast cancer, with 37,514 new cases registered in 2006: around two-thirds (23,384) in the colon and one-third (14,130) in the rectum (94). For patients with CRC there are a wide range of clinical scenarios and various treatment options with different timings, dependent on the stage of the cancer and also the extent of the cancer growth within each stage. Therefore, accurate staging once the cancer has been diagnosed is necessary to help identify the most appropriate patient treatment (95). The Tumour, Node, Metastases (TNM) staging classification is the internationally accepted cancer staging system. The TNM system classifies the extent of the tumour (T); the extent of spread (if any) to nearby lymph nodes (N); and whether or not the cancer has spread to other organs in the body, i.e.

the presence of metastases (M) (95;96). Numbers from 0 through to 4 appear after T, N and M to indicate the presence of and increasing severity of the cancer within each stage. The American Joint Committee on Cancer (AJCC) uses these TNM classifications to provide a uniform description of cancers within broader categories, known as the four AJCC stages: I II III and IV (95). Table 3-1 details the AJCC stages with the different TNM classifications for colorectal cancer, describing the extent of disease within each.

Table 3-1: Staging colorectal cancer using AJCC and TNM classifications

AJCC stage	Description of cancer	Tumour category	Node category	Metastases category
I	Invasion of inner layer of bowel & the bowel muscle	T1-T2	N0	M0
II	Invasion of outer lining of bowel wall & other parts of the bowel	T3-T4	N0	M0
III	Invasion of bowel & spread to lymph nodes	T1-T4	N1-N4	M0
IV	Invasion of bowel, spread to lymph nodes & distant metastases (spread to one or more other sites in the body)	T1-T4	N1-N4	M1a - M1b

Table developed based on information from Edge et al. (95) & CRUK (96)

Over the past two decades, a number of diagnostic tools have entered clinical practice and now facilitate the process of pre-operative staging for CRC. A number of imaging modalities are used including: computerised tomography (CT), contrast enhanced computerised tomography (ceCT), magnetic resonance imaging (MRI), ultrasound imaging, and positron emission tomography (PET). Recently there has been an emerging role for CT in combination with PET scanning ('hybrid' scan) in pre-operative staging for colorectal cancer.

PET and CT are complementary imaging techniques that, when combined, can maximise their individual advantages while minimising their respective disadvantages (97). Several studies have shown PET/CT to be more accurate than diagnostic CT and stand-alone PET for cancer staging including colorectal cancer (98;99), and the recommendation from the Royal College of Radiologists now is that every new PET scanner should be a PET/CT scanner and that every cancer network should have access to PET/CT services (97). However, PET/CT scanning is considerably more time consuming than either imaging device

alone, and for CRC it is recommended as an addition to routine imaging devices, and therefore as an 'add-on' technology it carries a heavy cost burden. Additionally, the diagnostic accuracy and therapeutic impacts of PET/CT in colorectal cancer are varied and currently unclear. Therefore, the HTA commissioned a systematic literature review of the evidence on effectiveness (diagnostic test accuracy) along with an economic modelling component to explore potential cost-effectiveness of the PET/CT hybrid scan as an addition to current practice. This is slightly different from a typical health sector evaluation whereby a new technology is compared directly against current practice, and cost-effectiveness is determined as a replacement technology. In this case, current practice with the addition of PET/CT was compared against current practice alone; therefore the cost-effectiveness of PET/CT was determined as an 'add-on' technology.

The pre-defined aim of the research was to determine whether PET/CT is cost-effective as an add-on test in comparison to routine imaging modalities; for pre-operative staging in patients with colorectal cancer. Given this aim of the research problem, identified by the funding body, probabilistic decision analytic modelling was undertaken using Monte Carlo simulation to address the following questions:

- Is PET/CT likely to be cost-effective as an addition to routine tests for pre-operative staging in CRC compared to current routine methods of diagnosis and staging, given current evidence and uncertainty?

- In which patients groups (i.e. primary, recurrent, metastatic disease)?

A value of information analysis was also undertaken to help inform whether there was potential worth in undertaking further research.

This chapter outlines the development of the economic component of the research, i.e. synthesising current evidence and clinical expertise to develop and populate the decision models. The next section provides an overview of the methods employed. Following this the design and development of each of the three models is detailed in turn. Subsequently, the evidence base and

data used to parameterise and populate the models is discussed for each model in turn and finally the deterministic results presented. The subsequent chapter (Chapter 4) explores the use of early decision analytic modelling to help inform the design of future research; presenting the probabilistic analysis results from the three models developed in this chapter and exploring uncertainty to assess whether there is any value in further research through value of information analyses.

3.2 Methods

3.2.1 Overview

The economic evaluation utilised existing evidence, incorporating decision modelling techniques to synthesise data from numerous sources (31;66). The evaluation was undertaken from the perspective of the UK NHS, reporting short term outcomes in terms of the incremental cost per correct diagnosis, and longer term outcomes in terms of the incremental cost per quality adjusted life year (QALY) gained. Reporting QALY outcomes enabled the analysis to incorporate the potential patient management implications of accurate and inaccurate diagnoses, particularly the implications for the patients' quality of life.

PET/CT and conventional imaging devices have different diagnostic test accuracies (DTA) for staging primary, recurrent and metastatic colorectal cancer, as such, in order to address the questions outlined in the aims, three separate economic models were designed. Patient management routes also differ between colon and rectal cancer, and therefore the primary and recurrent models were adapted to incorporate the specifics of rectal and colon cancer separately. The economic evaluation therefore involved five analyses, based on the three models, as detailed in Table 3-2.

Table 3-2: Model type and analysis undertaken

	Model	Analysis
1	Primary	rectal cancer
		colon cancer
2	Recurrent	rectal cancer
		colon cancer
3	Metastatic	metastatic disease

The cost-effectiveness of PET/CT as an add-on imaging device in pre-operative staging was assessed for: (i) primary rectal cancer, (ii) primary colon cancer, (iii) recurrent rectal cancer, (iv) recurrent colon cancer and (v) metastatic disease.

Each model was analysed probabilistically, using Monte Carlo simulation, to determine the expected cost, expected outcomes (correct diagnoses and QALYs) and the expected cost-effectiveness (cost per correct diagnosis and cost per QALY gained). The Monte Carlo simulations involved 2000 iterations for each model. Using any number of iterations greater than 1000 is generally considered acceptable (4), however to ensure a robust approach, the stability of the results was tested at 1000 iterations and 10,000 iterations for two of the analyses and found to be within reasonable bounds; i.e. there was only very slight changes to the incremental cost-effectiveness outcomes between 1000, 2000 and 10,000 iterations and therefore 2000 iterations was deemed acceptable.

3.2.2 Existing evidence

The economic models were designed, developed and populated based on a variety of information sources, in particular, from published data sources, literature and in consultation with clinical experts.

Previous economic evaluations of imaging devices for CRC were used to aid the design of the models, while the systematic review component of the research

project derived diagnostic test accuracy evidence for PET/CT and alternative imaging modalities. Economic and non-economic literature was required to inform specific model parameters, such as resource use, implications of diagnosis on patient management and therapeutic impact, quality of life and survival. Costing and resource use information was obtained from both the literature and UK NHS cost information sources such as the British National Formulary (100), Department of Health Reference Costs (101) and the Personal Social Services Research Unit (102).

Papers which were considered to be potentially relevant for the health economic evaluation were identified by the systematic reviewers during their screening process and passed on to the health economists as first-line literature to inform the development of the economic model. These initial papers provided an indication of the types of literature that were available and helped inform the design of the economic evaluations. Having established some first-line literature, a separate non-systematic literature search was undertaken in November 2009 to provide further information on the various parameters for the economic models. The objective was to search for and utilise information from economic evaluations and non-economic papers to develop and populate the economic models. Specifically, the search considered what evidence was available regarding the costs, treatment outcomes, management pathways, overall survival, quality of life and adverse events experienced by CRC patients undergoing pre-operative screening for primary, recurrent or metastatic colorectal cancer.

The following electronic databases were searched from beginning until November 2009: Medline, Embase, Web of Science, CINAHL Plus, Cochrane Library (NHSEED, HTA, CENTRAL, DARE), the Health Management Information Consortium (HMIC) and the CEA Registry. Specific searches were constructed for four main areas (PET/CT imaging for colorectal cancer, economics, adverse events or quality of life, and decision analysis) for each of the databases, as detailed in Appendix 1.   Inclusion criteria were applied to include relevant publications in any language that provided information on the diagnostic imaging devices PET/CT, contrast enhanced CT (ceCT) or MRI for detecting colorectal cancer, with regards to the topic areas of economic evaluation,

costing, patient management and therapeutic impact, quality of life and overall survival. Papers that only provided details on diagnostic test accuracy (DTA) were excluded, as this detail was being gathered by the systematic review team. Literature that was only detailed in conference proceedings or abstracts was also excluded. The search outputs are detailed in Table 3-3 below, resulting in a total 51 papers identified from the search which were deemed to be of relevance, plus an additional four quality of life papers identified through hand searching.

Table 3-3: Economics literature search

Search Stage	Search Strategy / Specified Criteria	No. of papers
Initial search	Identified 902 papers after de-duplication. Appendix 1 details the search for each database	902
Check titles & abstracts	Rejected 802 papers: All deemed irrelevant at title and abstract	100
Check full paper, apply inclusion criteria	Rejected 49 papers: Irrelevant / unavailable / abstract only / conference proceeding (21); Irrelevant / efficacy data alone (28)	51
Final papers	Costing studies (7) Economic evaluations (10) Quality of life (2) Survival (4) Management (28) Hand searched quality of life papers (4)	55

Information from this literature was used in consultation with the clinical experts involved with the project in order to design the models, in particular to identify appropriate comparators, management pathways and parameter estimates for each model.

3.3 Model design and development

Each of the three models was developed to incorporate the short term diagnostic pathway for the patient, the resultant treatment options and longer term outcomes such as patient survival and quality of life. Each model initially adopted a decision tree design to illustrate the patient pathway from suspected disease through to test outcome to distinguish between accurate and inaccurate disease staging. The costs and diagnostic test accuracy of the imaging devices were attributed to the appropriate branches in the trees. Following this, optimal treatment strategies with their associated costs and utilities were assigned dependent on the diagnosis from the decision tree, in order to account for treatment under both accurate and inaccurate staging. A longer term survival simulation was then undertaken to account for the costs, quality of life and survival impact of optimal versus received treatment. The diagnostic pathway for 'add-on' technologies will next be explored followed by details of the development of the Primary, Recurrent and Metastatic models respectively.

3.3.1 The diagnostic pathway

An economic evaluation for an intervention that is utilised as an add-on technology can be more complex than for a typical evaluation where cost-effectiveness is determined as a replacement technology. In this case, the evaluation involves the combination of two or more diagnostic tests and therefore presents some issues for mapping the diagnostic pathway and interpretation of test results, particularly if the second test refutes the outcomes from the first test. Therefore, prior to developing the decision models, the diagnostic pathway with a combination of two diagnostic tests needs to be explored.

In mapping the patient pathway for a diagnostic technology, decision trees are an ideal tool and typically follow the chronological order of the patient experience, for example, from a patient's initial examination, through to any tests, diagnosis and treatment, with the terminal node representing the true

patient disease status in order to distinguish between correct and incorrect diagnoses. However, in the case of diagnostic interventions, it may be more appropriate to structure the tree with disease prevalence at the beginning of the tree, prior to any diagnostic tests, in order to present a clearer, logical pathway. Phelps et al.(103) use such a strategy and portray a standard decision tree for a diagnostic test with disease prevalence at the first chance node, prior to the diagnostic test outcomes. Snider et al. (page 48) (61) also recommend this approach where appropriate.

> "Often the first chance node on a decision branch incorporates the characteristics of the population relevant to the prevention problem. The characteristics of the relevant population may be demographic, may be related to the prevalence of a disease or injury…"

Classifying the disease prevalence at the outset helps build a clear and logical structure in which the diagnostic test and its accuracy can then be identified, finishing up with correct or incorrect diagnosis at the terminal node. Fenwick (104) also prefers this approach for diagnostic interventions, noting that regardless of whether true disease status is mapped at the beginning or terminal nodes of the tree, the information and the probabilities associated with each pathway are the same. Therefore, in developing the decision tree a structure with disease prevalence at the outset was considered to be an appropriate approach.

Diagnostic testing is undertaken to aid in the diagnosis or detection of a disease, and the accuracy of the test depends on how well the technology or test correctly identifies or diagnoses the disease. The standard approach for measuring correct and incorrect diagnosis is the Diagnostic test accuracy (DTA), which is expressed through the test characteristics sensitivity and specificity. The sensitivity of a test is the probability that it can identify true positives, i.e. the proportion of patients who have the disease and are identified positively by the test (105).   The specificity of the test is the probability that the test correctly identifies true negatives, i.e. the proportion of patients without the disease who correctly receive a negative test result. These test characteristics are outlined in Table 3-4 for clarity.

Table 3-4: Diagnostic test accuracy

Test outcomes	Disease Prevalence	
	Disease +	Disease -
Test +	True Positive (sensitivity)	False Positive (1-specificity)
Test -	False Negative (1 – sensitivity)	True Negative (specificity)

As detailed in Table 3-4, incorrect diagnoses can fall into two categories: false positives or false negatives. When a patient has the disease but the diagnostic test (incorrectly) identifies no disease, this is known as a false negative, and the probability is equivalent to 1 minus the sensitivity of the test. When a patient does not have the disease, but a diagnostic device incorrectly identifies the patient as disease positive, this is known as a false positive, and the probability of occurring is equivalent to 1 minus the specificity.   As all diagnostic devices classify patients with some error, a clinician must select amongst various ways to interpret the diagnostic test results, trading-off the risks of false positives and false negatives (103). Therefore in an economic model it is important to analyse the longer term impact of diagnostic test accuracy, and specifically the impact on patient management in terms of false positives (where the patient may subsequently receive treatment for a disease they do not have) and false negatives (where patient may not receive treatment or experience a delay in treatment for their disease which was not identified).

If there is more than one diagnostic device used in the detection or diagnosis of a disease, such as in the case of add-on technologies, clinicians must deal with test error (false negatives and false positives) from both tests, and will require a strategy for dealing with conflicting test results (106;107). Figure 3-1 provides an example of a two test intervention, illustrating one branch from a decision tree for intervention (A) that involves two tests.

	Disease Prevalence	Test 1 DTA	Test 2 DTA	Alternative strategies	
				Confirmatory +ive	Confirmatory -ive
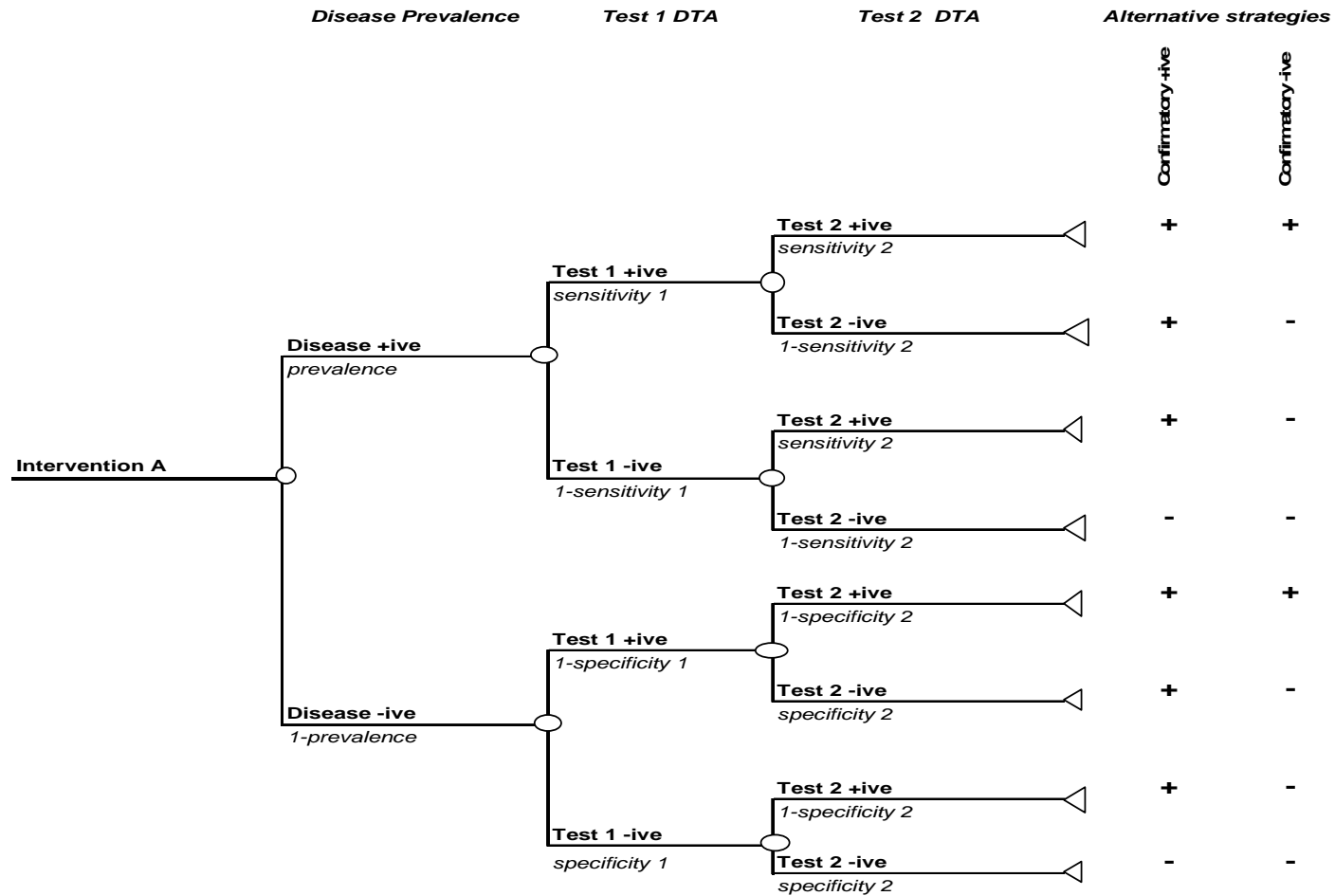
Figure 3-1: Diagnostic pathway for a two test intervention
This is an example with only 1 branch of a decision tree for an Intervention A, and does not include the comparator arm at the initial decision node

The branch for intervention A initially splits the patient population according to their true disease status, Disease positive (prevalence) or Disease negative (1 minus prevalence), and then incorporates the first diagnostic test, so that accurate (true positives and true negatives) and inaccurate (false positive and false negative) scan diagnoses can be identified from Test 1. The diagnostic test accuracy (sensitivity and specificity) for Test 1 is initially applied identifying true positives (sensitivity1) and true negatives (specificity1) for accurate diagnoses; and false negatives (1-sensitivity1) and false positives (1-specificity1) for inaccurate diagnoses. Following this, the second test is incorporated into the respective branches and the diagnostic test accuracy for Test 2 is applied (sensitivity2, 1-sensitivity2, specificity2, 1-specificity2). As can be seen in Figure 3-1, Test 2 either confirms or refutes the diagnoses from Test 1 depending on the diagnostic test accuracy of Test 2. In order to deal with the combined diagnostic test accuracy outcomes from the two tests, it is necessary to adopt a strategy to deal with conflicting results.

In the upper most layer of the branch in Figure 3-1, both Test 1 and Test 2 give positive disease confirmation, however, in the lower section of this upper branch, Test 2 is negative, refuting the positive Test 1 outcome. In this situation which test diagnosis should be accepted? If Test 2 outcomes always override Test 1 outcomes, then there is no need for Test 1 and the situation is no longer an add-on scenario but a replacement, and likewise if Test 1 outcomes are always chosen, there is no need for the add-on Test 2. Therefore, in a situation where the two tests are combined, there are two alternative strategies that can be adopted: confirmatory positive or confirmatory negative, as illustrated after the terminal node in Figure 3-1.

A confirmatory positive strategy would accept positive outcomes over negative results, i.e. where Test 2 is positive but Test 1 was negative the positive diagnosis from Test 2 would be accepted. Likewise where Test 1 is positive and Test 2 is negative, the positive result from Test 1 is accepted. This will have implications in terms of the proportion of false positives and false negatives identified from the tree. A confirmatory positive strategy benefits from the sensitivity of both tests, and may help identify a greater proportion of true positives than either test alone because false negatives from Test 1 would be

refuted by positive results from Test 2 in this strategy. However, this is at the expense of identifying fewer true negatives and also incurring a greater proportion of false positives i.e. trading-off specificity for improved sensitivity. The alternative strategies after the terminal node in Figure 3-1 shows that a confirmatory positive strategy results in mostly positive diagnoses from the tree, with a greater proportion of both true and false positives. In terms of patient management, the impact of additional false positives in this strategy must be considered, which is likely to lead to (over)treatment of patients for a disease that they don't actually have.

Alternatively, a confirmatory negative approach can be taken, whereby negative results overrule positive results, and in situations where the two tests contradict each other, the negative outcome is accepted. This is the second strategy illustrated at the terminal node in Figure 3-1, resulting in a greater proportion of negative outcomes. With this strategy more true negative cases will be identified, however, there will also be a greater proportion of false negatives identified. Therefore, this strategy benefits from the specificity of both tests, at the expense of lower overall sensitivity, and a greater proportion of false negative identifications. In terms of patient management, the impact of additional false negatives in this strategy must be considered, which is likely to lead to missed treatment (or at least delayed treatment) for patients with a disease, who were not identified correctly.

It is also important to note that if a confirmatory positive strategy is taken, then theoretically it would be more efficient to only incorporate Test 2 when Test 1 gives negative results. This strategy would give identical results as the confirmatory positive strategy but would only incur the cost of Test 2 in some rather than all cases. Likewise, if a confirmatory negative strategy is adopted, it would be more efficient and cost saving to only use Test 2 where Test 1 gives a positive result. While this is theoretically more efficient, in practice this approach may be less well accepted, and in the case of PET/CT, clinicians advised that all patients would be given a scan regardless of the conventional test outcomes.

In developing the model for the PET/CT as an add-on device in colorectal cancer staging, clinicians advised that in practice they adopt a confirmatory positive strategy for add-on PET/CT, i.e. positive results overrule negative results, and therefore results are only treated as negative when both the conventional and PET/CT test outcomes are negative. This strategy takes advantage of improved sensitivity from an add-on device, however this is at the risk of some over-staging (and subsequent unnecessary treatment) through false positives. By adopting a confirmatory positive strategy the clinicians indicate the importance of the benefit from improved sensitivity and are willing to accept the risk of incurring some additional false positives which may result in some patients receiving treatment unnecessarily, rather than adopt a strategy which would favour specificity and result in false negative identifications where patient would be untreated (or delayed treatment).

Therefore the baseline models adopted a confirmatory positive strategy, based on clinician advice, assuming that any positive test outcome overrides negative results. Negative results from the conventional imaging tests which are refuted by the PET/CT test are treated as positive, and likewise negative results from a PET/CT scan which conflict with prior positive results, are still treated as positive.   In situations where one of the tests is found to be superior to the other (in terms of both sensitivity and specificity) then the superior test can be taken to represent the combined DTA (108), therefore, in the PET/CT model a confirmatory positive strategy was adopted where there was no test superiority, and where one diagnostic test was found to be superior over the other, then this was taken to represent joint DTA in the add-on arm.

3.3.2 Model 1: Primary rectal & colon cancer

Diagnostic Pathway

The cost-effectiveness of PET/CT as an add-on device in primary (rectal and colon) cancer relates to the initial, pre-operative staging of primary patients. The value of incorporating a PET/CT scan in addition to conventional imaging in the primary disease stage is potentially greater sensitivity for the identification of nodal and metastatic disease (109-111); i.e. the N and M classifications in the TNM staging system, as detailed earlier in Table 3-2. The only diagnostic test accuracy evidence available for PET/CT in this context relates to the identification of lymph node involvement (109;110) and therefore the primary models were designed to evaluate PET/CT as an add-on device in nodal staging. In the AJCC staging system (detailed in Table 3-2), stages 1 and 2 have no nodal involvement, while both AJCC stages 3 and 4 can have some nodal involvement, and therefore the AJCC stages were considered to be an appropriate classification for distinguishing between nodal involvement in the primary model.

Figure 3-2 depicts the decision tree structure used for mapping the diagnostic pathway in the primary model. Due to the absence of economic models of PET/CT in primary CRC in the literature (none were identified from the systematic review or economics literature search), this model structure was informed primarily through consultation with clinical experts from the research team in order to accurately reflect the clinical pathway for nodal staging. The model was altered to include the disease specific criterion for rectal and colon cancers separately.
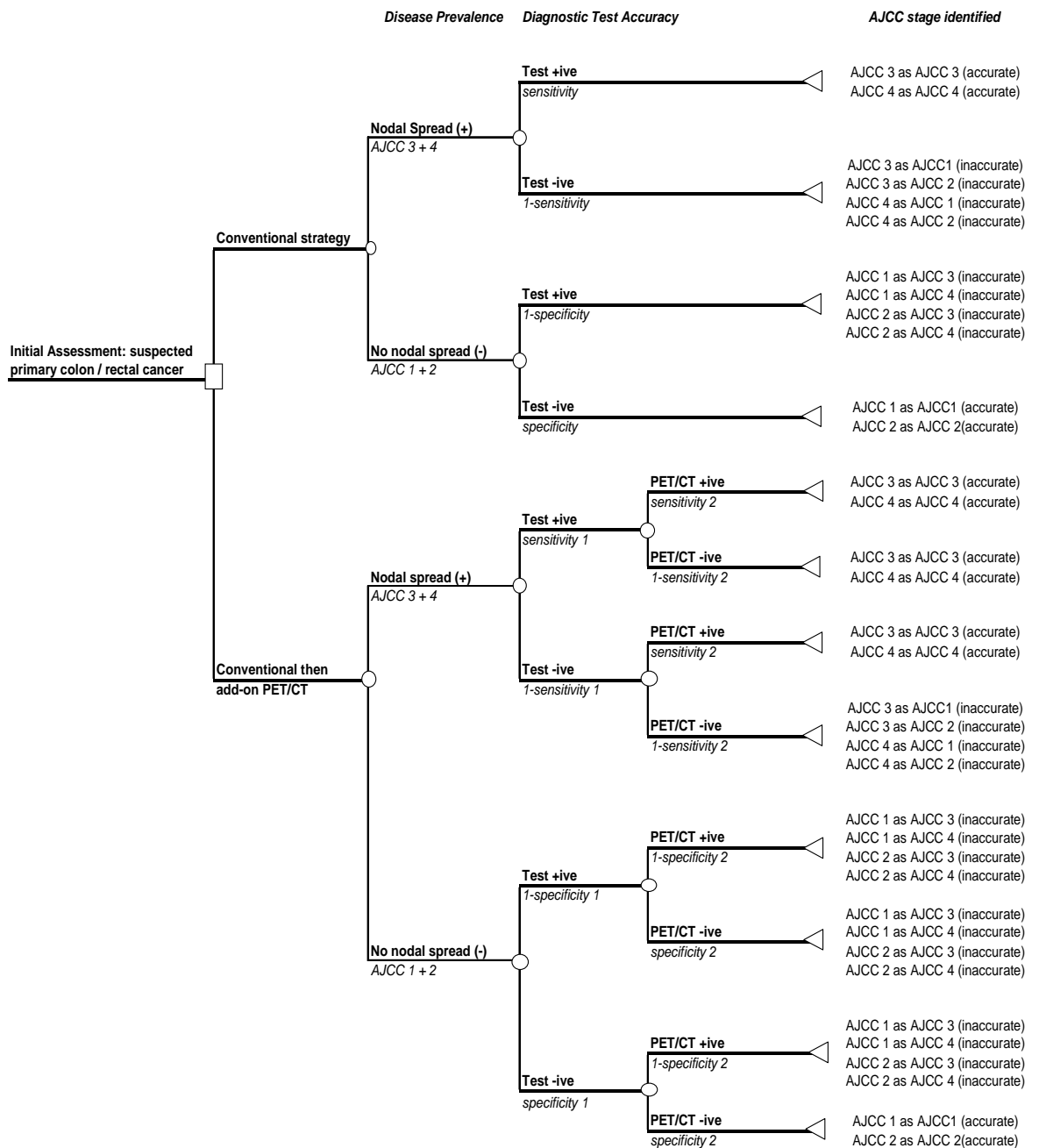
Figure 3-2: Decision tree for staging primary colorectal cancer

The decision tree model begins with patients who have had an initial assessment (involving a clinical examination, colonoscopy or sigmoidoscopy and a biopsy), which identified them as having primary (rectal or colon) cancer. The decision node in the tree represents the decision between the conventional strategy, where patients receive the standard imaging procedures, or the intervention strategy where PET/CT is added on. The

standard procedure for patients suspected of primary colon cancer is a contrast enhance computerised tomography (ceCT) scan of the chest, abdomen and pelvis to diagnose and/or stage the extent of the disease. In the case of primary rectal cancer, the conventional procedure involves a ceCT scan of the chest, abdomen and pelvis followed by an MRI scan of the pelvis[2]. This conventional pathway is represented in the top half of the tree in Figure 3-2. The intervention arm involves the conventional work up (ceCT for colon cancer or ceCT and an MRI for rectal cancer) followed by a PET/CT scan, which is depicted in the bottom half of the tree.

The primary decision tree model has been designed using actual CRC disease status at the outset, splitting the patient population according to the disease prevalence prior to the imaging scans, so that accurate and inaccurate scan diagnoses can be clearly identified. The objective of the scan in this model is to assess whether there is any nodal spread and therefore, after the initial decision node depicting the choice between the conventional or add-on PET/CT intervention, the tree divides the population according to actual nodal spread disease status using the AJCC colorectal cancer staging system (95) as detailed in Table 3-2. In the AJCC system, stages 1 and 2 have no nodal involvement, while both AJCC stages 3 and 4 can have some nodal involvement.

After dividing patients according to their true nodal spread disease status, the work-up of diagnostic tests are undertaken which will either identify nodal involvement (test positive), or no nodal involvement (test negative), depending on the sensitivity and specificity of the test. Having previously specified actual disease status, the top branch represents primary (rectal or colon) cancer with nodal spread (AJCC stages 3 and 4), and therefore at the test chance node the tree branch splits depending on whether the test was positive (accurately identified nodal involvement) or negative (inaccurately identifying no nodal involvement). The true positive outcomes correctly identify AJCC 3 as AJCC 3 and AJCC 4 as AJCC 4, however, the false negative outcomes lead to

---

[2] For simplification the conventional strategy is illustrated as one test and the intervention as two tests, however, in the case of primary rectal cancer the conventional strategy involves two tests and the intervention three tests.

inaccurate under-staging, identifying no nodal involvement (AJCC 1 and 2) when the patients do have nodal involvement (AJCC 3 and 4). The outcomes identified are detailed at the terminal node in Figure 3-2, where AJCC3 is identified as AJCC1 or AJCC2, and AJCC 4 is identified as AJCC 1 or AJCC2. The model does not allow for inaccuracies in staging between stages AJCC 3 and 4 or between AJCC1 and 2, as the purpose of the imaging scans (and therefore the model) is to identify spread to the lymph nodes. Therefore the model assumes that once lymph node spread has been identified as either positive or negative, the corresponding AJCC classifications (AJCC 1 & 2 for no nodal involvement or AJCC 3 & 4 for nodal involvement) are applied based on the AJCC stage specific prevalence.

The bottom half of the top branch, which split at the disease prevalence chance node, represents primary cancer with no nodal spread (AJCC 1 and 2). The tree depicts the negative test outcomes that were accurate (true negatives), i.e. patients who are staged as AJCC1 or 2 accurately, and also positive test outcomes which were inaccurate (false positives), again depending on the sensitivity and specificity of the test. These false positive outcomes diagnose nodal involvement, over-staging the extent of the disease which is actually no nodal involvement (inaccurately diagnosing AJCC 1 as AJCC 3 or AJCC 4, and AJCC2 as AJCC 3 or AJCC 4). In this way the decision tree separates out accurate and inaccurate diagnosis of nodal involvement.

Patients in the conventional arm of the model will be staged using the standard diagnostic test work-up described above (ceCT of the chest, abdomen and pelvis in colon cancer, and ceCT followed by a pelvic MRI in rectal cancer), represented by 'Test' in the top half of the tree in Figure 3-2. Patients in the intervention arm of the model are also given these conventional imaging tests represented by 'Test', followed by the addition of a PET/CT scan, represented as 'PET/CT' in the bottom half of Figure 3-2. In the intervention arm, the PET/CT test is given regardless of the outcomes from the initial test, and therefore the diagnostic test accuracies from both tests are combined. As PET/CT is an add-on test after the conventional test, a strategy is required to deal with the combined diagnostic test accuracy outcomes, particularly when one of the tests refutes the other (as discussed in section 3.3.1). The clinicians

advised that a confirmatory positive strategy is adopted in practice, and therefore the baseline model adopts such a strategy, whereby positive test results override negative results. Negative results from the conventional imaging tests which are refuted by the PET/CT test are treated as positive and vice versus, so that results are only treated as negative when both the conventional and PET/CT test outcomes are negative.

In the intervention arm of the tree, the outcomes are different from the top half of the tree, due to the combined DTA and the confirmatory positive strategy adopted. In Figure 3-2 the top branch in the intervention arm represents primary cancer with nodal spread (AJCC 3 and 4). The tree depicts the conventional test outcomes which are either positive (sensitivity 1) or negative (1-sensitivity 1) and is then directly followed by the PET/CT test, regardless of the outcomes from the conventional test. The DTA of the PET/CT test is then applied in the tree, distinguishing between positives (sensitivity 2) and negatives (1-sensitivity 2). In the top layer of this branch both tests were positive and therefore stages AJCC 3 and AJCC 4 are identified accurately. At the lower level of this branch, where the PET/CT chance node is negative, it conflicts with the conventional tests positive result, however, the positive outcome takes precedence due to the strategy adopted, and therefore this branch also identifies AJCC 3 and 4 accurately. Only when both the conventional and PET/CT tests report negative outcomes are the results treated as such. Figure 3-2 reports the AJCC stages identified, and whether these are accurate or inaccurate at the terminal node.

The proportion of accurate and inaccurate nodal staging outcomes at the end of the decision tree branches for the conventional arm of the model are compared against the intervention arm of the model and short-term cost-effectiveness can be assessed in terms of the incremental cost per accurate diagnosis. These interim decision model outcomes of accurate and inaccurate diagnosis for the four AJCC stages were then used to incorporate patient management in the model, as discussed next.

Treatment options

Following the development of a decision tree to map the diagnostic pathway for primary (rectal and colon) cancer, the economic model was adapted to incorporate the treatment impacts of accurate and inaccurate staging in primary colorectal cancer. The optimal treatment strategies for each AJCC stage were identified, along with their associated costs and utilities and applied in the model depending on the stage identified (diagnosis) from the decision tree.

Optimal treatment combinations for the four AJCC stages were determined through consideration of the literature (112-118) and in consultation with clinical experts. The various treatment options vary slightly between colon and rectal cancer, with the main difference being that radiotherapy is only appropriate for rectal cancer patients. Therefore the treatment combinations for both types of cancer were ascertained. Figure 3-3 details the optimal treatment options as decision tree branches, which were applied to each AJCC stage identified in the model for both rectal and colon cancer.

For primary colorectal cancer the model assumed that all AJCC 1 patients receive primary surgery; this is the only treatment option for patients with stage AJCC1. In the case of primary rectal cancer surgery entails rectal excision with lymphadenectomy, and for colon cancer a colonic resection with lymphadenectomy.

AJCC stage identified   **Primary Rectal treatment options**

AJCC 1   Surgery (rectal excision with lymphadenectomy)

                                Surgery

AJCC 2   Longcourse chemo-radiotherapy then surgery

                                Surgery followed by adjuvant chemotherapy

                                Surgery

AJCC 3   Longcourse chemo-radiotherapy then surgery

                                Surgery followed by adjuvant chemotherapy

                                Surgery

                                Longcourse chemo-radiotherapy then surgery

AJCC 4   Surgery followed by metastatic surgery

                                Surgery followed by palliative care

                                Palliative care

AJCC stage identified   **Primary Colon treatment options**

AJCC 1   Surgery (colonic resection with lymphadenectomy)

                                Surgery

AJCC 2   Surgery followed by adjuvant chemotherapy

                                Surgery

AJCC 3   Surgery followed by adjuvant chemotherapy

                                Surgery

                                Surgery followed by metastatic surgery

AJCC 4   Surgery followed by palliative care

                                Palliative care
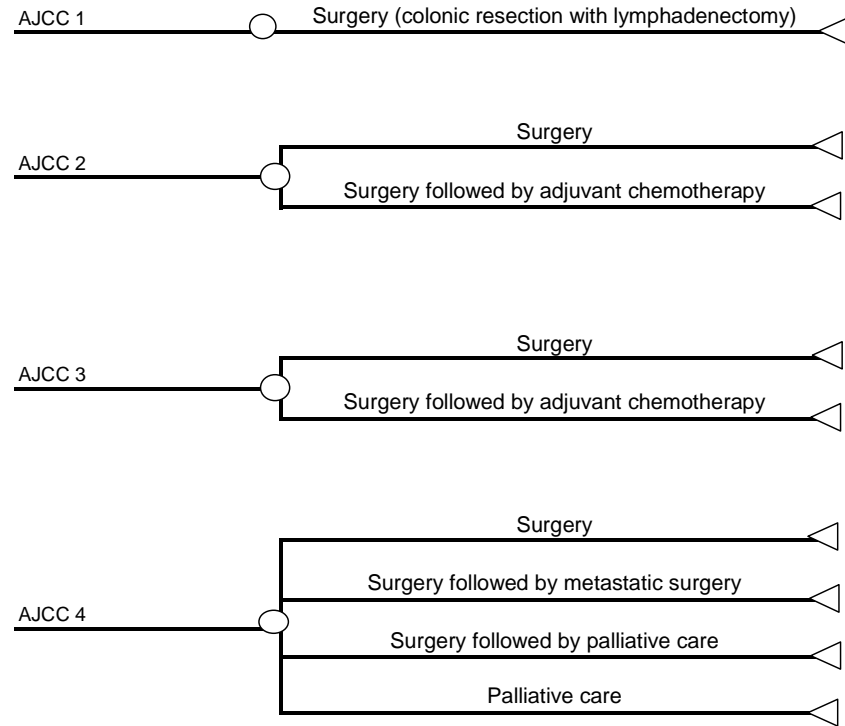
Figure 3-3: Treatment pathways for primary rectal & colon cancers

For patients with primary rectal cancer identified as stage AJCC2 or AJCC 3, there are three treatment options: surgery alone, long course chemo-radiotherapy prior to surgery or surgery followed by adjuvant chemotherapy. Surgery refers to a rectal excision with lymphadenectomy, long course chemo-radiotherapy refers to 5 weeks of radiotherapy (45Gy in 25 fractions over 5 weeks) followed by a 12 week course of chemotherapy of 5 Fluorouracil (5FU) (119;119), and adjuvant chemotherapy involves a 6 months course post surgery of intravenous 5FU and Oxaliplatin (100;120). Primary colon cancer patients identified as stage AJCC2 or AJCC 3, have two treatment options: surgery alone, or surgery followed by adjuvant chemotherapy. Surgery refers to a colonic resection with lymphadenectomy, and adjuvant chemotherapy involves a 6 months course post surgery of intravenous 5FU and Oxaliplatin (100;120).

AJCC4 rectal cancer patients can receive one of five treatment options: primary surgery alone, long course chemo-radiotherapy prior to primary surgery, primary surgery followed by metastatic surgery, primary surgery followed by palliative care or palliative care alone. Surgery refers to a rectal excision with lymphadenectomy, long course chemo-radiotherapy refers to 5 weeks of radiotherapy (45Gy in 25 fractions over 5 weeks) followed by a 12 week course of chemotherapy of 5 Fluorouracil (5FU) (97;119), and adjuvant chemotherapy refers to a 6 months course post surgery of intravenous 5FU and Oxaliplatin (100;120). Metastatic surgery refers to surgery at the metastatic site and palliative care, which is treatment designed to relieve symptoms and improve quality of life (as opposed to having a curative intent), represents an array of palliative treatments which may include chemotherapy or chemo-radiotherapy for rectal cancer patients. In the case of colon cancer AJCC 4 patients have four treatment options: primary surgery alone[3] (colonic resection with lymphadenectomy), primary surgery followed by metastatic surgery, primary surgery followed by palliative care or palliative care alone.

---

[3] There is typically a fifth treatment option of primary surgery combined with contemporaneous resection of metastases, however, after consultation with clinical experts it was decided this additional treatment option would only result in an unquantifiable and likely marginal effect on overall cost compared to primary surgery alone, hence these are considered together.

Assigning these optimal treatment options for each AJCC stage in the model means that patients in the decision tree who are accurately diagnosed will receive optimal treatment while patients who are inaccurately diagnosed (through false positive or false negative test outcomes) will receive suboptimal treatment, i.e. patients with no nodal involvement (AJCC1 or AJCC2 patients) who are inaccurately diagnosed as having nodal involvement (over-staged to either AJCC3 or 4) will receive unnecessary AJCC3 or AJCC4 treatments. In the case of inaccurate staging, the model assumes patients will receive the treatments for their (mis)diagnosed stage, but within a year their true diagnosis will be correctly identified and optimal treatment will then be given. This assumption was made in consultation with clinical experts and was considered to be valid with one year as an appropriate time scale for encompassing most cases of under-staging. In this way the model accounts for the cost and quality of life of optimal treatment and in the treatments which are received unnecessarily or initially failed to be received, due to over or under-staging. No transitions between nodal states are allowed during the year. The treatment profiles were assigned costs and utility weights in the model, which will be discussed later in section 3.4.1.

Survival

Following treatment for the diagnosed AJCC stage, the model incorporated a survival analysis in order to capture the mortality and quality of life impacts for the patients in each intervention. The survival analysis utilised a basic two-state Markov format for each of the four AJCC stages to calculate the average life expectancy and average quality adjusted life years (QALYs) in each AJCC stage. Figure 3-4 illustrates the model for the four AJCC colorectal cancer stages.   The first state represents the patients AJCC colorectal cancer stage, and the patient can either remain in this state or die moving into the death state.
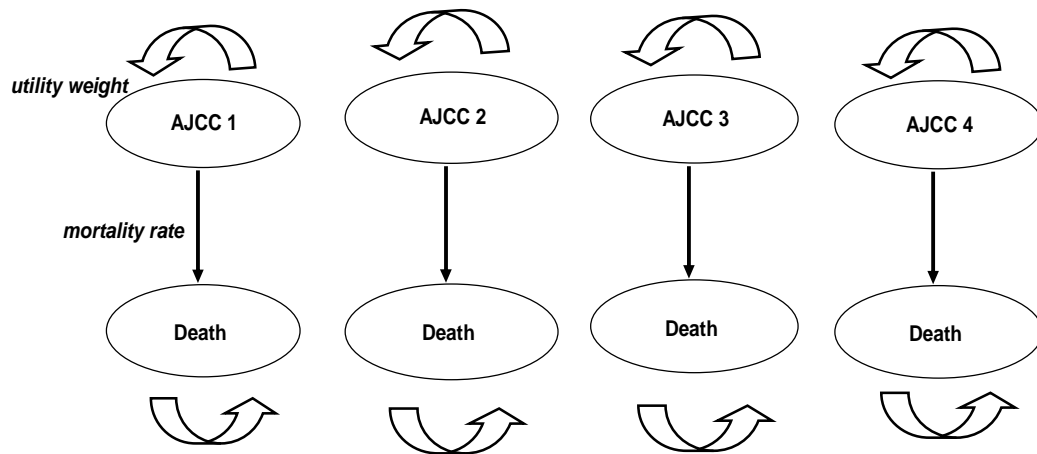
Figure 3-4: Two-state survival model for each AJCC colorectal cancer stage

The survival analysis assumed four patient cohorts, one for each AJCC stage, starting with the sample of patients in each stage in the Scottish CRC dataset (121). The model used annual cycles and assumed a starting age of 50 years[4]. A utility weight is applied to the AJCC state to reflect that state's average utility for five years post diagnosis; each of the four AJCC states has a different mean utility weight (122). Studies have shown (122;123) that five years post-diagnosis quality of life in colorectal cancer patients is on par with age specific general population utility weights, and therefore after five years in the AJCC state, it was assumed that patients who are still alive will have age specific population utility weights for the remainder of time in that state (124). Transition from the AJCC state to death was represented by the colorectal cancer AJCC stage specific morality rates for the first 10 years of the model (based on Scottish colorectal cancer five year overall survival estimates for each AJCC stage (121)). Returning to age specific population mortality rates five years post diagnosis was considered lenient as there is evidence of disease related death beyond five years (94), and therefore a more conservative approach was adopted whereby patients were considered to be cured 10 years post diagnosis. Beyond 10 years patients were assumed to have survived their

---

[4] The starting age of 50 years was used in the model as the dataset is based on the Scottish CRC population aged 50yrs and over (mean age 61 yrs). The models were also run using an older population (starting at age 70yrs), with the resultant effect of lowering life expectancy and quality adjusted life expectancy for patients in each AJCC stage, but no overall change to the incremental cost-effectiveness outcomes.

cancer and were assigned the age specific population level mortality rates (which were derived from routine data, and included all mortality including that from cancers)(125).

Figure 3-5 demonstrates the survival curves for each of the AJCC stages in patients suspected of primary rectal cancer. The population in each AJCC stage group was taken from the Scottish CRC dataset (121); however as the survival analysis is used to calculate average life expectancy in each group, the sample size at diagnosis (time zero) for each group in Figure 3-5 was assumed to be the same (n=1000) in order to clearly illustrate the difference in survival curves between the different AJCC groups.
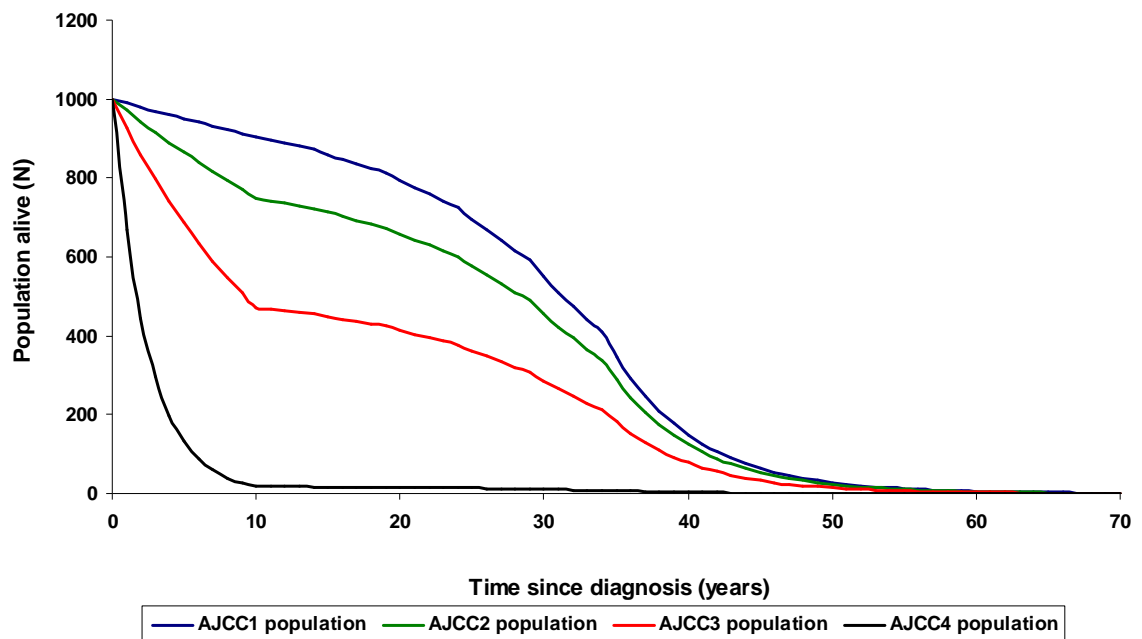
Figure 3-5: Overall survival for primary rectal cancer for each AJCC stage

As would be expected from AJCC 3 and AJCC4 patients, who have lymph node involvement in their primary cancer (and metastases to other sites in the body for AJCC 4), survival falls sharply in the first few years post-diagnosis, with only a small number of AJCC 4 patients surviving beyond ten years. The model starting age was set at 50 years for each AJCC group, based on the information from the Scottish dataset (121). The AJCC4 population have been diagnosed

with metastatic disease at their first (primary) diagnosis. This population have a very low life expectancy, and at 5 years post-diagnosis only 13% remain alive, and 10 years post diagnosis only 1.8% remain alive.

Note how AJCC stages 1 through to 3 survival curves all flatten out after 10 years; this is due to the change to population mortality rates. If they have survived this long, they are considered to be cured from their cancer and subject to same life expectancy as the general population, for their age. Ten years post diagnosis, the age specific population mortality rates kick in, and after 25 years the survival curves begin to decline more rapidly (population age 75 years). A very small percentage of the AJJC 1 & 2 populations live until approximately 100 years old. This is based purely on UK age specific population mortality rates, which incorporate all causes of death, including cancer

This survival analysis generated average life expectancies and average quality adjusted life years (QALYs) for each of the AJCC stages. These represent the average life expectancy and QALYs for patients who were accurately staged. It was assumed that patients who were inaccurately staged would be re-diagnosed within one year to their accurate diagnosis; however they would suffer negative impacts on their mortality and quality of life where appropriate. Publications (112;114;126-128) indicate that in comparison to conventional imaging modalities PET/CT scanning has no impact on overall survival; however, consultation with clinical experts highlighted that patients with AJCC 3 stage cancer (nodal involvement but no metastases) who fail to receive adjuvant chemotherapy due to inaccurate staging, may suffer a reduction in overall survival. This was incorporated into the model for AJCC 3 patients who were inaccurately under-staged as AJCC1 or 2, as a 25% reduction in overall survival. No other survival deductions were applied for inaccurate diagnoses; given the evidence form the literature and consultation with clinical experts.

With regards to adverse impacts on quality of life, a deduction was applied to utility to account for the impact of misdiagnosis where patients failed to receive treatment (due to false negative diagnoses) or received unnecessary

treatment (due to false positive diagnoses).   With regards to the disutility applied, this was used to account for the negative health impacts of delayed (or no) treatment, where it would have been optimal if given. Utility values are applied to the various (true disease) states to reflect quality of life in that state, given optimal treatment. A 6 month or 1 year disutility was added to cases that were inaccurately diagnosed, to reflect the negative impact on their health of failing to receive treatment. While in the short term invasive treatments may reduce quality of life (and is accounted for in the model), ultimately these treatments are used to improve the health of the patient. By failing to receive necessary treatment, it was felt that a reduction in quality of life was required to reflect the intermediate (up to 1 year) impact of this on the patients overall health state.

Table 3-5 details the sub-optimal treatments received due to inaccurate staging, through false positive and false negative diagnoses, and the duration of negative impact this has on quality of life in the model.

Table 3-5: Inappropriate treatments for inaccurate staging

Inaccurate staging	Inappropriate treatment	Impact on quality of life
AJCC 3 as 1	Fail to receive long course chemo-radiotherapy	6 months disutility
	Fail to receive adjuvant chemotherapy	6 months disutility
AJCC 3 as 2	Fail to receive adjuvant chemotherapy	6 months disutility
	Receive unnecessary long course chemo-radiotherapy	6 months alternative utility
AJCC4 as 1	Fail to receive long course chemo-radiotherapy	6 months disutility
	Fail to receive metastatic surgery and/or palliative care	1 year disutility
AJCC 4 as 2	Fail to receive metastatic surgery and/or palliative care	1 year disutility
AJCC 1 as 3	Receive unnecessary long course chemo-radiotherapy	6 months alternative utility
	Receive unnecessary adjuvant chemotherapy	6 months alternative utility
AJCC1 as 4	Receive unnecessary long course chemo-radiotherapy	6 months alternative utility
	Receive unnecessary metastatic surgery and/or palliative care	1 year alternative utility
AJCC 2 as 3	Fail to receive long course chemo-radiotherapy	6 months disutility
	Receive unnecessary adjuvant chemotherapy	6 months alternative utility
AJCC 2 as 4	Receive unnecessary metastatic surgery and/or palliative care	1 year alternative utility
	Fail to receive long course chemo-radiotherapy	6 months disutility
	Fail to receive adjuvant chemotherapy	6 months disutility

These inappropriate treatments correspond to the inaccurate diagnoses detailed in Figure 3-2, whereby false negative diagnoses lead to stage AJCC 3 being identified as AJCC 1 and AJCC 2, and stage AJCC 4 being identified as AJCC 1 and AJCC 2; while false positive diagnoses lead to stage AJCC 1 being identified as AJCC 3 and AJCC 4, and stage AJC C2 being identified as AJCC 3 and AJCC 4. The parameter values and references are detailed in section 3.4

The utility estimates were combined in the survival analysis and discounted at 3.5% (24) to derive discounted quality adjusted life expectancies for each AJCC stage.

3.3.3 Model 2: Recurrent colorectal cancer

The model developed for recurrent colorectal cancer follows a similar structure to the primary colorectal cancer model, incorporating the short term diagnostic pathway for the patient, the resultant treatment options and longer term outcomes such as patient survival and quality of life. Development of the recurrent model is now discussed.

Diagnostic Pathway

The recurrent model was developed to assess the cost-effectiveness of PET/CT as an add-on technology in detecting recurrent rectal and colon cancer. The additional value of incorporating a PET/CT scan to conventional imaging in this disease stage is through the ability to confirm or refute local recurrence and potentially identify metastatic recurrence.

Figure 3-6 depicts the decision tree structure used for the recurrent model. This was altered to include the disease specific criterion for rectal and colon cancer separately. The model structure was informed by the available literature (129;130) and based on consultation with clinical experts.

The recurrent decision tree model begins with patients who have previously had surgical treatment for primary rectal or colon cancer and in a routine follow-up assessment (involving clinical examination, routine imaging and CEA testing) were found to have rising CEA levels, which identified them as potentially having recurrent rectal or colon cancer. The decision tree then outlines the choice between conventional diagnostic testing and the add-on PET/CT strategy. The standard procedure for patients suspected of recurrent colon cancer involves a ceCT scan of the chest, abdomen and pelvis, while in recurrent rectal patients it involves a ceCT scans of the chest, abdomen and pelvis and an MRI scan of the pelvis, to confirm or refute local recurrence and assess whether this is an isolated recurrence or associated with distant metastases. The intervention arm also involves the conventional test, followed by a PET/CT test regardless of the conventional test outcomes.
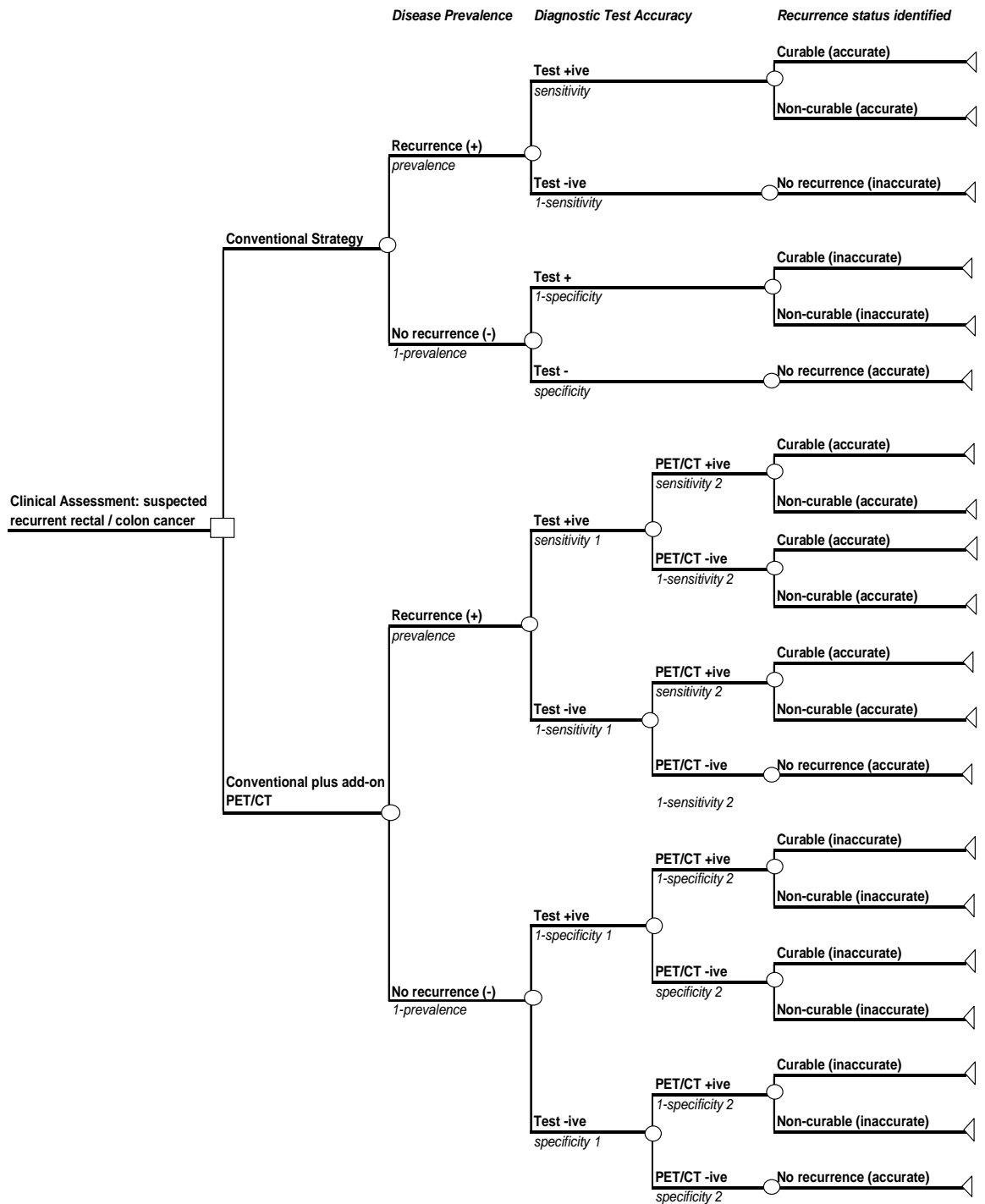
Figure 3-6: Decision tree for staging recurrent colorectal cancer

In Figure 3-6 the conventional pathway is represented in the top half of the tree, while the intervention pathway is represented in the bottom half.

Similar to the structure used in the primary models, this decision tree model has been designed using actual disease status, and therefore after the decision node depicting the choice between the conventional and intervention strategies, the tree splits the patient population according to their true disease status, so that accurate and inaccurate diagnoses can be identified. The objective of the scan in this model is to assess whether there has been any recurrence of disease and therefore the tree divides disease prevalence into recurrence (isolated local or local combined with distant metastases) and no recurrence populations. The standard work-up of diagnostic tests are then undertaken, represented by 'Test' in Figure 3-6, which will either identify recurrence (test positive), or no recurrence (test negative). Having previously specified actual recurrence status, the top branch for the conventional strategy represents recurrent cancer, and therefore the tree branch splits depending on whether the test was positive (accurately identified recurrence) or negative (inaccurately identifying no recurrence). Positively identified recurrence is then further separated into curable and non-curable recurrence which will involve different treatment options. In the top half of this branch, negative test outcomes represent false negatives, which lead to patients being inaccurately diagnosed as having no recurrence. The objective of the scan in this model is to assess whether or not there has been any recurrence of disease and therefore the model assumes that there is no inaccurate diagnoses between recurrence curable and recurrence incurable. Inaccurate diagnoses are dependent on the DTA for the test in identifying recurrence or nor recurrence.

The bottom branch in the top half of the tree represents the disease status of no recurrence, so negative test outcomes accurately indicate no recurrence. Positive test outcomes in this branch of the tree are false positives which inaccurately diagnose recurrence when there is no recurrence. This population is further divided into curable and non-curable recurrence in order to determine inaccurate treatment in the next stage of the model.

In the intervention arm of the tree, the conventional 'Test' is followed by the 'PET/CT' scan in Figure 3-6, and therefore the test outcomes must be combined, and a strategy adopted for dealing with conflicting results. As with

the primary model a confirmatory positive strategy was the baseline approach, however, if one test was found to have superior test performance characteristics over the other, i.e. superior in terms of both sensitivity and specificity, then the test with superior DTA can be used to represent the joint DTA. This approach of using superior test performance to represent joint imaging modalities has been used previously for in add-on PET/CT screening for cervical cancer (108).

The accurate and inaccurate identification of recurrence at the end of the decision tree branches for the conventional arm of the model can be compared against the intervention arm and assessed in terms of the incremental cost per accurate diagnosis. These interim outcomes of accurate and inaccurate diagnosis were then used to assess the impact on patient management in the model, incorporating optimal treatments for curable recurrence, non-curable recurrence and no recurrence, as discussed next.

Treatment Options

Following the diagnostic pathway for recurrent rectal and colon cancer, the model incorporates the treatment impacts of accurate and inaccurate diagnoses of recurrent colorectal cancer. Optimal treatment combinations for curable and non-curable recurrence were determined through the literature (115-118;131) and in consultation with clinical experts. The treatment options vary slightly between recurrent rectal and colon cancer and therefore treatment combinations for each type of cancer were ascertained. Figure 3-7 details the optimal treatment options as decision tree branches, which were applied to the diagnoses from the decision tree (Figure 3-6) for both recurrent rectal and colon cancer.

Diagnosis   **Recurrent Rectal treatment options**

Local surgery

Local surgery followed by adjuvant chemotherapy

Long course chemo-radiotherapy then surgery

Recurrence Curable   Local surgery followed by metastatic surgery

Local surgery followed by adjuvant chemotherapy then
metastatic surgery

Long course chemo-radiotherapy followed by local
surgery then metastatic surgery

Metastatic surgery followed by palliative care

Recurrence Incurable

Palliative care alone

No recurrence   Wait and watch with an annual follow-up

Diagnosis   **Recurrent Colon treatment options**

Local surgery

Local surgery followed by adjuvant chemotherapy

Recurrence Curable   Local surgery followed by metastatic surgery

Local surgery followed by adjuvant chemotherapy then
metastatic surgery

Metastatic surgery followed by palliative care

Recurrence Incurable

Palliative care alone

No recurrence   Wait and watch with an annual follow-up

Figure 3-7: Treatment pathways for recurrent rectal & colon cancers

Based on clinicians advice the model assumed that 40% of recurrent rectal cancer patients would have received radiotherapy as part of their treatment for primary cancer and therefore would not receive further radiotherapy, while the remaining 60% of those patients who subsequently developed local recurrence but who did not receive radiotherapy for their primary cancer, would receive long course chemo-radiotherapy prior to surgery for recurrent disease. Patients with curable rectal recurrence had one of six treatment options: local surgery alone, local surgery followed by adjuvant chemotherapy, long course chemo-radiotherapy prior to local surgery, local surgery followed by metastatic surgery, local surgery and adjuvant chemotherapy followed by metastatic surgery or long course chemo-radiotherapy prior to local surgery followed by metastatic surgery. As with the primary colon model, radiotherapy is not included as a treatment option for recurrent colon cancer. Therefore, patients with curable colon recurrence had one of four treatment options: local surgery alone, local surgery followed by adjuvant chemotherapy, local surgery followed by metastatic surgery or local surgery and adjuvant chemotherapy followed by metastatic surgery.

Local surgery refers to a rectal excision with lymphadenectomy for recurrent rectal cancer patients and a colonic resection with lymphadenectomy for recurrent colon cancer. Long course chemo-radiotherapy refers to 5 weeks of radiotherapy (45Gy in 25 fractions over 5 weeks) followed by a 12 week course of chemotherapy of 5 Fluorouracil (5FU) (97;119), and adjuvant chemotherapy involves a 6 months course post surgery of intravenous 5FU and Oxaliplatin (100;120). Metastatic surgery refers to surgery at the metastatic site.

Patients with incurable recurrence had one of two treatment options: metastatic surgery followed by palliative care, or palliative care alone. This is the same for both rectal and colon recurrence. Palliative care, involves treatments designed to relieve symptoms and improve quality of life (as opposed to having a curative intent), and therefore represents an array of palliative treatments which may include chemotherapy or chemo-radiotherapy for rectal cancer patients. Patients diagnosed with no recurrence are treated with a wait and watch strategy in which they have an annual oncology follow-up.

Assigning these optimal treatment options for each diagnoses (curable recurrence, incurable recurrence and no recurrence) in the model means that patients in the decision tree who are accurately diagnosed will receive optimal treatment while patients who are inaccurately staged (through false positive or false negative test outcomes) will receive suboptimal treatment i.e. patients with no recurrence who are inaccurately diagnosed as having curable recurrence will receive one of the curable treatment options unnecessarily. In the case of inaccurate diagnosis, the model assumes patients will receive the treatments for their (mis)diagnosed stage, but within a year their true diagnosis will be correctly identified and optimal treatment will then be given. This assumption was made in consultation with clinical experts and was considered to be valid with one year as an appropriate time scale for encompassing most cases of misdiagnosis. In this way the model accounts for the appropriate treatments and the treatments which are received unnecessarily or initially failed to be received, due to over or under-staging. The treatment profiles were assigned costs and utility weights in the model, and will be discussed in section 3.4.2.

Survival

Following treatment for diagnosis and misdiagnosis (recurrence curable, recurrence incurable, no recurrence), the model incorporated a lifetime analysis in order to capture the mortality and quality of life impacts for the patients in each intervention.

The lifetime analysis was implemented using an approach similar to that used in the primary model. A basic two-state Markov format was used for each of the three recurrence statuses, as depicted in Figure 3-8, to calculate average life expectancy and average quality adjusted life years in each status. The first state represents the patients' true diagnosis (no recurrence, recurrence curable, or recurrence incurable), and the patient can either remain in this state or die, moving into the death state. The model uses annual cycles and assumed a starting age of 50 years.
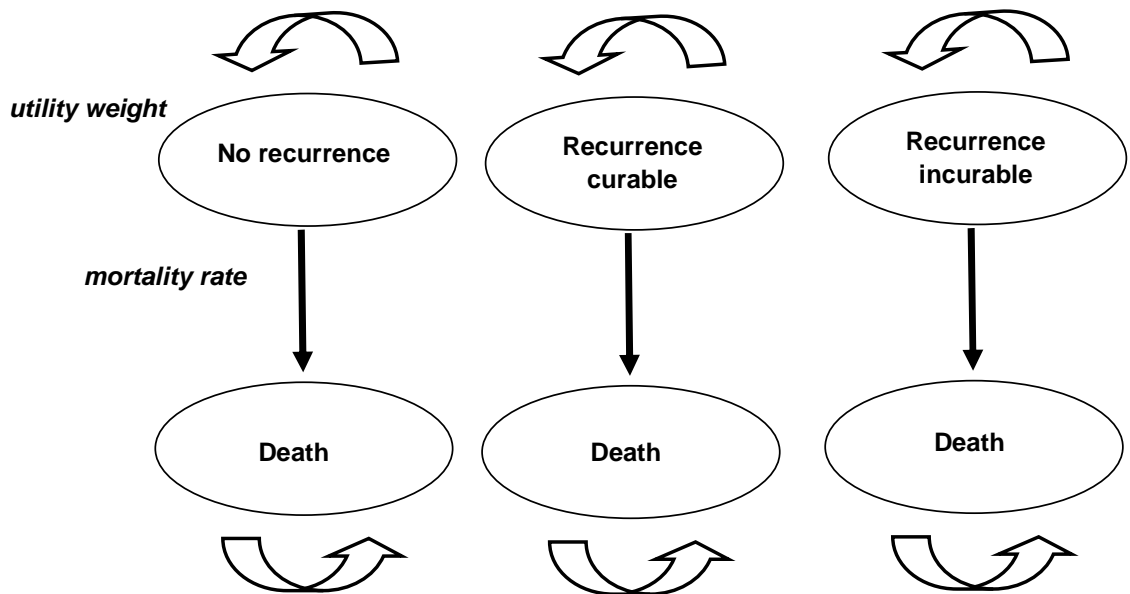
Figure 3-8: Two-state survival model for each recurrence status

A utility weight is applied to each of the disease states to reflect that states average utility for 5 years post diagnosis; each state has a different mean utility weight (122;132). After five years in the disease state, it was assumed that patients who are still alive will have age specific population utility weights for the remainder of time in that state. Five year overall survival estimates were determined from the literature for patients with no recurrence, recurrence that is curable and for non-curable recurrence (133). This 5 year overall survival data was used to determine an annual mortality rate under the assumption of an exponential survivor function, and was used to represent transitions from the recurrence state to death for the first 10 years of the model. Beyond ten years patients were assumed to have survived their cancer and were assigned the age specific population level mortality rates (125). The disease state utility weights were applied for five years post diagnosis, based on available evidence (122;123), however, disease state mortality rates were applied for 10 years post diagnosis, adopting a conservative approach which incorporates evidence of disease related death beyond five years (94). The model did not account for transitions between the various states as this analysis was used to determine the average life expectancy and QALYs for each state.

Figure 3-9 demonstrates the survival curves for each of the recurrence categories (no recurrence, recurrence curable, recurrence incurable) in patients suspected of recurrent rectal cancer. The Scottish network CRC dataset (121) was used to represent a cohort of AJCC1-3 patients who would be susceptible to colorectal recurrence (n=2409), and used disease prevalence estimates to determine no recurrence (n=723) and recurrence colorectal cancer (n=1686) samples, a subset of which was deemed to have curable recurrence (n=506), while the rest have incurable recurrence (n=1180). As the survival analysis is used to calculate average life expectancy in each group, the sample size at diagnosis (time zero) for each group in Figure 3-9 was assumed to be the same (n=1000) in order to clearly illustrate the difference in survival curves.
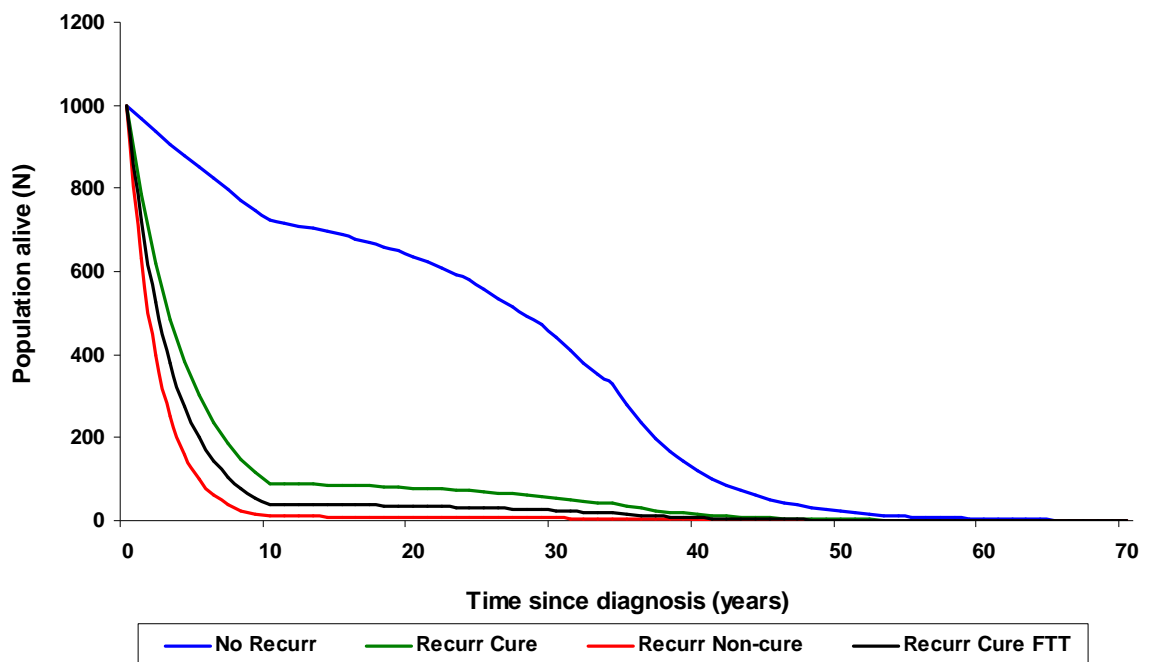
Figure 3-9: Overall survival for patients suspected of recurrent rectal cancer
Note: FTT= fail to treat, Recurr= recurrence

As would be expected the population with incurable recurrence have a drastic fall in their survival curve in the first few years post-diagnosis. The recurrence curable population see a steep decline in survival, while the no recurrence population have a steady decline, which flattens out for patients surviving beyond ten years, who then have mortality rates on par with the age specific general population.

This survival analysis generated average life expectancies and average quality adjusted life years (QALYs) for the no recurrence, recurrence curable and recurrence incurable states. These represent the average life expectancy and QALYs for patients who were accurately staged. It was assumed that patients who were inaccurately staged would be re-diagnosed within one year to their accurate diagnosis; however they would suffer negative impacts on their mortality and quality of life where appropriate.

Based on clinical advice, it was assumed that patients in the model who had curable recurrence but were inaccurately diagnosed as no recurrence (false negatives) and failed to receive treatment in the first year would have a negative impact on their mortality. Therefore five year mortality rates for these patients were determined to be midway between curable and non-curable mortality rates, and a fourth category was included in the survival analysis (and in Figure 3-9) to calculate the average life expectancy of patients with curable recurrences who initially fail to receive adjuvant chemotherapy negative diagnoses. As can be seen in Figure 3-9, using the same cohort population for curable recurrence, those who initially fail to receive treatment have a lower survival than those who are accurately staged and receive timely, optimal treatment. No other survival deductions were applied for inaccurate diagnoses; however there were implications in terms of patient quality of life.

With regards to adverse impacts on quality of life, a deduction was applied to utility to account for the impact of misdiagnosis where patients failed to receive treatment (due to false negative diagnoses) or received unnecessary treatment (due to false positive diagnoses). Table 3-6 details the sub-optimal treatments received due to inaccurate diagnoses, through false positive and false negative diagnoses, and the duration of negative impact this has on quality of life.

Table 3-6: Inappropriate treatments for inaccurate diagnosis

Inaccurate diagnosis	Inappropriate treatment	Impact on quality of life
Recurrence curable as No recurrence	Fail to receive local surgery	1 year disutility
	Fail to receive adjuvant chemotherapy	
	Fail to receive long course chemo-radiotherapy	
	Fail to receive metastatic surgery	
	Fail to receive palliative care	
Recurrence incurable as No recurrence	Fail to receive metastatic surgery	1 year disutility
	Fail to receive palliative care	
No recurrence as Recurrence curable	Receive unnecessary local surgery	1 year alternative utility
	Receive unnecessary adjuvant chemotherapy	
	Receive unnecessary long course chemo-radiotherapy	
	Receive unnecessary metastatic surgery	
	Receive unnecessary palliative care	
No recurrence as Recurrence incurable	Receive unnecessary metastatic surgery	1 year alternative utility
	Receive unnecessary palliative care	

These inappropriate treatments correspond to the inaccurate diagnoses detailed in Figure 3-6, whereby false negative diagnoses lead to Recurrence (curable and incurable) being misdiagnosed as No recurrence, and false positive diagnoses lead to No recurrence being misdiagnosed as Recurrence curable and incurable. The parameter values and references for this model are detailed in section 3.4.2.

3.3.4 Model 3: Metastatic cancer

The model developed for metastatic colorectal cancer follows the same structure as the first two models, incorporating the short term diagnostic pathway, the resultant treatment options and longer term outcomes such as patient survival and quality of life. The structure of the metastatic model is now discussed.

Diagnostic pathway

The metastatic model was undertaken to assess the cost-effectiveness of PET/CT as an add-on device in detecting metastatic cancer. The additional value of incorporating a PET/CT scan to conventional imaging in this disease stage is through its ability to detect unsuspected, metastatic disease and potentially identify unsalvageable extra metastases not detected by conventional imaging devices.

Figure 3-10 depicts the decision tree structure used for the metastatic model, informed by the literature (134-136) and based on consultation with clinical experts.
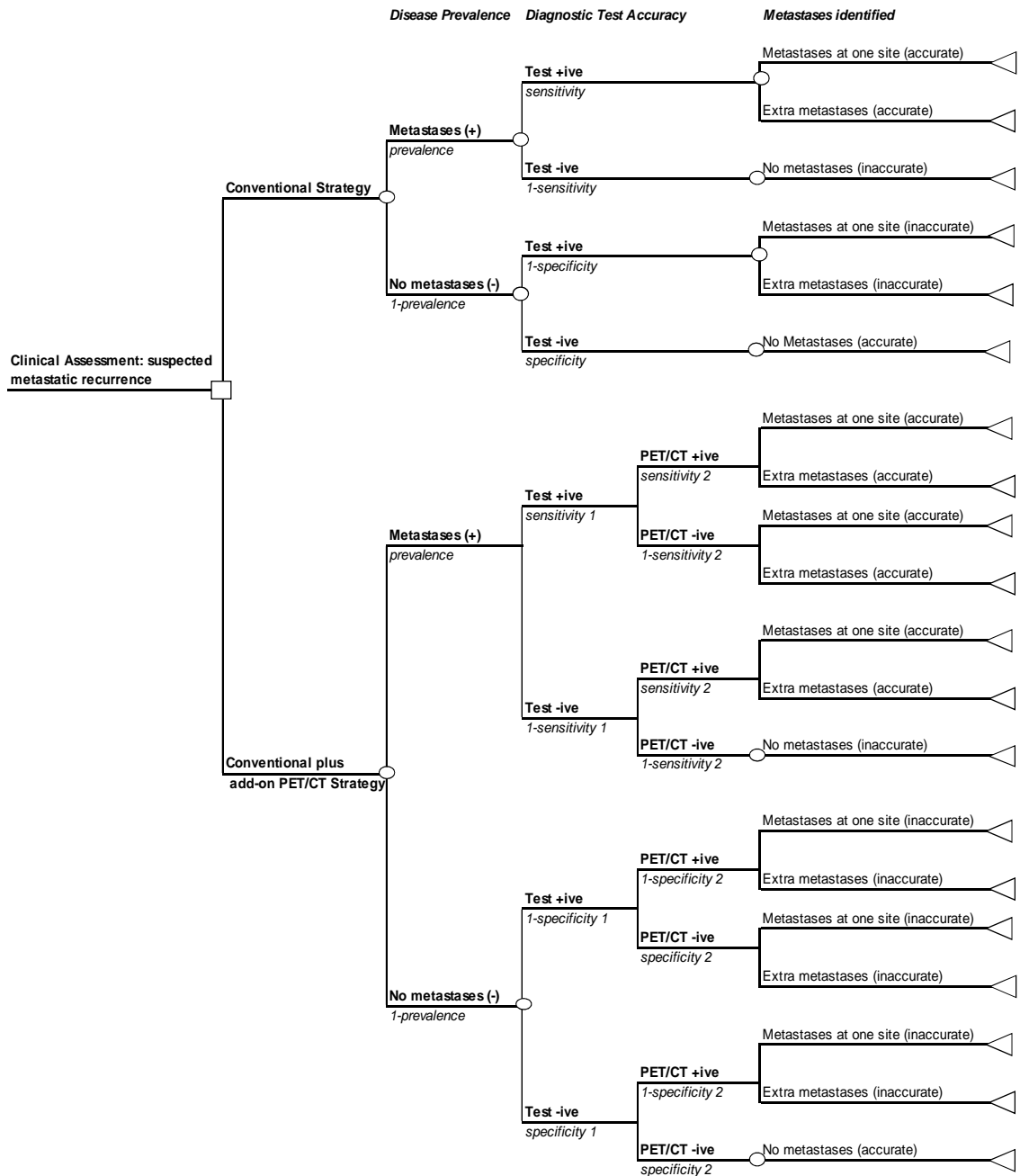
Figure 3-10: Decision tree for staging metastatic colorectal cancer

The metastatic decision tree begins with patients who have previously had surgical treatment for primary CRC and in a routine follow-up assessment (involving a clinical examination and CEA testing) were found to have rising CEA levels, and were identified as potentially having a metastatic recurrence. The decision node depicts the choice between the conventional or add-on PET/CT strategies. Similar to the structure used in the previous models, this decision tree has been designed using actual disease status, and therefore the

decision tree has split the patient population according to their true disease status (metastatic recurrence or no metastatic recurrence) prior to applying the DTA estimates for the tests, so that accurate and inaccurate diagnosis can be identified.

The conventional procedure for patients suspected of metastatic recurrence is to undertake a ceCT scan of the chest, abdomen and pelvis to confirm or refute metastatic recurrence and potentially identify additional sites of metastases. This is depicted as 'Test' in the upper branch of Figure 3-10, and will either identify metastases (test positive), or no metastases (test negative). In the conventional arm, having specified actual disease status, the top half of this branch represents metastatic recurrence, and therefore the tree branch splits depending on whether the test was positive (accurately identified metastatic recurrence) or negative (inaccurately identifying no metastatic recurrence). Positive identification of metastatic recurrence is further separated in this model to distinguish between metastases at one site or extra metastases at numerous sites, as the extent of the metastatic recurrence will impact on the treatment options in the longer-term model. The negative test outcomes in the top branch of the decision tree indicate a misdiagnosis of no metastatic recurrence (false negative).

The bottom half of the conventional tree branch represents the status of no metastatic recurrence, so negative test outcomes accurately indicate no metastases. Positive test outcomes in the bottom half of the tree are false positives which inaccurately diagnose metastatic recurrence when there is no recurrence. This population is then further divided to distinguish between inaccurate diagnosis of metastases at one site and inaccurate diagnosis of extra metastases at numerous sites. In this way the decision tree separates out accurate and inaccurate diagnoses of metastatic recurrence.

Patients in the 'conventional arm' of the model will be staged using the standard diagnostic test (ceCT chest, abdomen and pelvis), represented by 'Test' in the top half of Figure 3-10. Patients in the 'intervention arm' of the model will also be given the ceCT scan, followed by the addition of a PET/CT scan. This is represented in the bottom half of the tree, with the second 'PET-

CT' test added-on regardless of the conventional 'Test' outcomes. As this is an add-on strategy the test outcomes must be combined, and a strategy adopted for dealing with conflicting results. As with the primary and recurrent models a confirmatory positive strategy was the baseline approach and is depicted in the intervention arm in Figure 3-10. In the add-on strategy, if one test was found to have superior test performance characteristics over the other, i.e. superior in terms of both sensitivity and specificity, then the confirmatory positive strategy can be dropped and the test with superior DTA can be used to represent the joint DTA. This approach of using superior test performance to represent joint imaging modalities has been used previously in add-on PET/CT screening for cervical cancer (108).

The accurate and inaccurate identification of metastases at the end of the decision tree branches for the conventional arm of the model are compared against the intervention arm and assessed in terms of the incremental cost per accurate diagnosis. These interim outcomes of accurate and inaccurate diagnosis were then used to assess the impact on patient management in the model, incorporating optimal treatments for metastases at one site, extra metastases and no metastatic recurrence.

Treatment Options

Following the development of a diagnostic pathway for metastatic colorectal cancer, the model incorporates the treatment impacts of accurate and inaccurate diagnoses of metastatic recurrence. Optimal treatment combinations for metastatic recurrence at one site, extra metastases (at more than one site) and no metastatic recurrence were determined from the literature and in consultation with clinical experts. Figure 3-11 details the optimal treatment options as decision tree branches, which were applied to the diagnoses from the decision tree in Figure 3-10.

Diagnosis   **Metastatic treatment options**

Metastases at one site   Pre-operative chemotherapy followed by metastatic surgery

Pre-operative chemotherapy followed by metastatic surgery and palliative care

Extra metastases

Chemotherapy followed by palliative care

No metastases   Wait and watch with an annual follow-up
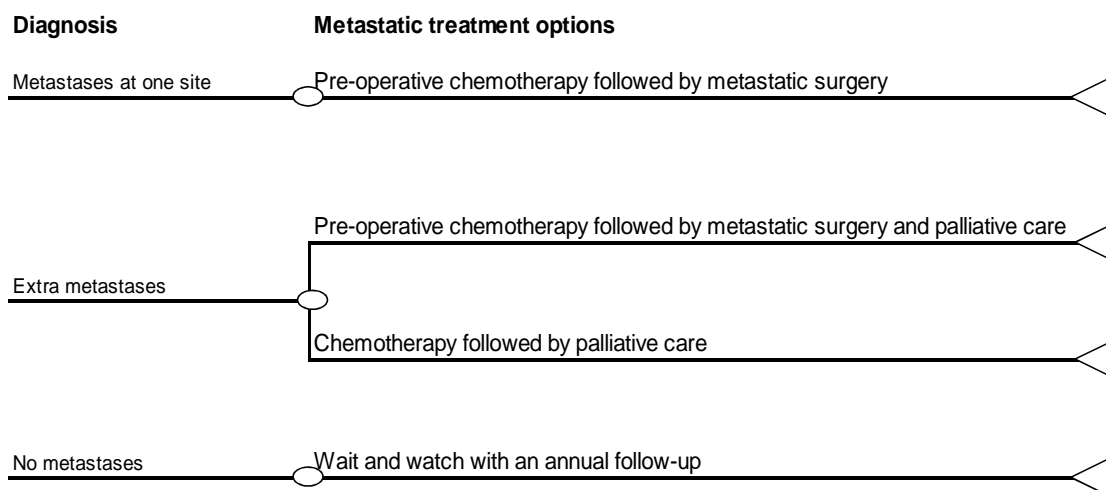
Figure 3-11: Treatment pathways for metastatic colorectal cancer

The model assumes that all patients with metastases at a single site will receive pre-operative chemotherapy and metastatic surgery. Pre-operative chemotherapy involves a 6 months course of intravenous 5FU and Oxaliplatin (100;120), and metastatic surgery refers to surgery at the metastatic site.

Patients with extra metastases are assumed to be non-curable and will receive one of two treatment options: pre-operative chemotherapy followed by metastatic surgery and palliative care, or chemotherapy and palliative care. Palliative care, involves treatments designed to relieve symptoms and improve quality of life, and therefore represents an array of palliative treatments.

It was assumed that all patients identified as having no metastatic recurrence would be treated with a wait and watch strategy in which they would be followed-up annually.

Assigning these optimal treatment options for each of the diagnoses means that patients in the decision tree who are accurately diagnosed will receive optimal treatment while patients who are inaccurately staged (through false positive or false negative test outcomes) will receive suboptimal treatment i.e. patients with no metastatic recurrence who are inaccurately diagnosed as having metastases will receive treatment for either metastases or extra metastases unnecessarily. In the case of inaccurate diagnosis, the model

assumes patients will receive the treatments for their (mis)diagnosed stage, but within a year their true diagnosis will be correctly identified and optimal treatment will then be given. This assumption was made in consultation with clinical experts and was considered to be valid with one year as an appropriate time scale for encompassing most cases of misdiagnosis. In this way the model accounts for the appropriate treatments and the treatments which are received unnecessarily or initially failed to be received, false positive and false negative diagnoses. The treatment profiles were assigned costs and utility weights in the model, which are discussed in section 3.4.3.

Survival

The survival analysis was implemented employing an approach similar to that used in the primary and recurrent models. Following treatment for diagnosis and misdiagnosis (no metastases, metastases at one site, extra metastases), the model incorporated a lifetime analysis in order to capture the mortality and quality of life impacts for the patients in each intervention. A basic two-state Markov format was used for each of the three disease states, as depicted in Figure 3-12, to calculate average life expectancy and average quality adjusted life years in each status.   The first state represents the patients' true diagnosis (no metastases, metastases at one site, or extra metastases), and the patient can either remain in this state or die, moving into the death state. The model uses annual cycles and assumed a starting age of 50 years.
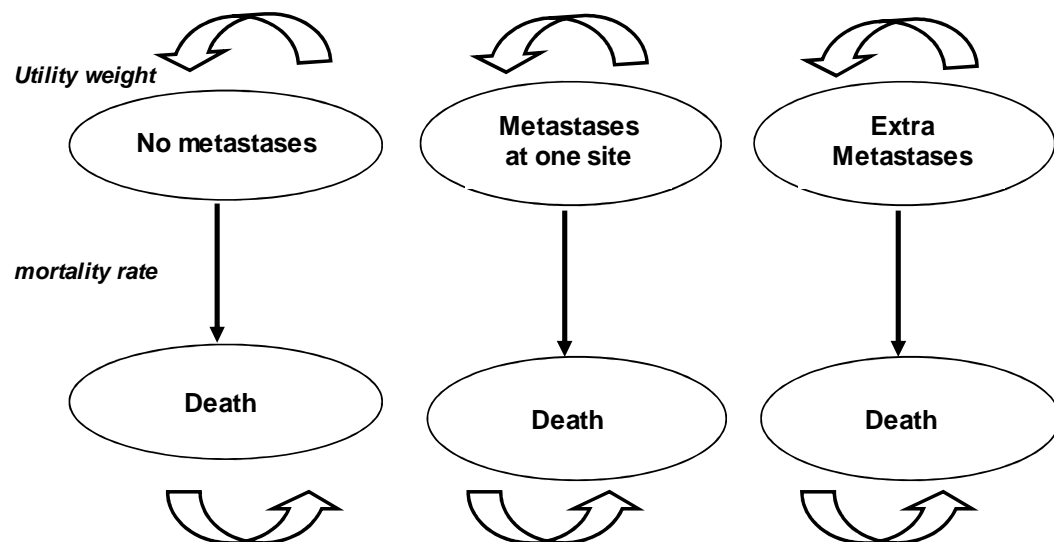
Figure 3-12: Two-stage survival model for metastatic disease status

A utility weight is applied to each disease state to reflect that states average utility for five years post diagnosis (122;132;137). After five years in each state it was assumed that patients who are still alive will have a quality of life similar to the general population and therefore age specific population utility weights were applied for the remainder of time in that state.

Transition from the disease state to death is represented by mortality rates. Five year overall survival estimates were determined from the literature for the no metastases and metastases at one site states (133;138). The extra metastases state was split between the two different treatment options for this state (surgery or palliative care alone), assigning a different five year overall survival estimates dependent on the treatment. This is because patients with extra metastases who receive metastatic surgery with palliative intent will have a greater five year survival estimate than patients with extra metastases who receive palliative care alone (133;138). The five year overall survival estimates were used to determine annual mortality rates under the assumption of an exponential survivor function. The disease state mortality rates (for no metastases, metastases at one site, extra metastases with surgery, and extra metastases palliative care only) were applied for the first 10 years in each state, and following this age specific population mortality rates

(125) were assigned for the remained of time in that state. The disease state mortality rates were applied for 10 years post diagnosis, adopting a conservative approach as in the primary and recurrent models.

Figure 3-13 demonstrates the resultant survival curves for the three disease states in patients suspected of metastatic recurrence, including two survival curves for extra metastases (surgery and palliative alone), which were assumed to have different five year overall survival estimates. As the survival analysis is used to calculate average life expectancy and QALYs in each group, the same sample size (n=600) was assumed for each group in order to clearly illustrate the difference in survival curves[5].
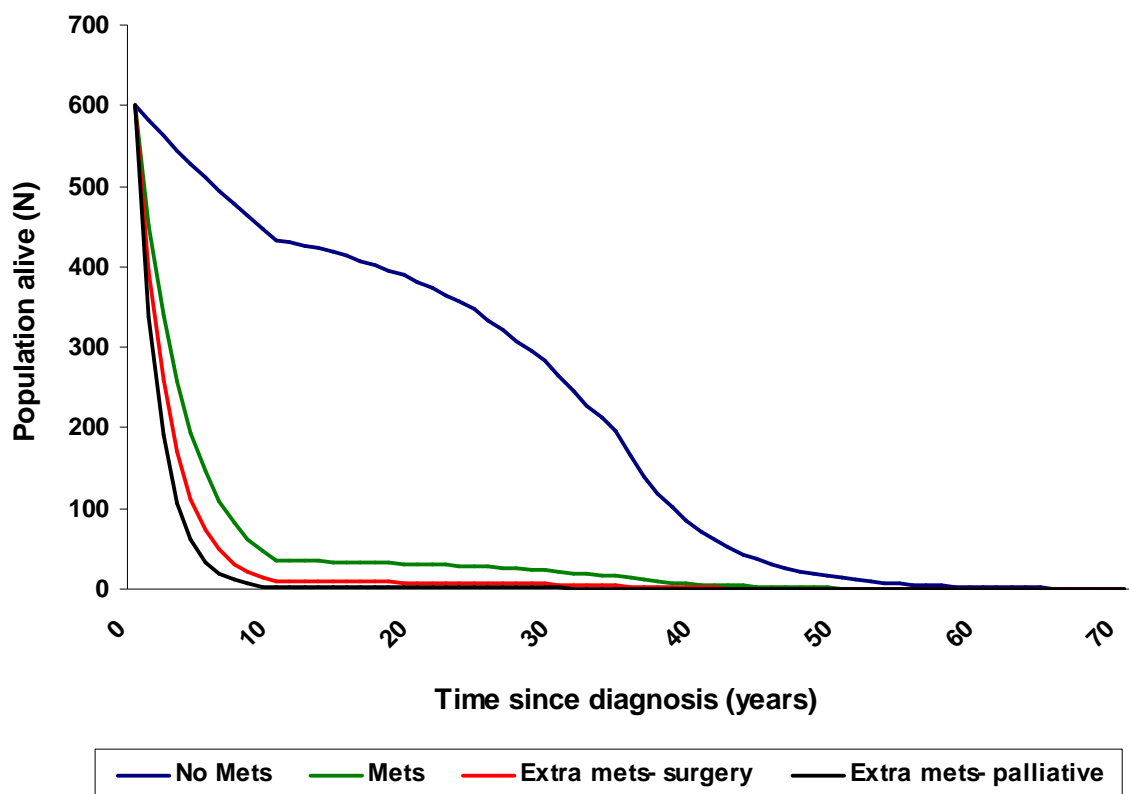
Figure 3-13: Overall survival for patients suspected of metastatic recurrence
No Mets=no metastases, Mets=metastases at one site

---

[5] Alternatively the Scottish network CRC dataset (121) can be used to represent a cohort of AJCC1-3 patients who would be susceptible to metastatic recurrence (n=2409), and use the disease prevalence estimates to determine no metastases (n=1445), metastases at one site (n=289) and extra metastases (n=675) populations. The same mean life expectancy would be derived regardless of the sample size, as it is the average.

As would be expected the no metastases population have a steady declining survival curve for the first ten years, which then levels off as the age specific general population mortality rates set in, and then begins to decline again about 25 years post diagnosis, when the cohort population age is 75. The population with extra metastases have the sharpest declining survival cure (with palliative intent alone falling sharper than those who receive surgery), while the sample with metastases at one site, also have a sharp drop in survival, but live longer, and many survive beyond ten years.

This survival analysis generated average life expectancies and average quality adjusted life years (QALYs) for each of the model diagnoses. These represent the average life expectancy and QALYs for patients who were accurately staged. It was assumed that patients who were inaccurately staged would be re-diagnosed within one year to their accurate diagnosis; however they would suffer negative impacts on their quality of life where appropriate. No survival deductions were applied for inaccurate diagnoses; given the evidence form the literature and consultation with clinical experts.

With regards to adverse impacts on quality of life, a deduction was applied to utility to account for the impact of misdiagnosis where patients failed to receive treatment (due to false negative diagnoses) or received unnecessary treatment (due to false positive diagnoses). Table 3-7 details the sub-optimal treatments received due to inaccurate diagnoses (false positives and false negatives) and the duration of negative impact this has on quality of life.

Table 3-7: Inappropriate treatments for inaccurate diagnosis

Inaccurate diagnosis	Inappropriate treatment	Impact on quality of life
Metastases 1 site as No metastases	Fail to receive pre-operative chemotherapy and metastatic surgery	1 year disutility
Extra metastases as No metastases	Fail to receive pre-operative chemotherapy, metastatic surgery and palliative care	1 year disutility
	Fail to receive chemo and palliative care	1 year disutility
No metastases as Metastases at 1 site	Receive unnecessary pre-operative chemotherapy and metastatic surgery	1 year alternative utility
No metastases as Extra metastases	Receive unnecessary pre-operative chemo, metastatic surgery and palliative care	1 year disutility
	Receive unnecessary chemo & palliative care	1 year disutility

These inappropriate treatments correspond to the inaccurate diagnoses detailed in the metastatic decision tree in Figure 3-10, whereby false negative diagnoses lead to Metastases (at one site and extra metastases) being misdiagnosed as No metastases, and false positive diagnoses lead to No metastases being misdiagnosed as Metastases at one site and extra metastases.

Having outlined the design and development of the three models (for primary, recurrent and metastatic colorectal cancer), the following section discusses the parameterisation and population for each of the models.

3.4 Model Parameterisation and population

3.4.1 Model 1: Primary rectal and colon

The model for primary colorectal cancer was populated with parameters representing the prevalence of AJCC disease status, the diagnostic test accuracy of the conventional and intervention imaging devices, the various treatment options, survival estimates and quality of life. Each of these is discussed in turn, followed by details of the associated costs.

Disease Prevalence

In the primary colorectal cancer model disease prevalence was defined as the spread of the primary cancer into the lymph nodes (nodal involvement) and therefore the AJCC stages were used to distinguish between nodal involvement (AJCC stages 3 & 4) and no nodal involvement (AJCC stages 1 & 2).

Prevalence data for the four AJCC stages was derived from a Scottish network dataset (121) provided by the clinical experts in the research team. The data comprises detailed clinico-pathological and imaging staging from an on-going study involving 2,383 Scottish CRC patients (average age 61 yrs). The dataset is a prospective series which identifies all cases of CRC in Scotland by direct clinical and nurse contact, through pathology department returns, managed clinical networks, cancer registration and death registrations. This series is considered to represent the generality of CRC in the UK as any differences in the epidemiology of CRC between Scotland and the rest of the UK will be marginal. The dataset provided information on CRC disease status using the AJCC colorectal cancer staging system along with five year overall survival data for each of the four AJCC stages. This dataset is discussed in full in a recent publication detailing the population background characteristics and survival analysis outcomes (121). A previous analysis of a subset of the dataset was published in 2006 (139). Table 3-8 presents the AJCC stage cancer prevalence point estimates, derived from this dataset.

Table 3-8: Primary CRC model prevalence parameter estimates

Parameter	Estimates for Primary colon & rectal cancer			
	Point estimate	N patients	Probabilistic distribution	Data Source
Cancer Prevalence AJCC stage 1 (T1,T2, no nodes, no metastases)	0.19	n = 541*	Dirichlet	Clinical experts 2010 (121)
AJCC stage 2 (T3,T4, no nodes, no metastases)	0.34	n = 977*	Dirichlet	(121)
AJCC stage 3 (T, nodal spread, no metastases)	0.31	n = 891*	Dirichlet	(121)
AJCC stage 4 (T, nodal spread, metastases)	0.15	n = 429*	Dirichlet	(121)

* Number of patients in each AJCC stage, derived from the Scottish dataset (121)

The AJCC stage prevalence data was incorporated into the model under the assumption that AJCC stages 1 and 2 represent patients with no nodal involvement (1518, 53%), and AJCC 3 and 4 represent patients with nodal involvement (n = 1320, 47%). The prevalence and number of patients in each AJCC stage in the dataset is detailed in Table 3-8. Having merged the AJCC data to distinguish disease in terms of nodal involvement to synchronise with the diagnostic test outcomes, the decision tree then separates the data back into the individual AJCC stages in the final branches, using the prevalence point estimates for each stage, in order to assign treatment strategies for each AJCC stage. Therefore, patients who are inaccurately staged are done so according to that disease stage prevalence, i.e. patients who have nodal involvement (AJCC 3 and 4) but have false negative test results, are inaccurately staged as either AJCC1 or AJCC2 based on the prevalence of AJCC 1 and 2. AJCC2 is more prevalent than AJCC1, as such this assumption ensures that when false negatives inaccurately under-stage patients as AJCC 2 and AJCC1 (instead of AJCC3 and 4), a greater proportion of patients will be inaccurately staged as AJCC2 than inaccurately staged as AJCC1.   It is also more likely that an AJCC3 patient would be under-staged to AJCC2 than to AJCC1. In the case of false positives which lead to over-staging, as the prevalence of AJCC3 is greater than that of AJCC 4, a greater proportion of

inaccurate over-staging will be attributed to AJCC 3 than AJCC4 due to greater prevalence.

For the probabilistic analysis, Dirichlet distributions (4;75) were deemed to be appropriate due to the nature of the prevalence data. The point estimates were derived from a cohort of 2838 patients in the Scottish colorectal cancer dataset (121), within which the sample of patients was split into four categories representing the four AJCC stages. When dealing with multinomial data such as this, the Dirichlet distribution which is a multinomial version of the Beta distribution, is recommended (4). Applying the Dirichlet distribution interprets the prevalence parameters as sample sizes, allowing the sample size within each AJCC stage to vary, but in relation to each other so that the total sample size remains constant.

Diagnostic Test Accuracy

The systematic review undertaken by the research team was intended to yield data on diagnostic test accuracy (DTA) for the various imaging technologies, which would be pooled in a meta-analysis to inform the main parameters for the economic models. However, the systematic review found limited evidence and inadequacies and reporting bias in published papers for all stages of CRC disease.

For primary colorectal cancer, evidence was required regarding the diagnostic test accuracy of ceCT, MRI and PET/CT scanning for staging primary colorectal cancer.   Only two papers were identified which evaluated the DTA of PET/CT for the detection of primary colorectal cancer (109;110); however one of the papers (110) did not include a comparator and the other (109) compared PET/CT against contrast enhanced PET/CT (cePET/CT) which is not available in the UK. Both studies were conducted in Japan in cancer research centres, and therefore neither study included ceCT or MRI which are the conventional imaging technologies used for staging colorectal cancer in the UK. Neither study reported including a consecutive series of patients or a random sample and therefore the studies may be unrepresentative of the test in clinical practice. In both studies the assessors were blind to the clinical information

and results of other studies, however there was no mention of whether the individuals reviewing the scans were blinded.

Due to the lack of papers it was deemed inappropriate to undertake a meta-analysis in primary colorectal cancer. Therefore, the two papers were considered one by one, along with other literature identified through the economic search and considerable input from the clinical experts, in order to decide which data to incorporate in the economic models. One of the studies (110) only reported DTA data at a lesion level and was therefore not useful to the model; however, the other study by Tateishi et al. (109) (which compared PET/CT to cePET/CT) reported patient level data on the sensitivity and specificity of PET/CT for staging nodal involvement and provided confidence intervals. No distinction was made between colon and rectal cancer in either paper, and due to this and the lack of alternative information, the Tateishi paper PET/CT estimates were used in both the colon and rectal analyses.

Table 3-9 details the DTA parameter estimates for the ceCT, MRI and PET/CT imaging technologies used in the primary colon and rectal analyses.

Table 3-9: Primary colon & rectal cancer DTA parameter estimates

DTA Parameter	Primary Rectal Model			Primary Colon Model			Data Source
	Point estimate	se	Prob dist	Point estimate	se	Prob dist	
ceCT sensitivity	0.55	0.06	Beta	0.55	0.06	Beta	(140)
ceCT specificity	0.74	0.04	Beta	0.74	0.04	Beta	(140)
MRI sensitivity	0.66	0.06	Beta	-	-	-	(140)
MRI specificity	0.76	0.09	Beta	-	-	-	(140)
PET/CT sensitivity	0.85	0.08	Beta	0.85	0.08	Beta	(109)
PET/CT specificity	0.42	0.10	Beta	0.42	0.10	Beta	(109)

Note: se=standard error, Prob dist = probability distribution

The MRI scan is not used in the assessment of primary colon cancer and therefore the primary colon analysis incorporates only ceCT as the conventional imaging modality. As previously discussed, the DTA literature made few distinctions between colon and rectal cancer, and therefore due to this and the lack of alternative information, the PET/CT and ceCT estimates were used to represent DTA for both colon and rectal cancer.

Sensitivity and specificity are probability parameters and therefore as the data is binomial, Beta distributions were used for the probabilistic analysis. The Tateishi paper (109) reported the mean DTA value with the associated confidence intervals (CI) for sensitivity: 0.85 (95% CI 0.69, 0.936) and specificity: 0.42 (95% CI 0.23, 0.637). Therefore the Method of Moments was used to fit the Beta distributions (4), utilising the mean DTA estimates and the corresponding standard errors, which were calculated using the lower confidence limit[6]. An independent probability distribution was assumed for the sensitivity and specificity estimates.

DTA data for ceCT and MRI was taken from a study (140) that undertook a meta-analysis in primary CRC and reported DTA estimates with confidence intervals for these imaging modalities for staging nodal involvement. This was the only study which provided patient level DTA estimates for identifying lymph node involvement. Beta distributions were applied in the probabilistic analysis, and therefore the lower confidence interval was used to calculate a standard error for use in the Method of Moments for fitting the Beta distributions. PET/CT was not included in this meta-analysis, however, as the study detailed the sensitivity and specificity of ceCT and MRI specifically for nodal involvement, it is reasonable to enter these estimates into the primary models, to compare with the addition of PET/CT using the Tateishi (109) estimates specifically for staging nodal involvement.

In the primary colon cancer analysis, there is only one test (ceCT) in the conventional arm and therefore the DTA and corresponding standard error reported in Table 3-9 were used to represent this. The primary rectal analysis (which involves a ceCT scan and an MRI scan in the conventional arm) used the DTA estimates for MRI to represent the joint (ceCT + MRI) DTA, as overall it has superior test performance characteristics for lymph node involvement, i.e. both sensitivity and specificity of MRI is superior to ceCT (140). This approach of using superior test performance to represent joint imaging modalities has been used by others (108) and is also reasonable given the evidence identified

---

[6] Both the lower and upper confidence limits were used to determine the standard error, however the lower confidence limit generated a larger standard error than the upper limit, and therefore the standard error from the lower limit was used to represent the widest range of uncertainty.

in the systematic review which favoured MRI in the identification of nodal involvement (141;142). By adopting the superior tests' DTA to represent joint DTA, conflicting test outcomes are avoided and a confirmatory positive (or negative) strategy, as discussed in section 3.3.1, is not required in the conventional arm for rectal cancer. In the intervention arm, the DTA for PET/CT is added-on after the conventional test, and a confirmatory positive strategy was adopted whereby the model assumes that positive test outcomes override negative test outcomes. Therefore, negative results from the conventional test which are refuted by the PET/CT test are treated as positive. Results are only treated as negative when both the conventional and PET/CT tests result are negative.

Treatments

The systematic review and the non-systematic economics search identified some literature on therapeutic impact and patient management in primary colorectal cancer. This literature found that while PET/CT impacts in terms of more accurate staging of primary colorectal cancers, it had only a minor impact on changing patient management (112-114;143), and none of the studies reported the impact that a change in management had in terms of patient outcomes.

Optimal treatment combinations for the four AJCC stages were determined through consideration of the literature (112-118) and in consultation with clinical experts, and are detailed earlier in section 3.3.2. The proportions of patients receiving each treatment within each AJCC stage were informed by publications reporting treatment and therapeutic impacts for primary rectal and primary colon cancer (112;113;118;134;144) however the point estimates were assigned based on judgement from the clinical experts on the research team. Table 3-10 details the point estimates for the probability of receiving the treatment options within each AJCC stage, for rectal and colon cancer.

Table 3-10: Primary colon & rectal cancer Treatment parameter estimates

Treatment Parameters	Primary Rectal Model			Primary Colon Model		
	Point estimate	N patients	Prob dist	Point estimate	N patients	Prob dist
AJCC1						
Surgery	1.00		-	1.00		-
AJCC2						
Surgery alone	0.38	n=977	Dirichlet	0.80	n=977	Dirichlet
LCR plus surgery	0.46		Dirichlet	-		-
Surgery & adjuvant chemo	0.15		Dirichlet	0.20		Dirichlet
AJCC 3						
Surgery alone	0.34	n=891	Dirichlet	0.34	n=891	Dirichlet
LCR & surgery	0.37		Dirichlet	-	-	-
Surgery & adjuvant chemo	0.29		Dirichlet	0.66		Dirichlet
AJCC 4						
Surgery alone	0.08	n=429	Dirichlet	0.09	n=429	Dirichlet
LCR & surgery	0.11		Dirichlet	-	-	-
Surgery & metastatic surgery	0.16		Dirichlet	0.19		Dirichlet
Surgery & palliative care	0.53		Dirichlet	0.63		Dirichlet
Palliative care alone	0.13		Dirichlet	0.09		Dirichlet

Data sources: Clinician advice for point estimates supported by publications (113); Dirichlet distributions & N patients derived from Scottish CRC dataset (121) Prob dist = probability distribution, LCR = long course chemo-radiotherapy, chemo = chemotherapy

For patients diagnosed with AJCC1 colon and rectal cancer, the only treatment option is local surgery, with a probability of 1. There is no uncertainty in this parameter, everyone diagnosed with AJCC 1 will receive this treatment, and therefore this was held constant in the probabilistic analysis. With regards to stages AJCC 2, 3 and 4, the various treatment options were assigned as probabilities, summing to 1 within each stage. These are multinomial parameters and therefore it is appropriate to consider the Dirichlet distribution to represent uncertainty in the probabilistic analysis. The sample of patients from the Scottish CRC dataset (121) was utilised to represent a cohort of patients for each AJCC stage, so that a series of Dirichlet distributions within each AJCC stage could be used to incorporate uncertainty.

Survival

The Scottish CRC network dataset (121) (2,328 CRC patients average age 61 yrs) detailed the five year overall survival of patients for each AJCC stage. This data was used to determine annual mortality rates under the assumption of an exponential survivor function, for use in the survival analysis, as detailed in section 3.3.2. The five year survival data was derived from both rectal and colon patients in the colorectal cancer dataset, and therefore the same survival estimates were used for both the colon and rectal analyses. Table 3-11 details the survival parameters for each of the AJCC stages.

Table 3-11: Primary colorectal cancer Survival parameter estimates

Survival Parameters	Estimates for Primary colon & rectal cancer			
	Point estimate	*Std error*	*Prob dist*	*Data Source*
5 year overall survival AJCC 1	0.95	0.01	Beta	Data provided by clinical experts Feb 2010
5 year overall survival AJCC 2	0.86	0.01	Beta	
5 year overall survival AJCC 3	0.69	0.02	Beta	
5 year overall survival AJCC 4	0.13	0.02	Beta	
Reduction 5yr overall survival For fail to receive AJCC 3 chemo	0.25	0.05	Beta	Author Assumption

Data source: Scottish CRC overall survival data provided by clinical experts Feb 2010. Published reference (121)

The data was binomial, in the form of the probability of survival 5 years post-diagnosis, and therefore Beta distributions were deemed appropriate to represent uncertainty in the probabilistic analysis. The dataset provided the sample size for each AJCC stage and the number of events (deaths) in each, and therefore the Beta distribution was fitted in the standard way whereby the number of events represents alpha and the sample size minus the events is equivalent to beta. Given the sample size and the number of events, the standard error was calculated using the moments of the Beta distribution (4). The point estimates and standard errors are presented in Table 3-11.

Publications (126-128) indicate that in comparison to conventional imaging modalities PET/CT scanning has no impact on overall survival; however, consultation with clinical experts highlighted that patients with AJCC3 stage cancer (nodal involvement but no metastases) who fail to receive adjuvant

chemotherapy due to inaccurate staging, may suffer a reduction in overall survival. This was incorporated into the model for AJCC3 patients who were inaccurately under-staged as AJCC1 or 2, as a 25% reduction in overall survival. The 25% reduction in five year overall survival was an author assumption based on advice from the clinical experts. A corresponding standard error of 0.05 was assumed, which was greater than the standard errors for the survival data, in order to represent the greater uncertainty in this parameter estimate.

Quality of life / Utility

Average utility estimates were used to represent quality of life for each of the four AJCC stages, while adjustments were made for inaccurate staging, which resulted in failing to receive treatment (disutility for 1 year), or receiving unnecessary treatment (alternative utility for duration of treatment). Table 3-12 details the utility and disutility values used for the rectal and colon analyses.

Ramsey et al. (122) report utility values for different stages of CRC at various time points' post-diagnosis. The mean 5 year utility values and corresponding standard errors were used to represent quality of life for each AJCC stage. Patients who were correctly diagnosed in the model received the average utility for their state for the first 5 years, followed by age adjusted population utility weights (124).

Patients incorrectly diagnosed received their true disease stage utility, but with a disutility relating to the inappropriate treatment they received for a specified duration as detailed previously in Table 3-5. It was assumed that patients who were inaccurately staged and failed to receive either long course chemo-radiotherapy or adjuvant chemotherapy post surgery received a disutility of 0.20 for a 6 month duration, while patients who were inaccurately diagnosed and failed to receive metastatic surgery or palliative care were assumed to receive a disutility of 0.3 for a year, reflecting the large impact on quality of life for delayed treatment. The disutility values and standard errors were assigned at the authors discretion, based on advice from clinicians and related utility information provided in various papers (122;132;137).

Table 3-12: Primary colon & rectal cancer Utility parameter estimates

Utility Parameters	Primary Rectal Cancer			Primary Colon Cancer			Data Source
	Point est	se	Prob dist	Point est	se	Prob dist	
AJCC1							
Mean 5 year utility	0.84	0.17	Gamma on disutility	0.84	0.17	Gamma on disutility	(122)
AJCC2							
Mean 5 year utility	0.86	0.14	Gamma on disutility	0.86	0.14	Gamma on disutility	(122)
AJCC 3							
Mean 5 year utility	0.85	0.14	Gamma on disutility	0.85	0.14	Gamma on disutility	(122)
AJCC4							
Mean 5 year utility	0.84	0.12	Gamma on disutility	0.84	0.12	Gamma on disutility	(122)
Fail to receive							
LCR or adjuvant chemotherapy	0.20	0.08	Gamma	0.20	0.08	Gamma	AA based on (132)
metastatic surgery / palliative care	0.30	0.08	Gamma	0.30	0.08	Gamma	AA based on (132)
Receive unnecessary							
LCR	0.74	0.14	Gamma on disutility	-	-	-	(122;132)
adjuvant chemo	0.80	0.14	Gamma on disutility	0.80	0.14	Gamma on disutility	(122)
metastatic surgery	0.74	0.21	Gamma on disutility	0.74	0.21	Gamma on disutility	(137)

LCR = long course chemo-radiotherapy, chemo = chemotherapy, AA = authors assumption, se=standard error, est=estimate, prob=probability

Patients who received unnecessary long course chemo-radiotherapy or unnecessary adjuvant chemotherapy received an alternative, lower utility value (122;132) for six months to reflect the impact on their quality of life during their unnecessary treatment. Patients who received unnecessary metastatic surgery or palliative care had a lower, alternative utility value (137) for one year.

For probabilistic analysis of utilities it is common to assume that zero is the worst possible health state (no negatives) and fit Beta distributions, however, due to the nature of the utility values for cancer, where quality of life can be very low (and even negative) during treatment or in advanced stages of cancer, the Gamma distribution was considered to be more appropriate. The transformation method was used whereby the Gamma distribution was set on

disutility (Disutility= 1-Utility) (4), so that the distribution was constrained on the interval zero to infinity, allowing for a right skew of the data which represents very low and even negative utility values. The mean utility values and corresponding standard errors were reported in the literature, and therefore the method of moments for gamma was used to calculate the shape (alpha) and scale (beta) parameters to fit the gamma distributions.

Costs

The costs for the economic model are attributed to the cost of the alternative imaging devices as a cost per scan and the cost of the various treatments for each AJCC stage. NHS reference costs data was used (101;145) along with various other data sources for the AJCC stage treatment options (97;100;102;119). The various cost items are detailed in Table 3-13.

The cost of the imaging devices was incorporated as a cost per scan, representing staff time, and use of the imaging machinery. Cost details regarding ceCT and MRI scans were available in NHS reference costs, however, no details were provided for the cost of PET/CT scanning in either the Department of Health or the Scottish ISD reference costs. Various literature reports the cost of a PET/CT scan in the UK as between £750 and £1000 per scan (120;146-148). It is also widely reported that PET/CT scans generally have a duration of 20-40 minutes on equipment costing 2-3 times that of CT scanners, which can perform scans on a patient every 5-10mins (99) therefore, assigning a cost of £800 per PET/CT scan seemed appropriate. A standard error for this baseline cost was derived using the upper and lower price range reported for an PET/CT scan (148).

Table 3-13: Primary CRC model costs

| Item | Primary Rectal Cancer | | | Primary Colon Cancer | | | |
	Unit Cost (£)	*se*	*Prob dist*	*Unit cost (£)*	*se*	*Prob dist*	*Data Source*
Imaging Devices							
ce CT scan (chest abdomen & pelvis)	143	22	Normal	143	22	Normal	(101)
MRI scan (pelvis)	179	24	Normal	-	-		(101)
PET/CT scan	800	100	Normal	800	100	Normal	(101)
Treatments							
Primary surgery (rectal excision with lymphadenectomy):	5637	677	Normal	-	-	-	(51;145)
Primary surgery (colonic resection with lymphadenectomy)	-	-	-	5893	746	Normal	(51;145)
LCR	13721	-	-	-	-	-	(97;100;119;120;145)
Adjuvant chemotherapy	11532	-	-	11532	-	-	(100;120;145)
Palliative Care	2468	494	Normal	2468	494	Normal	(149)
Metastatic surgery	9134	1827	Normal	9134	1827	Normal	(145)

Prob dist=probabilistic distribution, se=standard error, LCR=long course chemo-radiotherapy

The cost of primary rectal surgery (rectal excision with lymphadenectomy) involves the cost of a distal colon procedure (including surgical consultation, theatre time and staff costs), an average hospital inpatient stay of 6 days and a CRC surgery consultant follow-up. Primary colon surgery (colonic resection with lymphadenectomy) involves the cost of a proximal procedure (including surgical consultation, theatre time and staff costs), an average hospital inpatient stay of 6 days and a CRC surgery consultant follow-up.

The cost of long course chemo-radiotherapy treatment includes the radiotherapy drugs (45Gy in 25 fractions given over five weeks),12 weeks of intravenous 5-fluorouracil (5FU) for chemotherapy and the associated administration and hospital stay costs (97;120). The cost of adjuvant chemotherapy treatment incorporated a six month course of intravenous 5FU plus Oxilaplatin for 24 weeks (100;120) and the associated administration and hospital costs. The cost of metastatic surgery was represented by the Information Services Division reference cost of surgical specialities in medical

oncology (145), which includes the cost of surgery including theatre time, surgical consultation and follow-up, and an average inpatient stay in hospital of 10 days. Resource use and costs for palliative care were taken from a study which assessed the cost to the NHS of palliative care in colorectal cancer (149). The costs of palliative care were reported at price year 2000/01, and therefore the HCHS pay and price index was used to adjust this to price year 2009 (102).

The unit costs, standard errors and distributions used in the probabilistic analysis are reported in Table 3-13. In the probabilistic analysis normal distributions were considered to be appropriate for representing the unit cost parameters, as the point estimates were derived from UK and Scottish reference costs sources, and as these are very large data sources they can be considered to have sufficient sample sizes for the central limit theorem to apply. The unit costs for long course chemo-radiotherapy and chemotherapy remained constant in the probabilistic analysis.

The average cost per AJCC stage was calculated using the proportion of patients receiving each treatment option within each AJCC stage. In the model, if a patient was staged accurately they would receive their optimal treatment option and be assigned the average cost of treatment for that stage. The model also incorporates the extra costs incurred through inaccurate staging. If a patient is inaccurately diagnosed they incur the cost of the misdiagnosed treatment, followed by the discounted cost of treatment for their true stage the following year (i.e. it is assumed that the true disease stage will be identified within a year). Costs were discounted at 3.5% (24).

Scenario analysis: cePET/CT as a lone technology

The systematic literature review of PET/CT found suggestions within the literature that as PET/CT technology improves with the development and introduction of contrast enhanced PET/CT (cePET/CT) scanners (109), it may be possible to use this higher quality technology as an alternative to CT or ceCT in primary CRC rather than using PET/CT as an add-on imaging device.

While the scope of the research undertaken was focussed on PET/CT as an add-on device, a scenario analysis was undertaken for the primary colorectal model in which PET/CT was considered to be a replacement for conventional ceCT, rather than as an add-on device. The Tateishi paper (109) which provided DTA evidence for PET/CT also provided patient level DTA estimates for cePET/CT in nodal staging (with equivalent sensitivity to PET/CT but improved specificity). These contrast enhanced DTA estimates and confidence intervals were used in the scenario analysis to portray the future potential of improved PET/CT imaging. For the primary rectal scenario, the conventional strategy (ceCT followed by MRI) was compared against a cePET/CT replacement strategy (cePET/CT followed by MRI); and for the primary colon scenario, the conventional ceCT was compared with cePET/CT alone. All model parameters remain as detailed in Tables 3-8 to 3-13 with the exception of the DTA estimates and the cost of cePET/CT. The cePET/CT DTA estimates and 95% confidence intervals for sensitivity: 0.85 (0.69, 0.93) and specificity: 0.68 (0.46, 0.85) were used (109) and a cost for the cePET/CT scan was incorporated, assuming an increase of 20% to the PET/CT scan cost to reflect the cost of this more expensive technology.

3.4.2 Model 2: Recurrent disease

The model for recurrent colorectal cancer was populated with parameters representing the prevalence of recurrent colorectal cancer, the diagnostic test accuracy of the conventional and intervention imaging technologies for diagnosing recurrent rectal and colon cancer, and the various treatment options, survival estimates and quality of life. Each of these is discussed in turn, followed by details of the associated costs. The parameter tables distinguish the estimates for rectal and colon cancer separately where the details are different, and jointly as colorectal cancer where the same estimates were used for both analyses.

Disease Prevalence

Some publications indicate that local recurrence in rectal cancer is more common than local recurrence in colon cancer, however, data for the UK indicate only a very small difference in local recurrence for rectal and colon cancers (120). Therefore, the recurrent model assumed the same probability of recurrence for both the rectal and colon analyses.

The literature identified in the economics search and the systematic review was used to provide disease prevalence evidence for the recurrent model. Disease prevalence data on recurrence in CRC (150) determined a 30% probability of local recurrence and a 40% probability of metastatic recurrence for patients previously treated for primary colorectal cancer. These point estimates were used and therefore the probability of no recurrence is 30%, as detailed in Table 3-14.

Table 3-14: Recurrent CRC model prevalence parameter estimates

| | Recurrent Rectal and Colon Cancer | | | |
Prevalence Parameters	Point estimate	Standard error	Probabilistic distribution	Data source
Local recurrence	0.3	n=723	Dirichlet	(150)
Metastatic recurrence	0.4	n=964	Dirichlet	(150)
No recurrence	0.3	n=723	Dirichlet	
Recurrence curable	0.3	0.1	Beta	(135)
Recurrence non-curable	0.7		1- above	

* These are not standard errors; they are the mean number of patients in each sample for the Dirichlet distributions

It was assumed that a cohort of patients who were diagnosed as AJCC1, AJCC2 or AJCC3 for primary CRC would be susceptible to recurrence. The Scottish network CRC dataset (121) of patients (n=2838) was used to determine a sample of patients diagnosed with stages AJCC 1, 2 & 3 (n=2409). This sample was then used to represent a cohort susceptible to recurrence, and the point estimate probabilities of recurrence (150) were applied; so that 30% of the n=2409 patients would have local recurrence, 40% would have metastatic recurrence, and the remaining 30% no recurrence. Uncertainty in these prevalence point estimates was incorporated by applying a Dirichlet distribution to the cohort, so that uncertainty between the three prevalence categories: no recurrence, local recurrence, or metastatic recurrence was accounted for. As recurrence falls into three categories in this model, it is appropriate to use the Dirichlet distribution, so that the prevalence parameters are interpreted as sample sizes, allowing the sample size within each recurrence category to vary, but in relation to each other so that the total sample size remains constant.

Once the prevalence of recurrence had been established, the model splits recurrence in to curable and non-curable recurrence in order to incorporate appropriate treatment options. This model structure is similar to the structure used in two other economic evaluations which assessed the value of using PET in the identification of recurrent colorectal cancer (129;130). These two models also incorporated patient management and quality of life impacts to their models by including a probability of curable and non-curable recurrence

to the recurrent population (129;130). Table 3-14 details the point estimates used to identify curable and non-curable recurrence (135). The probabilistic analysis accounted for the uncertainty in the parameter estimate by fitting a Beta distribution. As there are two categories, summing to a probability of 1, Beta distributions were fitted, using the method of moments to find alpha and beta, given the reported mean value and confidence intervals (135).

Diagnostic Test Accuracy

The systematic review undertaken by the research team was intended to yield data on diagnostic test accuracy (DTA) for the various imaging technologies, which would be pooled in a meta-analysis to inform the main parameters for the economic models. The meta-analysis was undertaken by the systematic review team in order to elicit pooled diagnostic test accuracy estimates of PET/CT for recurrent colorectal cancer (151). The quality of evidence from literature identified in the systematic review was poor and subject to reporting bias and only five studies were considered to be eligible for meta-analysis (152-156). All of the studies were retrospective, patient populations were not well described in terms of disease classification or primary diagnosis, and all included only a small number of patients. Figure 3-14 details the forest plot presenting the accuracy data of PET/CT in identifying recurrent disease in 276 patients from the five studies.

	Sens	STUDY	Spec	
	6/8	Sarikaya 2007	0/4	
	14/15	Bellomi 2007	51/52	
	40/45	Votrubova 2006	36/39	
	23/24	Even-Sapir 2004	34/38	
	17/18	Schmidt 2009	6/6	
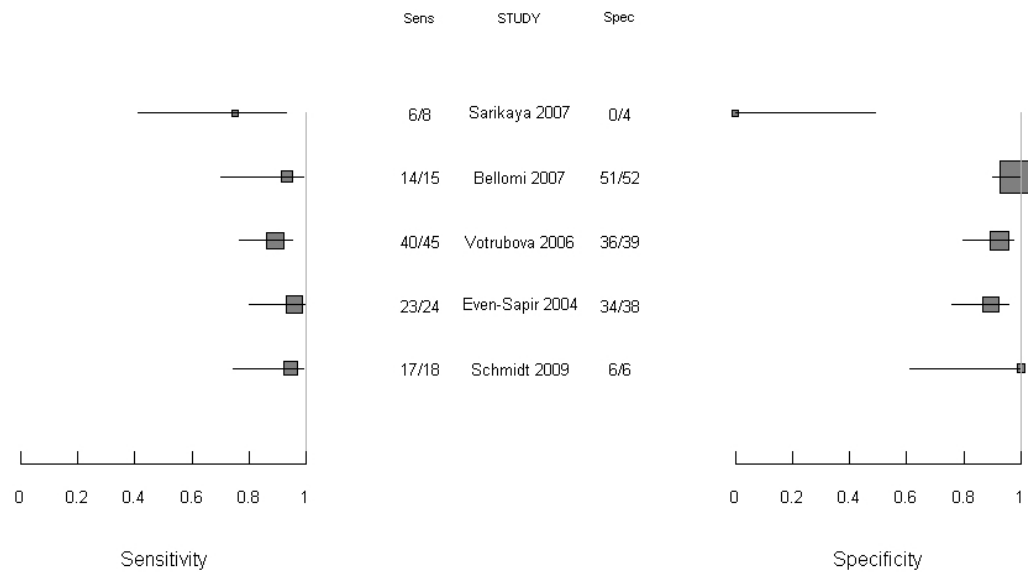
Figure 3-14: Accuracy data of PET/CT in the detection of recurrent CRC

There was little evidence of heterogeneity in the sensitivity estimates, and therefore a fixed effects meta-analysis was undertaken, resulting in an overall estimate of sensitivity of 0.91 (95% CI 0.87 to 0.95). There was some evidence of heterogeneity in the specificity estimates from these papers and therefore the random effects Bivariate/HSROC method was intended to be used, but the data were not adequate to fit the Bivariate/HSROC model and therefore two separate univariate meta-analyses were used. The overall estimate of specificity is 0.91 (95% CI 0.85 to 0.95).

These pooled estimates for PET/CT may not be an accurate reflection of diagnostic test accuracy due to the inadequacies and reporting bias in the identified papers. The pooled estimates also give tight confidence intervals which do not fully represent the wide uncertainty in the mean estimates. Therefore, it was decided that the pooled DTA estimates were not adequate for use in the baseline economic model. The papers identified by the systematic review were considered one by one, along with other papers identified through the economic search and considerable input from the clinical experts, in order to decide which data to incorporate in the economic

models and to find reasonable estimates of DTA for the economic models with wide uncertainty intervals.

Three papers provided DTA evidence of PET/CT as an add-on device for diagnosis of recurrent colorectal cancer. One paper (156) compared PET/CT with whole body MRI, but reporting bias was evident in this study. In addition, the DTA for whole body MRI was inappropriate for the model which incorporates pelvic MRI rather than whole body MRI. A second paper (157) provided evidence for ceCT in comparison to PET/CT, but the point estimates assigned appeared to be biased in favour of PET/CT (reporting a sensitivity of zero for ceCT, but with a confidence interval range up to 0.65). The final paper (158) provided DTA evidence for ceCT in comparison to PET/CT; however they did not report any confidence intervals or other measures of uncertainty. The point estimates from this paper (158) were deemed to be the best reflection of mean DTA and were therefore used in the model along with the wide confidence intervals from Ramos et al. (157) to ensure a suitably wide range to reflect the considerable uncertainty surrounding the mean DTA estimates. The pooled meta-analysis DTA estimates had more restrictive confidence limits, and were therefore deemed inappropriate to accurately reflect uncertainty in the economic models. There were no reliable estimates of pelvic MRI diagnostic test accuracy for recurrent rectal cancer reported, as such an estimate was taken from the DTA of MRI used in another (134) economic evaluation. DTA estimates, their standard errors and the distributions used in the probabilistic model are detailed in Table 3-15.

Table 3-15: Recurrent CRC model Diagnostic Test Accuracy estimates

| DTA Parameters | Recurrent Rectal Cancer | | | Recurrent Colon Cancer | | | Data source |
	Point estimate	se	Prob dist	Point estimate	se	Prob dist	
ceCT sensitivity	0.53	0.06	Beta	0.53	0.06	Beta	(157;158)
ceCT specificity	0.98	0.015	Beta	0.98	0.015	Beta	(157;158)
MRI sensitivity	0.85	0.03	Beta	-	-	-	(134)
MRI specificity	0.95	0.08	Beta	-	-	-	(134)
PET/CT sensitivity	0.93	0.069	Beta	0.93	0.069	Beta	(157;158)
PET/CT specificity	0.98	0.025	Beta	0.98	0.025	Beta	(157;158)

Prob dist=probability distribution, se=standard error

As the diagnostic test accuracy parameters are binomial in nature, Beta distributions were fit using the Method of Moments (4) with the mean estimates and standard errors derived from the reported confidence intervals.

With regards to dealing with joint test outcomes (for the intervention arm of the model and in the conventional arm for the rectal cancer analysis), where there was superiority in one test, this was taken to represent combined DTA (108), and where there was no test superiority, a confirmatory positive approach was adopted.

Treatments

The optimal treatment combinations for patients with curable and incurable recurrent colorectal cancer were determined through consideration of the literature, and in consultation with clinical experts and are reported earlier in Figure 3.7.

The proportions of patients receiving each treatment for curable recurrence were assigned based on consultations with clinical experts and the publications reporting treatment and therapeutic impacts for recurrent colorectal cancer (117;118;131). As patients with recurrent colon cancer do not receive long course chemo-radiotherapy, the treatments and probabilities of receiving the treatments vary from the rectal analysis. The proportions of patients receiving each treatment for incurable recurrence were informed by literature and previous economic models for recurrent colorectal cancer (129;159). No recurrence had only one treatment option and was therefore assigned a probability of 1. Table 3-16 details the parameter estimates for the recurrent colon and rectal cancer treatment options.

Table 3-16: Recurrent CRC model Optimal Treatment estimates

Treatment Parameters	Recurrent Rectal Cancer			Recurrent Colon Cancer		
	Point estimate	*se*	*Prob dist*	*Point estimate*	*se*	*Prob dist*
Recurrence Curable						
Surgery alone	0.05		Dirichlet	0.05		Dirichlet
Surgery & adjuvant chemo	0.1		Dirichlet	0.35		Dirichlet
LCR then surgery	0.25		Dirichlet	-		-
Surgery (local) & metastatic surgery	0.1		Dirichlet	0.1		Dirichlet
Surgery (local), adjuvant chemo & metastatic surgery	0.15		Dirichlet	0.50		Dirichlet
LCR, surgery & metastatic surgery	0.35		Dirichlet	-	-	-
Recurrence Incurable						
Metastatic surgery & palliative care	0.2	0.04	Beta	0.2	0.04	Beta
Palliative care alone	0.8		1- above	0.8		1- above
No recurrence						
Wait & watch	1	-	-	1	-	-

Data Sources: Recurrence Curable clinician advice & various refs (112;113;118), Recurrence Incurable (129), No recurrence: author assumption. Chemo=chemotherapy, LCR=long course chemo-radiotherapy, se=standard error, Prob dist=probability distribution

With regards to uncertainty in the treatment options for curable recurrence, Dirichlet distributions were considered to be appropriate given the categorical nature of the parameters. Therefore, the Scottish network CRC dataset (121) was used to derive a cohort of patients with AJCC stages 1-3 (n=2409) who would be susceptible to colorectal recurrence. The recurrent colorectal cancer sample was then derived using the probabilities of local and metastatic recurrence from Table 3.14 (n=1686), and a subset of this population was deemed to have curable recurrence (n=505), again using the prevalence probabilities assigned in the model (detailed in Table 3-14). A Dirichlet distribution was applied to this subset, in order to capture the uncertainty surrounding the treatment allocations.

With regards to incurable recurrence, there were only two treatment options, making this a binomial parameter and therefore a Beta distribution was fitted using the mean and standard error reported in the literature (129).

Survival

Five year survival estimates for recurrent and non recurrent colorectal cancer patients are reported by the American Cancer Society (133). This data was used to determine annual mortality rates under the assumption of an exponential survivor function, for use in the survival analysis, as detailed in section 3.3.3. As the survival data represents colorectal cancer, not distinguishing between colon and rectal patients, the same survival estimates were used for both the colon and rectal analyses. Table 3-17 details the survival parameters for the recurrent colorectal cancer model.

Table 3-17: Recurrent CRC model Survival estimates

| Survival Parameters | Recurrent Rectal and Colon | | | |
	Point estimate	*Std error*	*Probability distribution*	*Data source*
5 year overall Survival				
No recurrence	0.85	0.01	Beta	(133)
Recurrence curable	0.3	0.02	Beta	(133)
Recurrence non-curable	0.1	0.01	Beta	(133)
Recurrence curable (fail to treat)	0.2			Author Assumption

Std = standard

The data was binomial, in the form of the probability of survival 5 years post-diagnosis, and therefore Beta distributions were deemed appropriate to represent uncertainty in the probabilistic analysis. The Scottish network CRC dataset (121) was used to represent a cohort of AJCC1-3 patients who would be susceptible to colorectal recurrence (n=2409). The prevalence estimates (detailed in Table 3-14) were then applied to derive a population for no recurrence (n=723) and recurrence (n=1686), within which there would be curable patients (n=506) and incurable patients (n=1180). These populations were taken to represent sample sizes for the three recurrence categories, and the number of events (deaths) within each sample was determined from the 5 years survival estimates (133). Beta distributions were then fit to this data employing the Method of Moments (4) and using the number of deaths within

each recurrence category to represent the alpha parameter, and the sample size minus the events to represent the Beta parameter.

Patients who had curable recurrence but were inaccurately diagnosed as no recurrence (false negatives) and failed to receive treatment in the first year would have a negative impact on their mortality, as detailed in section 3.3.3. This was based on clinical advice, and as there was no survival estimates for such a scenario in the available literature, it was assumed that five year survival for these patients would be midway between the curable and non-curable survival estimates (0.2), as detailed in Table 3-17.

Quality of life / Utility

Average utility estimates were used to represent quality of life in the no recurrence, curable recurrence and incurable recurrence groups. Adjustments were made for inaccurate diagnoses, which resulted in failing to receive treatment (disutility for that year to account for the negative impact on the patients' quality of life), or receiving unnecessary curative or non-curative treatments (an alternative, lower utility for 1 year). Table 3-18 details the utility and disutility values which were used for the rectal and colon analyses.

Table 3-18: Recurrent CRC Utility parameter estimates

Utility Parameters	**Recurrent Rectal and Colon**			
	Point estimate	*Standard error*	*Probability distribution*	*Data source*
No Recurrence mean 5 year utility	0.91	0.11	Gamma (disutility)	(122)
Curable Recurrence mean 5 year utility	0.84	0.12	Gamma (disutility)	(122)
Incurable Recurrence mean 5 year utility	0.52	0.08	Gamma (disutility)	(132)
Fail to receive				
Curable treatment	0.3	0.08	Gamma	AA based on (132)
Incurable treatment	0.2	0.08	Gamma	AA based on (132)
Receive unnecessary				
Curable treatment	0.74	0.14	Gamma (disutility)	AA based on (132)
Incurable treatment	0.61	0.2	Gamma (disutility)	(137)

Utility estimates and standard errors reported in Ramsey et al. (122) were used to represent mean quality of life for patients with no recurrence and curable recurrence, for five years post-diagnosis. A utility estimate reported in Tengs & Wallace (132) for colorectal patients who receive best palliative/supportive care was taken to represent mean quality of life for patients with incurable recurrence, for five years post-diagnosis.   No standard error or confidence intervals were given to support this point estimate, and therefore the authors assumed a standard error of 0.08. Following the first 5 years in the survival analysis, UK age adjusted population utility weights were applied (124).

Patients incorrectly diagnosed received their true disease stage utility, but with a disutility relating to the inappropriate treatment they received for a specified duration as detailed previously in Table 3-6. It was assumed that patients who were inaccurately diagnosed and failed to receive curable treatments (such as surgery, adjuvant chemotherapy, long course chemo-radiotherapy, metastatic surgery and palliative care) were assumed to receive a disutility of 0.2 for a year, and patients who failed to receive incurable treatments (metastatic surgery and palliative care) received a disutility of 0.3 for a year, reflecting the large impact on quality of life for delayed treatment. The disutility values and standard errors were assigned at the authors' discretion, based on advice from clinicians and related utility information provided in various papers (132). The authors assumed a corresponding standard error of 0.08.

Patients who received unnecessary curative treatment (local surgery, adjuvant chemotherapy, long course chemo-radiotherapy, metastatic surgery) received an alternative, lower utility value of 0.74 (132) for one year to reflect the impact on their quality of life during their unnecessary treatment. No standard error or confidence intervals were given to support this point estimate, and therefore the authors assumed a standard error of 0.14 to represent uncertainty in the point estimate. Patients who received unnecessary incurable treatments were assigned an even lower alternative utility value (0.61 with a corresponding standard error 0.2 (137) for one year, to reflect the considerable impact on quality of life of receiving unnecessary treatment for misdiagnosed incurable disease.

Similar to the approach adopted for the primary CRC model, Gamma distributions were considered to be appropriate for the probabilistic analysis of utilities. The transformation method was used whereby the Gamma distribution was set on disutility (Disutility= 1-Utility) (4), so that the distribution was constrained on the interval zero to infinity, allowing for a right skew of the data which represents very low and even negative utility values. The mean utility values and corresponding standard errors were reported in the literature (or derived based on author assumption where appropriate), and therefore the Method of Moments for gamma was used to calculate the shape (alpha) and scale (beta) parameters to fit the gamma distributions.

Costs

The costs for the recurrent model are attributed to the alternative imaging devices and the treatment options for diagnoses of no recurrence, recurrence curable and recurrence incurable, as detailed in Table 3-19.

Table 3-19: Recurrent CRC model costs

Item	Rectal Cancer			Colon Cancer			Data Source
	Unit cost £	se	Prob dist	Unit cost £	se	Prob dist	
Imaging Devices							
ce CT scan (chest abdomen & pelvis)	143	22	Normal	143	22	Normal	(101)
MRI scan (pelvis)	179	24	Normal	-	-		(101)
PET/CT scan	800	100	Normal	800	100	Normal	(148)
Treatments							
Local surgery (rectal excision with lymphadenectomy)	5637	677	Normal	-	-	-	(101;145)
Local surgery (colonic resection with lymphadenectomy)	-	-	-	5893	746	Normal	(101;145)
LCR	13721	-	-	-	-	-	(97;100;119;120; 145)
Adjuvant chemotherapy	11532	-	-	11532	-	-	(100;120;145)
Palliative care	2468	494	Normal	2468	494	Normal	(149)
Metastatic surgery	9134	1827	Normal	9134	1827	Normal	(145)

Prob dist=probabilistic distribution, se=standard error, LCR=long course chemo-radiotherapy

Table 3-19 details the various cost items, unit costs and standard errors. NHS reference costs data was used (101;145) along with other data sources for the various treatment options (97;100;102;119;120).

The costs of imaging devices are the same as those used in the primary model. The treatment option combinations for the recurrent rectal model are different to those in the primary model; however the cost of the component treatments were assigned in the same way. For example, the cost of local recurrent surgery involves the cost of the procedure (including surgical consultation, theatre time and staff costs), an average hospital inpatient stay of 6 days and a CRC surgery consultant follow-up. The cost of long course chemo-radiotherapy treatment, adjuvant chemotherapy treatment, metastatic surgery and palliative care were also determined by the same means as in the primary model.

In the probabilistic analysis normal distributions were considered to be appropriate for representing the unit cost parameters, as in the primary CRC model. The point estimates were derived from UK and Scottish reference costs sources, and as these are very large data sources they can be considered to have sufficient sample sizes for the central limit theorem to apply.

The expected costs in the no recurrence, recurrence curable and recurrence incurable groups were calculated using the proportion of patients receiving each treatment option within each different group. In the model, if a patient was diagnosed accurately they would receive their optimal treatment option and the associated costs of that treatment. If a patient is inaccurately diagnosed they incur the cost of the diagnosed group treatment, followed by the discounted cost of treatment for their true diagnosis the following year (i.e. it is assumed that the true diagnosis would be identified within a year). Costs were discounted at 3.5% (24).

3.4.3 Model 3: Metastatic disease

The model for metastatic colorectal cancer was populated with parameters representing the prevalence of metastatic colorectal cancer, the diagnostic test accuracy of the conventional and intervention imaging technologies for diagnosing metastases, and the various treatment options, survival estimates and quality of life. Each of these is discussed in turn, followed by details of the associated costs.

Disease Prevalence

The literature identified in the economics search and the systematic review was used to provide disease prevalence evidence for the metastatic model. Estimates provided by Saunders et al (150) were used for the prevalence point estimates of metastatic recurrence for patients previously treated for primary colorectal cancer. The point estimates and probabilistic distributions are detailed in Table 3-20.

Table 3-20: Metastatic CRC model Prevalence parameters

Prevalence Parameters	Point estimate	Standard error	Probabilistic distribution	Data Source
No metastases	0.6	n=1445*	Dirichlet	(150)
Metastases	0.4	n=964*	Dirichlet	(150)
Metastases at one site	0.3	0.1	Beta	(135)
Extra metastases	0.7		1 - above	

* These are not standard errors; they are mean number of patients assumed in each sample for Dirichlet distributions

It was assumed that a cohort of patients who were diagnosed as AJCC1, AJCC2 or AJCC3 for primary CRC would be susceptible to metastatic recurrence. Using the Scottish network CRC dataset (121) to represent this cohort (n=2409) and assigning the probability of metastatic recurrence (150), a cohort of patients with no metastases (n=1445) and metastases (n=964) was generated. In the probabilistic analysis, uncertainty was incorporated by applying a

Dirichlet distribution to the cohort, to represent the uncertainty in the mean sample sizes for no metastases and metastases which in turn reflects uncertainty in the prevalence point estimates. As there are two categories the point estimates are binomial in nature, and therefore a Beta distribution would typically be used to represent uncertainty, however, there was no standard error and as Dirichlet distribution had been used to represent uncertainty in the prevalence parameters for the primary and recurrent CRC models, a consistent approach was adopted.

Once the prevalence of metastases had been established, the model splits metastases in to metastases at one site and extra metastases (at more than one site), in order to incorporate appropriate treatment options. This model structure is similar to that used by previous economic evaluations assessing the cost-effectiveness of using add-on PET/CT in the identification of metastatic disease (134;135). Previous models have attempted to incorporate patient management and quality of life impacts by distinguishing between resectable and unresectable metastases (135) or by distinguishing between hepatic metastases and extra metastases (134). Our evaluation distinguished between metastases at one site and at multiple sites (extra metastases), assigning a probability for each in the overall metastatic recurrence population (135). In this way the model could distinguish between metastatic and extra metastatic disease, even though the DTA estimates referred only to identification of metastases.

The probabilistic analysis accounted for uncertainty in these parameter estimates by fitting a Beta distribution. As there are two categories, summing to a probability of 1, Beta distributions were fitted, using the method of moments to find alpha and beta, given the reported mean value and confidence intervals (135).

Diagnostic Test Accuracy

The systematic review undertaken by the research team was intended to yield data on diagnostic test accuracy (DTA) for the various imaging technologies, which would be pooled in a meta-analysis to inform the main parameters for the economic models. The meta-analysis was undertaken by the systematic review team in order to elicit pooled diagnostic test accuracy estimates of PET/CT for metastatic colorectal cancer. The quality of evidence from literature identified in the systematic review was poor and out of 16 studies identified only seven studies were considered to be eligible for meta-analysis (158;160-165). The majority of papers did not reveal how sample patients were recruited, but three did report taking a consecutive approach (158;163;164). Figure 3-15 details the forest plot presenting the accuracy data of PET/CT in identifying metastatic CRC in n=281 patients from the seven studies.

Sens	STUDY	Spec
7/8	D'Souza 2009	0/0
63/67	Chua 2007	6/8
60/61	Kong 2008	4/4
60/66	Selzner 2004	9/10
26/28	Rappeport 2007	3/3
23/24	Coenegrachts 2009	0/0
17/27	Wildi 2008	0/4

Sensitivity
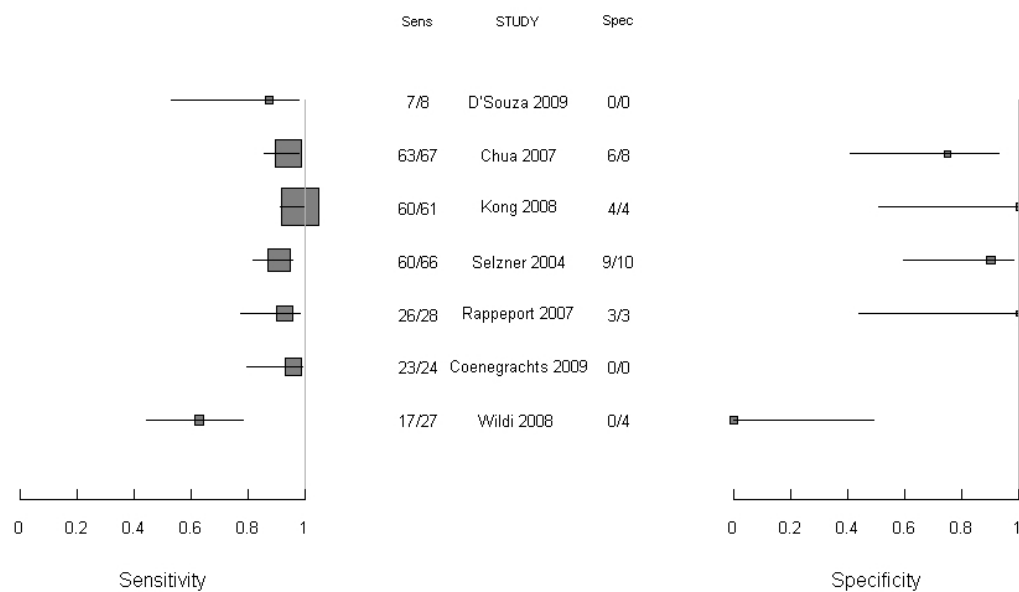
Specificity

Figure 3-15: Accuracy of PET/CT in detecting hepatic metastases

The systematic review team undertook two univariate meta-analyses for sensitivity and specificity separately. There was little evidence of heterogeneity in sensitivity estimates, and therefore a fixed effects model was

used, giving an overall estimate of sensitivity of 0.91 (95% CI 0.87 to 0.94). There was evidence of some heterogeneity in the specificity estimates, so a random effects model was used and the overall estimate of specificity was found to be 0.76 (95% CI 0.58 to 0.88). The validity of these pooled estimated is compromised due to the retrospective study designs (case series, audits), the variation in the types of reference standard used (differential verification bias) which undermines the estimates and several other types of bias including review bias.

Due to the inadequacies and reporting bias in the identified papers, the pooled estimates for PET/CT were considered unlikely to be an accurate reflection of the mean diagnostic test accuracy. The confidence intervals for the pooled estimates were also tight around the pooled mean, restricting the level of uncertainty represented. Therefore, the meta-analysis of DTA data was deemed to be inappropriate for use in the baseline economic model and papers identified by the systematic review were considered along with papers previously identified through the economic search, to find reasonable estimates of DTA for the economic models.

Four papers provided DTA evidence of PET/CT at the patient level for diagnosing metastatic recurrence in comparison to ceCT (158;161-163). These papers were all deemed to be of variable quality. The DTA evidence reported by Chau et al. (161) compared ceCT and PET/CT in identifying hepatic metastases and appeared to provide reasonable point estimates and confidence intervals for PET/CT, however, the point estimate of specificity for ceCT was very low (0.25) compared with that reported in other studies, and appeared to bias the results in favour of PET/CT. Therefore, the point estimates and confidence intervals for PET/CT and the sensitivity for ceCT were taken from Chau et al. (161), with the point estimate for ceCT specificity and accompanying wide confidence interval from Selzner et al. (158). The DTA estimates, their standard errors and the distributions used in the probabilistic model are detailed in Table 3-21.

Table 3-21: Metastatic CRC model Diagnostic Test Accuracy parameters

Diagnostic Parameters	Point estimate	Standard error	Probabilistic distribution	Data Source
ceCT sensitivity	0.91	0.05	Beta	(161)
ceCT specificity	0.7	0.15	Beta	(158)
PET/CT sensitivity	0.94	0.04	Beta	(161)
PET/CT specificity	0.75	0.17	Beta	(161)

As the diagnostic test accuracy parameters are binomial in nature, Beta distributions were fit using the Method of Moments (4) with the mean estimates and standard errors derived from the reported confidence intervals.

With regards to dealing with joint test outcomes for the intervention arm of the model where PET/CT is an add-on after the conventional ceCT test; the PET/CT test characteristics were superior to the conventional test, and this was taken to represent combined DTA, as adopted in other studies involving combined tests (108).

Treatments

The treatment combinations for metastatic diagnosis (metastatic recurrence at one site, extra metastases and no metastatic recurrence) were determined from the literature and in consultation with clinical experts, as detailed previously in Figure 3-11. For a diagnosis of metastases (at one site) there is only one treatment option (pre-operative chemotherapy followed by metastatic surgery) and therefore this treatment is assigned a probability of 1. Likewise, patients with no metastatic recurrence are treated as wait and watch with an annual follow-up, and this treatment option was also assigned a probability of 1. The proportion of patients receiving each of the two treatment options for extra metastases was determined from previous economic evaluations (129;130). Table 3-22 details these treatment options.

Table 3-22: Metastatic CRC model Treatment Parameters

Treatment Parameters	Point estimate	Standard error	Probabilistic distribution	Data Source
Metastases (1 site)				
Pre-op chemo & metastatic surgery	1			Author assumption
Extra metastases				
Pre-op chemo & metastatic surgery	0.2	0.04	Beta	(129)
Palliative care & chemotherapy	0.8		1 - above	(129)
No metastases				
Wait Watch; annual follow up	1			Author assumption

Pre-op=pre-operative, chemo=chemotherapy

The treatment parameter estimates for metastases and no metastases were kept constant in the probabilistic analysis, as everyone diagnosed as such will receive these respective treatments. Uncertainty in the extra metastases parameter estimates was accounted for by fitting a Beta distribution. As there are two treatment options summing to a probability of 1, Beta distributions were fitted, using the method of moments to find alpha and beta, given the reported mean value and standard error (129).

Survival

Five year survival estimates for CRC patients who do not experience metastases were reported by the American Cancer Society (133); this estimate is the same as that used for no recurrence in the recurrent CRC model. Survival estimates for metastases at one site and extra metastases (for surgery with palliative intent and palliative care alone) were taken from the American Joint Committee on Cancer staging manual (138). These estimates were used to determine annual mortality rates under the assumption of an exponential survivor function, for use in the survival analysis, as detailed in section 3.3.4. Table 3-23 details the survival parameters for the metastatic colorectal cancer model.

Table 3-23: Metastatic CRC model Survival Parameters

5 year Survival Parameters	Point estimate	Standard error	Probabilistic distribution	Data Source
Metastases	0.24	0.03	Beta	(138)
Extra metastases (surgery)	0.12	0.04	Beta	(138)
Extra metastases (palliative)	0.06	0.04	Beta	(138)
No Metastases	0.85	0.01	Beta	(133)

The data was binomial, in the form of the probability of survival 5 years post-diagnosis, and therefore Beta distributions were applied for the probabilistic analysis. The Method of Moments was used to fit the distributions, using the mean survival estimate and standard errors reported for Metastases and Extra metastases (surgery with palliative intent). No standard error was given for the Extra metastases - palliative alone estimate, so the standard error was assumed to be the same as that for Extra metastases - surgery. No standard error was provided with the 5 yr survival estimate for patients who have No metastases (133) and therefore, a sample size was derived from the Scottish network CRC dataset (121). This data was used to represent a cohort of AJCC1-3 CRC patients who would be susceptible to metastases (n=2409). The prevalence estimates (detailed in Table 3-17) were then applied to derive a population for no metastases (n=1445), and the number of events (deaths) was determined from the 5 years survival estimate (133). Beta distributions were then fit to this data employing the Method of Moments (4) using the number of deaths to represent the alpha parameter, and the sample size minus the events to represent the Beta parameter.

Quality of life / Utility

Utility estimates were incorporated into the model, representing the average quality of life for patients in the no metastatic recurrence, metastases at one site and extra metastases groups. Patients who were inaccurately diagnosed as no metastatic recurrence (false negatives) and therefore failed to receive treatment for either metastases or extra metastases in the first year were assigned a disutility for that year to account for the negative impact on the patients' quality of life. Likewise, patients who were inaccurately diagnosed

as metastatic (false positives) and received unnecessary metastatic surgery or treatments for extra metastases were assigned a lower utility status for that year to account for the negative impact of unnecessary treatment on quality of life.   Table 3-24 details the utility parameter estimates, standard errors and distributions used in the probabilistic model.

Table 3-24: Metastatic CRC model Utility Parameters

Utility Parameters	Point estimate	Standard error	Probabilistic distribution	Data Source
No Metastases	0.91	0.11	Gamma (disutility)	(122)
Metastases	0.84	0.12	Gamma (disutility)	(122)
Extra Metastases				
Palliative care alone	0.52	0.08	Gamma (disutility)	(132)
Metastatic surgery & palliative	0.74	0.21	Gamma (disutility)	(137)
Fail to receive				
Metastases treatment	0.3	0.08	Gamma	Assumption (132)
Extra metastases treatment	0.2	0.08	Gamma	Assumption (132)
Receive unnecessarily				
Metastases treatment	0.74	0.14	Gamma (disutility)	(137)
Extra metastases treatment	0.61	0.2	Gamma (disutility)	(132)

Utility estimates and standard errors reported in Ramsey et al. (122) were used to represent the mean quality of life for patients with no metastases and metastases at one site, for five years post-diagnosis. A utility estimate reported in Tengs & Wallace (132) for colorectal patients who receive best palliative/supportive care was taken to represent mean quality of life for patients with extra metastases who receive palliative care alone. Patient with extra metastases who receive surgery with palliative intent are assigned a slightly higher utility value (137) to represent their slightly better scenario, as recommended by clinical experts. These were applied in the model for five years post-diagnosis, followed by UK age adjusted population utility weights for the remainder of time alive in the model (124).

Patients incorrectly diagnosed received their true disease stage utility, but with a disutility relating to the inappropriate treatment they received for a

specified duration as detailed previously in Table 3-7. Patients who were inaccurately diagnosed and failed to receive treatment for metastases at one site (pre-operative chemotherapy and metastatic surgery) were assumed to receive a disutility of 0.3 for a year, and patients who failed to receive treatment for Extra metastases received a disutility of 0.2 for a year, reflecting the impact on quality of life for delayed treatment. The disutility values and standard errors were assigned at the authors discretion, based on advice from clinicians and related utility information provided in various papers (132). The authors assumed a corresponding standard error of 0.08.

Patients who received unnecessary metastatic treatment received an alternative, lower utility value of 0.74 (132) for one year to reflect the impact on their quality of life during their unnecessary treatment, a standard error of 0.14 was assumed to represent uncertainty in the point estimate. Patients who received unnecessary extra metastases treatments were assigned an even lower alternative utility value (0.61 with a corresponding standard error 0.2 (137)) for one year, to reflect the considerable impact on quality of life of receiving unnecessary treatment for misdiagnosed extra metastases.

Similar to the approach adopted for the primary and recurrent models, Gamma distributions were set on disutility (Disutility= 1-Utility) so that the distribution was constrained on the interval zero to infinity, allowing very low and even negative utility values. The mean utility values and corresponding standard errors were reported in the literature (or derived based on author assumption where appropriate), and therefore the Method of Moments for gamma was used to calculate the shape (alpha) and scale (beta) parameters to fit the gamma distributions.

Costs

As in the previous models, the costs for the metastatic model are attributed to the alternative imaging devices (as a cost per scan) and the various treatment options for diagnoses of no metastases, metastases at one site, extra metastases. NHS reference costs data was used (101;145) along with other data sources for the various treatment options (100;120;149). The various unit costs

point estimates, standard errors and the distributions used in the probabilistic analysis are detailed in Table 3-25.

Table 3-25: Metastatic CRC model costs

Item	Unit Cost (£)	Standard error (£)	Probabilistic distribution	Data Source
Imaging Devices				
ce CT scan (chest, abdomen, pelvis)	143	22	Normal	(101)
MRI scan (pelvis)	179	24	Normal	(101)
PET/CT scan	800	100	Normal	(146;148)
Treatments				
Chemotherapy	11532	-	-	(100;120;145)
Palliative Care	2468	494	Normal	(149)
Metastatic surgery	9134	1827	Normal	(145)
Wait & watch: annual follow-up	60	13	Normal	(101)

The cost of the imaging devices are the same as those used in the previous models. The treatment option combinations for the metastatic model are different to those in the primary and recurrent models; however the cost of the component treatments were assigned in the same way. For example, the cost of metastatic surgery includes surgical consultation, theatre time, staff costs and an average inpatient stay of 10 days; pre-operative chemotherapy includes six months treatment with 5-FU and Oxilaplatin, and the wait and watch treatment involves an annual oncology consultation.

For the probabilistic analysis normal distributions were considered to be appropriate for representing the unit cost parameters, as in the previous two models. The point estimates were derived from UK and Scottish reference costs sources, and as these are very large data sources they can be considered to have sufficient sample sizes for the central limit theorem to apply.

The expected costs of treatment for the groups were calculated using the proportion of patients receiving each treatment option within each different group. In the model, if a patient was diagnosed accurately they would receive

their optimal treatment option and the associated costs of that treatment. If a patient is inaccurately diagnosed they incur the cost of the treatment for the (mis)diagnosed group, followed by the discounted cost of treatment for their true diagnosis the following year (i.e. it is assumed that the true diagnosis would be identified within a year if the patient were still alive). Costs were discounted at 3.5% (24).

3.5 Deterministic Results

Having detailed the development of the three models and how they were parameterised, this section now reports the deterministic results for each model reporting the incremental cost per correct diagnosis and incremental cost per QALY. The exploration of uncertainty through probabilistic sensitivity analysis and value of information analysis is reported in Chapter 4.

The evaluations were undertaken from the perspective of the UK NHS, reporting short term outcomes in terms of the incremental cost per correct diagnosis, and longer term outcomes in terms of the incremental cost per quality adjusted life year (QALY) gained. For the three models, five evaluations were undertaken to assess the cost-effectiveness of PET/CT as an add-on imaging device in pre-operative staging for (i) primary rectal cancer, (ii) primary colon cancer, (iii) recurrent rectal cancer, (iv) recurrent colon cancer and (v) metastatic disease. In the primary colon and rectal models a scenario analysis was undertaken to assess the future potential (if and when the technology becomes readily available in the UK) for contrast enhanced PET/CT (cePET/CT) as a replacement for conventional ceCT, rather than as an add-on imaging device.

3.5.1 Primary colorectal cancer results

Primary Rectal Cancer

Table 3-26 details the expected costs of the imaging involved in the conventional strategy and the intervention strategy, the expected probability of a correct diagnosis under each strategy and the probability of a true positive diagnosis. Cost-effectiveness for these two alternative measures of effectiveness is presented as incremental cost-effectiveness ratios (ICER). Correct diagnosis embodies the probability of identifying both true positives and true negatives in the model, and on this basis, the addition of PET/CT is dominated by the conventional strategy. PET/CT is both more expensive and less effective in terms of the probability for correct diagnosis.

Table 3-26: Primary rectal cancer – cost per correct diagnosis

Diagnostic Tool	Mean cost per scan £	Probability correct diagnosis	Probability true positive
CT	£322	0.71	0.31
CT + PET/CT	£1,122	0.61	0.44
Difference	£800	-0.10	0.13
ICER		CT dominates	£5,931

This is due to the add-on nature of the PET/CT intervention, and the confirmatory positive strategy adopted, which identifies lots of true positives (and false positives) but results in fewer true negatives being identified, and overall fewer correct diagnoses. In using an add-on technology there is a trade-off between specificity and sensitivity, and in this case, adopting a confirmatory positive strategy resulted in improvements in sensitivity, at the expense of a reduction in specificity.

Reporting the overall result of correct diagnoses (true positives plus true negatives) implicitly gives an equal weight to the incorrect diagnoses: false positive and false negatives, however, they are not equal. The confirmatory positive strategy recommended by the clinicians confirms that in this case false negatives are considered to be worse than false positives; as clinicians would rather gain additional true positive cases and over-treat additional cases incorrectly (false positives) than risk some cases going undetected (false negatives). To understand the correct diagnoses outcome more clearly, it is worthwhile looking at the probabilities of each of the DTA characteristics. Table 3-27 illustrates the outcomes in terms of the probability of correct diagnoses, true positives, false negatives, false positives and true negatives, for the conventional, and add-on PET/CT strategy adopted in this model. The table also illustrates outcomes if a confirmatory negative approach had been adopted, and if PET/CT had only been implemented when the conventional test gave a negative result.

Table 3-27: Primary rectal cancer – DTA outcomes under different strategies

Outcomes	Conventional	add on PET/CT confirmatory +	add on PET/CT confirmatory -	PET/CT only if conventional -ive
Correct diagnoses	0.71	0.61	0.72	0.61
True positives (TP)	0.31	0.44	0.26	0.44
False negatives (FN)	0.16	0.02	0.20	0.02
False positives (FP)	0.13	0.36	0.07	0.36
True negatives (TN)	0.41	0.17	0.46	0.17
Cost	£322	£1,122	£1,122	£774
Cost-effectiveness		**ICER**	**ICER**	**ICER**
Correct diagnoses		Dominated	£89,695	Dominated
True positives		£5,931	Dominated	£3,349

Table 3-27 shows that while the conventional strategy results in a greater probability of correct diagnosis than the add-on PET/CT intervention, in terms of true positives (TP), PET/CT has the better outcome. The conventional strategy identifies more true negatives (TN) than add-on PET/CT, and once they are combined with TP the conventional strategy is the most effective in terms of correct diagnosis. The add-on PET/CT strategy identified more true positives, but at the expense of false positives, which have a higher probability than in the conventional arm, yet in the conventional arm, many true negatives are identified, but at the expense of a greater proportion of false negatives.

Table 3-27 also illustrates the potential outcomes if a confirmatory negative strategy had been adopted for add-on PET/CT. In such a strategy (which favours specificity over sensitivity), even fewer true positives are identified (0.26) than in the confirmatory positive intervention (0.44), and the conventional (0.31), however, more true negatives are identified (0.46) resulting in an overall correct diagnosis slightly greater than the conventional strategy. The alternative strategy in which add-on PET/CT is only used when

the conventional test is negative, results in the same DTA outcomes as the confirmatory positive strategy.

Table 3-27 also details the cost of each strategy and reports the incremental cost-effectiveness ratios (ICER) with effectiveness measured by correct diagnoses and by true positives. The confirmatory negative strategy may have the best outcome in terms of correct diagnoses, however, it is only marginally so and this is reflected in a high ICER of approximately £90,000. If true positives were the measure of effectiveness, then the confirmatory negative strategy would be dominated by both the conventional strategy and the confirmatory positive strategy. When a correct diagnosis is the outcome of interest, add-on PET/CT with confirmatory positive is dominated by the conventional strategy, however under the true positives outcome, it has an ICER of approximately £6000. When PET/CT is used only when the conventional test is negative, the DTA results are the same as confirmatory positive strategy, however cost-effectiveness is improved with an ICER of approximately £3000, as the cost of PET/CT would only be incurred in some of the patients, those who get a negative result from the conventional test (56%). This is a more efficient research design and has been used by others (134) however, in practice this approach may be less well accepted and in the case of PET/CT, clinicians advised that all patients would be given a scan regardless of the conventional test outcomes.

Given the add-on PET/CT intervention with a confirmatory positive strategy compared with the conventional strategy, the cost-effectiveness depends on whether the effectiveness outcome of interest is 'probability of correct diagnosis' (conventional strategy dominates) or 'probability of true positives' (PET/CT cost-effective at £5,900 per true positive identified). These interim analysis outcomes detailed in Tables 3-26 appear to ignore the impact of false positives and false negatives (or implicitly value them as equal), which is not the case. They are incorporated in the longer term analysis through the treatment given under correct and incorrect diagnosis. The longer term model also captures the resultant life expectancy and quality of life for patients, reporting the incremental cost per QALY gained, which incorporates the

diagnoses which were true positives, true negatives, false positives and false negatives.

Table 3-28 details the long term cost-effectiveness outcomes. The expected cost per person incorporates the imaging device and the subsequent treatment costs associated with the conventional and intervention strategies. The outcomes are reported in terms of QALYs under each strategy and cost-effectiveness is reported in terms of the incremental cost per QALY gain for primary rectal cancer.

Table 3-28: Primary rectal cancer – cost per QALY gain

Diagnostic Tool	Cost per person (£)	QALY gain
MRI + CT	£15,151	9.42
MRI + CT + PET/CT	£17,418	9.43
Difference	£2,267	0.0053
ICER	**£431,691**	

These outcomes reflect the cost and quality of life implications of false negative and false positive diagnoses from the model. Tables 3-26 and 3-27 showed that under an interim effectiveness outcome of correct diagnoses, the conventional strategy dominated the add-on PET/CT intervention. However, Table 3-28 illustrates an ICER outcome, in which the intervention arm is no longer dominated. This is due to the superior sensitivity from the PET/CT intervention which resulted in a greater probability of true positives than in the conventional strategy. However, the impact of the additional false positives and false negative outcomes are also incorporated here. The low specificity of PET/CT as an add-on technology (under a confirmatory positive strategy) results in a greater number of false positive outcomes, in which patients are over-staged and incur additional costs and suffer quality of life impacts for unnecessary treatments. The resultant impact in terms of QALYs is marginal, with the PET/CT intervention gaining only 0.005 QALYs per person. This is reflected in the extremely high ICER outcome from the primary rectal cancer analysis.

The addition of PET/CT to the conventional strategy involves an additional cost of approximately £432,000 per QALY gained and therefore would not be considered cost-effective in the UK under the usual definition of between £20,000 and £30,000 per QALY (£20k/QALY < ICER < £30k/QALY) (24).

Primary Colon Cancer

The results from the primary colon cancer analysis are now presented. Table 3-29 details the outcomes in terms of both the incremental cost per correct diagnosis and the incremental cost per true positive diagnosis. The expected costs of the imaging devices involved in each strategy are presented along with the expected probability of a correct diagnosis, and probability of a true positives diagnosis under each strategy.

Table 3-29: Primary colon cancer – cost per correct diagnosis

Diagnostic Tool	Mean cost per scan £	Probability correct diagnosis	Probability true positive
CT	£143	0.65	0.26
CT + PET/CT	£943	0.60	0.43
Difference	£800	-0.05	0.18
ICER		**CT dominates**	**£4,481**

The addition of PET/CT is dominated by the conventional strategy in terms of correct diagnosis, i.e. PET/CT is both more expensive and less effective than CT alone. The add-on nature of the PET/CT strategy gives rise to a combined DTA where more cases of true positives are identified than the conventional strategy, but far fewer true negatives and therefore has a lower overall probability of correct diagnosis. However, as with primary rectal cancer, when the probability of a true positive diagnosis is used as a measure of effectiveness the PET/CT intervention has an ICER, in this case of approximately £4000 per true positive diagnosis. These interim outcomes do not reflect the inaccurate diagnoses (FP and FN test outcomes) under either strategy and therefore, it is the longer term QALY analysis that is more informative.

Table 3-30 details the cost per QALY outcomes. The expected costs incorporate the cost of the imaging device as well as the subsequent costs of treatment following the outcomes of each strategy. The expected outcomes are reported in terms of QALYs under each strategy and the cost-effectiveness in terms of incremental cost per QALY gain for primary colon cancer.

Table 3-30: Primary colon cancer – cost per QALY gain

Diagnostic Tool	Cost per person (£)	QALY gain
CT	£12,815	9.41
CT + PET/CT	£15,066	9.42
Difference	£2,253	0.013
ICER	**£171,018**	

The impact of the additional false positives and false negative in the model are incorporated in these outcomes. The low specificity of PET/CT as an add-on technology (under a confirmatory positive strategy) results in a greater number of false positive outcomes, in which patients are over-staged and incur additional costs and suffer quality of life impacts for unnecessary treatments. However, in the conventional arm there are a greater proportion of false negative diagnoses where patients fail to receive necessary treatment for a year, and incur negative impacts on their survival and quality of life impacts. The resultant impact in terms of QALYs is marginal, with the PET/CT intervention gaining only 0.01 QALYs per person. This is reflected in a high ICER (£170,000), which is lower than the ICER for PET/CT in the primary rectal analysis (£430,000) but still far too high to be considered cost-effective.

The addition of PET/CT to the conventional strategy involves an additional cost of approximately £171,000 per QALY gained and would not be considered cost-effective under the typical UK definition of £20,000 to £30,000 per QALY (24).

Primary Colorectal Cancer Scenario: cePET/CT as a lone technology

PET/CT technology is rapidly advancing (120;148) and contrast enhanced PET/CT (cePET/CT) is already being explored in Japan (109;110). Improvements in PET/CT technology may lead to improved accuracy of the CT component, so that if and when cePET/CT becomes readily available for use, it could potentially be used as a replacement for conventional ceCT, rather than as an add-on technology.

Primary Rectal Cancer Scenario

The primary rectal scenario explored the potential cost-effectiveness of such an advanced cePET/CT technology; replacing standard ceCT test with contrast enhanced PET/CT (cePET/CT) in addition to an MRI scan. The results are detailed in Table 3-31, and offer an improvement in cost-effectiveness compared to the baseline add-on PET/CT results detailed in Table 3-28.

Table 3-31: Primary rectal cancer scenario – cost per QALY

Diagnostic Tool	Cost per person (£)	QALY gain
MRI+ CT	£15,120	9.43
MRI+ cePET/CT	£16,095	9.44
Difference	£975	0.009
ICER	**£107,652**	

In primary rectal cancer, an advanced cePET/CT technology would be useful but would not negate the requirement for an MRI scan, and therefore, the additional benefit in terms of QALYs is marginal, while there would be an additional cost of the cePET/CT scan. This scenario results in an ICER of £107,600 and therefore this potential future strategy of cePET/CT as a replacement for contrast enhanced CT in primary rectal cancer is not considered to be cost-effective under the typical willingness to pay threshold in the UK, i.e. £20,000 per QALY < ICER < £30,000 per QALY (24).

Primary Colon Cancer Scenario

MRI scans are not used for the detection of colon cancer, and therefore cePET/CT could be used as a complete replacement for the conventional strategy (ceCT), if and when it becomes available in the UK. With regards to the primary colon scenario, the results indicate that there is potential for this strategy to be highly cost-effective. Table 3-32 details the cost-effectiveness results.

Table 3-32: Primary colon cancer scenario – cost per QALY

Diagnostic Tool	Cost per person (£)	QALY gain
CT	£12,766	9.45
cePET/CT	£12,972	9.47
Difference	£206	0.02
ICER	**£12,832**	

The incremental cost of cePET/CT is £200 and the improved diagnostic test accuracy of cePET/CT is reflected in a QALY gain of 0.02. This results in an ICER of £12,800 which is considerably below the typical UK willingness to pay threshold of £20,000 to £30,000 per QALY (24). Therefore, using contrast enhanced PET/CT as a replacement technology for contrast enhanced CT is likely to be cost-effective, given current evidence.

3.5.2 Recurrent colorectal cancer results

Recurrent Rectal Cancer

The cost-effectiveness results from the recurrent rectal cancer analysis are detailed in Table 3-33, in terms of the incremental cost per correct diagnosis and the incremental cost per true positive diagnosis. The expected costs of the imaging involved in both the conventional and intervention test strategies are detailed along with the expected probabilities.

Table 3-33: Recurrent rectal cancer – cost per correct diagnosis

Diagnostic Tool	Mean cost per scan £	Probability correct diagnosis	Probability true positive
MRI + CT	£322	0.88	0.60
MRI+CT + PET/CT	£1,122	0.95	0.65
Difference	£800	0.07	0.06
ICER		**£11,713**	**£13,769**

The inclusion of a PET/CT scan involves an additional cost of £800 but this also leads to an increased probability of correct diagnosis (true positives and true negatives), resulting in an ICER of approximately £12,000 per correct diagnosis. If the probability of a true positive diagnosis is the outcome of interest, then the add-on PET/CT strategy is also more effective than the conventional strategy, resulting in an ICER of approximately £14,000 per true diagnosis. The PET/CT intervention is superior to the conventional test in terms of both sensitivity and specificity and therefore rather than adopting a confirmatory positive strategy (as was done in the primary CRC model) the PET/CT diagnostic test accuracy was taken to represent joint DTA. As PET/CT has superior DTA characteristics, the interim cost-effectiveness outcomes reflect this. Therefore, given current information, the MRI, CT and PET/CT scan strategy would be considered cost-effective compared with the conventional strategy under the typical UK willingness to pay threshold of £20,000 to 30,000 per QALY gained (24).

These interim outcomes don't reveal the impact of false positives and false negative diagnoses, and therefore the longer term outcomes reporting incremental cost per QALY gains are more informative. The lifetime analysis results are presented in Table 3-34. The expected cost per person is presented which includes the cost of the imaging devices along with the subsequent costs of treatment associated with each of the test strategies. The QALYs under each strategy are presented and cost-effectiveness is reported in terms of cost per QALY gain for recurrent rectal cancer.

Table 3-34: Recurrent rectal cancer – cost per QALY gain

Diagnostic Tool	Cost per person (£)	QALY gain
MRI + CT	£7,243	4.56
MRI + CT + PET/CT	£7,955	4.59
Difference	£712	0.03
ICER	**£21,409**	

The addition of PET/CT to the conventional strategy involves an additional cost of approximately £700, with an increase of 0.03 QALYs. The ICER is approximately £21,500 and would therefore be considered cost-effective under the usual UK definition (£20k/QALY < ICER < £30k/QALY) (24).

Recurrent Colon Cancer

The results from the recurrent colon model are now presented. Table 3-35 details the interim analysis outcomes for probability of correct diagnosis and probability of true positive diagnosis. The expected costs of the imaging devices for each strategy are presented along with the expected probability of a correct diagnosis under each strategy and the cost-effectiveness in terms of incremental cost per correct diagnosis for recurrent colon cancer.

Table 3-35: Recurrent colon cancer – cost per correct diagnosis

Diagnostic Tool	Mean cost per scan £	Probability correct diagnosis	Probability true positive
CT	£143	0.67	0.37
CT + PET/CT	£943	0.94	0.65
Difference	£800	0.27	0.28
ICER		**£2,999**	**£2,857**

The PET/CT intervention is superior to the conventional test in terms of both sensitivity and specificity and this is reflected in the correct diagnosis and true positive outcomes. The benefit of add-on PET/CT is through improved sensitivity, with an increase of 0.28 in the probability of true positive diagnosis and resulting in fewer false negative diagnoses than in the conventional arm. As the DTA of PET/CT is superior to that of ceCT, PET/CT was taken to represent joint DTA in the intervention arm. Including a PET/CT scan as an add-on technology involves an additional cost of £800, but this strategy also improves the probability of correct diagnosis by 28%. The resultant incremental cost-effectiveness ratio is approximately £3,000 per correct diagnosis and per true positive diagnosis. This ratio would be considered cost-effective compared with the conventional strategy under the typical UK threshold for cost-effectiveness, i.e. £20,000 to £30,000 per QALY (24).

The outcomes in terms of incremental cost per QALY gain are presented in Table 3-36. The expected costs incorporate the cost of the imaging devices as well as the longer term costs associated with treatment under each strategy. The expected outcomes are in terms of QALYs under each strategy.

Table 3-36: Recurrent colon cancer – cost per QALY gain

Diagnostic Tool	Cost per person (£)	QALY gain
CT	£6,677	4.44
CT + PET/CT	£7,543	4.58
Difference	£866	0.14
ICER	**£6,189**	

In the longer term analysis the addition of a PET/CT scan improves QALYs by 0.14 QALYs at an additional cost of £860. This translates to an ICER of approximately £6,000 per QALY gain and would be considered highly cost-effective under the typical UK definition (£20k/QALY < ICER < £30k/QALY) (24).

3.5.3 Metastatic colorectal cancer results

The deterministic results from the metastatic model are now presented. Table 3-37 reports the incremental cost per correct diagnosis and per true positive diagnosis. The expected cost represents the cost of the imaging devices involved in each strategy and the outcome is the expected probability of a correct diagnosis under each strategy. As with the recurrent colorectal cancer analyses, DTA for PET/CT was found to be superior to ceCT in terms of both sensitivity and specificity and therefore, was used to represent joint DTA in the intervention arm.

Table 3-37: Metastatic cancer – cost per correct diagnosis

Diagnostic Tool	Mean cost per scan £	Probability correct diagnosis	Probability true positive
CT	£143	0.78	0.36
CT + PET/CT	£943	0.83	0.38
Difference	£800	0.04	0.01
ICER		**£19,048**	**£66,667**

The intervention strategy of PET/CT in additional to contrast enhance CT involves an additional cost of £800 and results in an expected improvement in correct diagnoses of 4%. The correct diagnosis outcome incorporates both true positive and true negative outcomes, and as can be seen the intervention arm benefits from improved sensitivity and specificity over the conventional strategy. In terms of true positive outcomes alone, the intervention strategy is only marginally more effective, with an improvement in probability of correct diagnosis of 0.01, leading to a higher ICER of £66,000 which would not be considered cost-effective in the UK. The correct diagnosis outcome reflects the superiority of the PET/CT test in both the sensitivity and specificity characteristics, and results in an ICER of approximately £19,000 which would be considered cost-effective compared with the conventional strategy under the usual definition of cost-effectiveness in the UK (24).

The longer term analysis formally incorporates the impact of false negative and false positive outcomes, and the results are presented in Table 3-38. The expected costs incorporate the cost of the imaging devices as well as the subsequent treatment costs associated with each strategy, while QALYs incorporate the quality of life and life expectancy for patients diagnosed accurately and inaccurately.

Table 3-38: Metastatic cancer – cost per QALY gain

Diagnostic Tool	Cost per person (£)	QALY gain
CT	£10,184	7.48
CT + PET/CT	£10,460	7.49
Difference	£276	0.01
ICER	**£21,434**	

The addition of PET/CT to the conventional strategy involves an additional cost of approximately £300 and results in a gain of 0.01 QALYs. The lifetime QALY impact of the addition of a PET/CT scan is marginal; however impact on cost is also small resulting in an incremental cost-effectiveness ratio of £21,000 per QALY gained. This is likely to be considered cost-effective under the usual definition of £20,000 to £30,000 per QALY (24).

3.6 Summary

This chapter outlined the development of three economic models using existing evidence and clinical opinion as part of an early stage economic evaluation to determine the potential cost-effectiveness of PET/CT as an additional to routine care for staging colorectal cancer. This case study demonstrated the benefit and ease of synthesising current evidence and clinical expertise to develop early stage models, for an explorative economic evaluation, i.e. stage 2 of the iterative economic approach. Systematic literature reviewing (166) is an efficient, structured technique which ensures all relevant literature has been captured, and following this with meta-analysis techniques (167) is an ideal means of establishing pooled estimates of key parameters which can be used in economic models. However, decision analytic modelling in itself is a mechanism for synthesising information, and this chapter highlighted that even when existing evidence is of poor quality and clearly subject to various forms of bias (including reporting bias), clinical expertise and research group consensus decision making can be used to decide on appropriate mean point estimates and wide uncertainty intervals to capture the present uncertainty. This is particularly of use when meta-analysis cannot be undertaken due to lack of data or (in the case of recurrent and metastatic CRC) where meta-analysis results do not credibly represent the wide range of uncertainty. Early decision analytic models may be undertaken prior to the design of large scale trials, and require to capture cost-effectiveness based on current (and potentially limited or poor quality) evidence. Just because evidence is poor, does not mean that the economic evaluation will be poor quality.

With regards to the alternative interventions and delivery strategies compared in the PET/CT models, it is of interest to consider the extent to which this was led by the current clinical context. One of the key advantages of early DAM is that it enables explicit strategy identification through exploring a wide range of alternative interventions; however, in practice implicit judgements are often made concerning which strategies should be included, based on arbitrary judgements. Recommended good practice for decision modelling (30;37)

emphasises that all relevant alternatives and strategies should be incorporated into a decision model.

In the PET/CT assessment, the choice of comparators was led primarily by the research question set by the funders and the clinicians involved in the study. As such the comparators, structure and boundaries of the decision model were potentially constrained by the current clinical context, limiting the range of strategies compared in the decision model. For example, early into the research project it became evident that as an add-on technology PET/CT was not efficient and would have limited benefit in primary CRC; while an advanced cePET/CT technology (which may become available in the UK in the near future) has potential value as a replacement technology in primary CRC. The original scope of the research question, defined by the funding body, was restrictive with regards to specifying PET-CT as an add-on device in primary colorectal cancer and therefore limited the baseline analysis. To counter-act this limitation, a scenario analyses was incorporated in the primary CRC assessment, to model the potential impact of cePET-CT as a stand alone device.

With regards to clinician led approaches potentially restricting the strategies modelled there were two examples. The confirmatory negative strategy for the add-on test was not incorporated as an alternative strategy in the baseline model as clinicians identified that this would not be considered in practice. Additionally, a potentially much more efficient strategy was identified as appropriate for the economic evaluation, in which PET-CT is used as an add-on test only when the conventional test is negative. This strategy was dismissed by the clinicians due to their concerns over the practicalities and possibly a preoccupation with current practice.

The confirmatory positive strategy recommended by the clinicians confirms that false negatives are considered to be worse than false positives; as clinicians would rather gain additional true positive cases and over-treat additional cases incorrectly (false positives) than risk some cases going undetected (false negatives). This is also of interest when the possibility for litigation proceedings to the NHS is considered. A confirmatory negative

approach has a much higher proportion of false negative cases (as detailed in Table 3-27) than the other strategies, and therefore may result in a greater proportion of unhappy patients attempting to sue the NHS for false negative results. The potential importance of litigation costs to the NHS could be explored in the model by incorporating an additional scenario analysis. By making an assumption about the proportion of false negative cases that will proceed to legal/court proceedings (based on morbidity through failing to detect/early treatment of cancer) the potential additional litigation costs can be incorporated.

While the confirmatory positive approach was not considered to be the most efficient from an economic standpoint (the 'PET-CT only if conventional test is negative' was the most efficient approach as demonstrated in Table 3-27), by insisting on a confirmatory positive approach, doctors are implicitly incorporating an additional source of benefit (through less false negatives, and less unhappy patients). An evaluation may miss this implicit benefit in the confirmatory positive approach, which may only become apparent when the (importance of) potential costs of litigation proceedings are assessed in the economic model.

3.6.1 Primary CRC

Few diagnostic tests have both high sensitivity and specificity and therefore combining two tests can help improve DTA, but creates a problem where results from two tests conflict with one another. Given the add-on nature of the PET/CT intervention, a strategy must be adopted to deal with conflicting test results. A confirmatory positive strategy, also known as the 'either positive' strategy (106), accepts positive results when either of the tests give a positive result; while a confirmatory negative strategy favours negative results and therefore a positive result is only accepted when both tests report a positive. Adopting one of these strategies involves a trade-off between sensitivity and specificity. The confirmatory positive strategy results in an overall larger number of positive outcomes, both true positives and false positives, and a reduction in negative outcomes, while the confirmatory negative approach favours specificity and results in a greater proportion of both true and false negative outcomes (as demonstrated in Table 3-27). In the PET/CT case study, a confirmatory positive approach was adopted as recommended by clinicians. This implies a preference for improving true positive outcomes and that false negative results (in which patients with the disease would be undetected or under-staged) are considered to be worse than false positive outcomes where patients would be treated for a disease they do not have.

An alternative strategy, in which add-on PET/CT is only used when the conventional test is negative, is likely to be a much more efficient research design. As demonstrated in Table 3-27, such a strategy would give the same DTA outcomes as a confirmatory positive strategy, however, the cost of a PET/CT scan (which is approximately four times that of the conventional scans) and the inconvenience to the patient would only be incurred when conventional imaging results are negative, reducing the overall cost of the strategy. Table 3-27 illustrated this for primary rectal cancer, showing that using PET/CT only when conventional tests gave negative results would improve short-term cost-effectiveness (where probability of a true positive diagnosis is the effectiveness endpoint of interest) with an ICER of

approximately £3,000 per true positive diagnosis rather than £5,900 under a test all approach. This is a more efficient research design and has been used by others (134) however, in practice this approach may be less well accepted and in the case of PET/CT, clinicians advised that all patients would be given a scan regardless of the conventional test outcomes. The add-on nature of the PET/CT technology also causes difficulty for interpreting short term cost-effectiveness based on DTA. The cost-effectiveness of the PET/CT intervention in primary CRC depends on whether the effectiveness outcome of interest is the probability of correct diagnosis, whereby the conventional strategy dominates, or probability of a true positive diagnosis, in which case the PET/CT intervention is cost-effective at £5,900 per true positive identified. The probability of correct diagnosis considers the overall advantage of the combined tests, in terms of both true positives and true negatives, whereas the probability of a true positive outcome is only interested in the sensitivity of the tests. As a confirmatory positive strategy was adopted, it could be argued that identification of true positive cases is the outcome of interest; however, it is the longer term outcomes that need to be considered to give a more appropriate measure of cost-effectiveness.   The lifetime analysis formally incorporates the impact of accurate and inaccurate (false positive and false negative) diagnoses, in terms of the costs of the inappropriate and appropriate treatments and the resultant impact on quality of life.

The cost per QALY analyses for primary rectal and colon cancer showed that the confirmatory positive strategy and low specificity of PET/CT in primary CRC results in a greater number of false positive outcomes, in which patients are over-staged and incur additional costs and suffer quality of life impacts for unnecessary treatments. This is reflected in the extremely high incremental cost per QALY outcomes. Further, the therapeutic impact literature identified in the systematic review (112-114;143) found that while PET/CT may potentially impact on accurate staging of primary colorectal cancer, it had only a minor impact on changing patient management. In the decision analytic models developed for this analysis, both the rectal and colon primary models identified an incremental QALY gain of only 0.005 and 0.01 for rectal and colon

cancer respectively, indicating that PET/CT as an add-on imaging device in primary CRC does not have any overall impact on patient outcomes.

The cost per QALY results for both the primary rectal and primary colon evaluations are extremely high, greater than £400,000 per QALY and greater than £170,000 per QALY respectively. As such, PET/CT is not cost-effective in either primary rectal or primary colon cancer given the UK recommended QALY threshold of £20k-£30k per QALY (24).

3.6.2 Recurrent CRC

The recurrent CRC model found PET/CT as an add-on imaging device to have an ICER of £21,409 for rectal cancer and £6,189 for colon cancer. Considering the UK monetary threshold of £20-30,000 per QALY, these can be considered to be cost-effective.

The ICER for the recurrent colon cancer analysis is considerably lower than that for recurrent rectal cancer, indicating that PET/CT is more cost-effective in the assessment of colon recurrence than in rectal recurrence. This difference is likely to be due to the sensitivity estimate for the CT diagnostic test parameter, which has considerably lower sensitivity than the PET/CT sensitivity estimate. Uncertainty around these parameter estimates is explored in Chapter 4 in the probabilistic sensitivity analysis.

Meta-analyses were undertaken using relevant papers identified from the systematic review to elicit pooled diagnostic test accuracy estimates of PET/CT for the recurrent colorectal cancer model. Due to inadequacies and reporting bias in the identified papers, the pooled estimates for PET/CT were considered to be an inaccurate reflection of the diagnostic test accuracy and the confidence intervals were tight around the pooled means, which was considered to be restrictive in terms of capturing the wide range of uncertainty. Therefore, expert judgment was used to determine point estimates and wide uncertainty intervals from the literature.

Most previous economic evaluations undertaken for recurrent CRC have been specifically interested in hepatic metastases. Two papers were identified which were interested in assessing recurrence. A cost-effectiveness analysis undertaken in Canada (130) considered PET/CT in comparison to CT for diagnosing colorectal recurrence, and an Australian based study (129) undertook a cost-consequence analysis of PET versus no PET for diagnosing local recurrence. The decision analytic model developed for this piece of research adds to this literature, providing an assessment of the cost-effectiveness of PET/CT as an add-on imaging device for diagnosing both recurrent rectal and recurrent colon cancer.

The Canadian CEA (130) reports cost savings with the PET/CT approach through avoidance of unnecessary surgeries. The paper does not report the number of unnecessary surgeries avoided in each strategy, just the cost savings. After considering the parameter estimates used in their model, it can be seen that the DTA estimates assigned to PET/CT are superior to those in the CT comparator arm by a wide margin, so it is no surprise that the PET/CT intervention was found to dominate CT. The recurrent model developed for the present analysis utilised DTA estimates from three trials (156-158) and based on this evidence adopted a more conservative approach, assigning the same specificity values to ceCT and PET/CT. This resulted in no difference between the alternative strategies in terms of the unnecessary surgeries outcome in the recurrent colon model; however, the recurrent rectal model did indicate reductions in unnecessary surgeries with the PET/CT intervention. The Australian publication (129) also report cost savings through the use of PET in comparison to a no PET strategy, however, few details are provided as to what the no PET strategy entails.

In comparison to other economic evaluations undertaken in this disease area, the model developed for this piece of research appears to have adopted a more conservative approach in assigning DTA estimates and through incorporating quality of life impacts and overall survival impacts in a cost per QALY outcome. This conservative approach attempted to minimise bias in the model to avoid unfairly favouring the intervention arm (add-on PET/CT).

3.6.3 Metastatic CRC

The metastatic model found PET/CT as an add-on device to have an ICER of £21,434 per QALY gained. This ICER value is within the UK monetary threshold range of £20-30,000 per QALY for determining cost-effectiveness (24).

Most of the existing publications which have undertaken economic evaluations of PET for CRC have been specifically interested in hepatic metastases. An American study (134) developed a decision model to determine the cost-effectiveness of PET and CT imaging in comparison to CT alone. They evaluated outcomes in terms of life year gains and report an incremental cost per life year gained of $16,437. This paper is the most similar to our model, but does not incorporate quality of life impacts.

Another economic evaluation in metastatic CRC was undertaken in France from the national health insurance perspective (135), reporting cost savings of €2671 ($3213) with no change in life expectancy when PET/CT was compared against CT in staging metastatic colorectal cancer. Another American study (136) assessed the cost-effectiveness of PET/CT in comparison to CT for identifying the presence of extra hepatic metastases. They report a cost saving of $5,269, due to unnecessary surgeries avoided, however they provide few details of how their model was constructed.   Details were not provided of the diagnostic test accuracy estimates used in the model or how the impact on patient management was incorporated. None of these metastatic models used probabilistic analysis to incorporate uncertainty to each of the model parameters. Uncertainty around the parameter estimates in the present analysis is explored in Chapter 4 in the probabilistic sensitivity analysis.

3.6.4 Next steps in the analysis

The economic models developed in this chapter relied heavily on clinical expertise, given the dearth of previous economic evaluations and poor quality, biased DTA evidence in this area. Some reviewers may consider this to be a major drawback to this type of decision analytic modelling approach; however, this is precisely the point of early stage decision models: attempting to capture

what knowledge is available and what is not, in a systematic manner.   Just because the existing evidence is poor, does not necessarily mean the outcomes of the economic evaluation are poor. It is entirely appropriate to utilise clinical expertise in such situations, however it is important to capture the uncertainty in these parameters, and explore what this means for uncertainty in terms of the model outputs.

In the PET-CT models, the parameters for the various treatments at diagnosis, probability of receiving those treatments and assumptions regarding the negative health impacts of incorrect diagnosis, were informed directly by the clinical experts on the team, due to a lack of published evidence.   Ideally formal elicitation of clinical opinion would have been undertaken (using Delphi techniques); however, due to time constraints on the research project a slightly less formal elicitation process was used. Tables were created for each of the treatment parameters and the two senior clinicians on the research team completed them with their estimates and surrounding uncertainty estimates. Uncertainty in these model parameters was dealt with in probabilistic sensitivity analysis with wide uncertainty intervals. The structural uncertainty regarding these could also potentially have been dealt with through scenario analyses. For example, alternative scenarios for greater and lower disutility impacts for receiving unnecessary treatment, or failing to receive treatment could have been modelled to explore the impact of this structural uncertainty on the model outcomes. Due to time constraints this was not included in the scenario analyses undertaken, however, given the wide uncertainty around the mean estimates used in the PSA, it is unlikely that such scenario analyses would have changed the conclusions around decision uncertainty.

Chapter 4 reports the outcomes of the probabilistic analyses, explores uncertainty in the cost-effectiveness outcomes and undertakes value of information analyses to demonstrate the advantage of early stage decision modelling in exploring whether further research is worthwhile, and if so, what type of research is appropriate.

4 Setting research priorities

4.1 Introduction

Following on from the preceding chapter where three economic models were developed in order to explore the potential cost-effectiveness of PET/CT as an add-on staging technology for colorectal cancer; this chapter explores uncertainty in the cost-effectiveness outcomes. Value of information analyses are undertaken to explore decision uncertainty further, in terms of whether further research is potentially worthwhile, and if so, what type of research is appropriate. This chapter uses the case study of PET/CT to demonstrate the rationale for funding evidence synthesis and early decision analytic modelling prior to primary research by highlighting the importance of this type of research in determining future research priorities.

4.1.1 Identifying the decision problem

The Health Technology Assessment (HTA) is a programme funded by the National Institute for Health Research (NIHR) (52) which produces independent research about the effectiveness, cost-effectiveness and broader impacts of health care treatments and technologies. The research generated is intended for an NHS audience, particularly those who plan and provide care in the NHS. They commission both primary research and evidence synthesis research, and the topics are identified and prioritised to meet the needs of the NHS. The HTA identify evidence gaps or decision problems for which they commission research through various means, such as consultation with key NHS stakeholders, extracting research recommendations from various resources, direct recommendations and communication with researchers (52). This process is akin to Stage 1 of the iterative approach to economic appraisal (1), but rather than the researchers exploring and forming the research question, the HTA programme have already defined the decision problem, specifying the intervention of interest and the relevant patient groups. It is commonly the case that funding bodies such as the HTA define an overall research question and put out a call for bids to undertake the research. Research bodies are,

arguably, well placed to do this as they are likely to have formal criteria for setting research priorities, as well as the ability to engage with various stakeholders, and also may have financing available to review current evidence in order to define an appropriate research question. However, a call for research from a funding body does not necessarily mean that an appropriate research question has been defined. The call may well be too prescriptive, or alternatively too broad to adequately define the problem, or it may well require some explorative research to define the appropriate question. In the context of the iterative economic framework, stage 1 in itself may involve an iterative process where exploration of the literature leads to updating and re-defining of the research question for evidence synthesis and modelling in stage 2.

In the context of the PET/CT case study, the HTA identified a decision problem for PET/CT in colorectal cancer and commissioned a research brief to answer the question: "In which groups of patients with colorectal cancer does the addition of PET/CT to preoperative staging add most value?"

In the United Kingdom colorectal cancer is the third commonest malignancy behind lung and breast cancer with 37,514 new cases registered in 2006 (94). For patients with colorectal cancer there are a wide range of clinical scenarios and various treatment options with different timings, dependent on the stage of the cancer and also the extent of the cancer growth within each stage. Over the past two decades, a number of diagnostic tools (such as ultrasound, CT, ceCT, MRI, and PET/CT) have entered clinical practice and now facilitate the process of pre-operative staging for CRC. Largely due to its wider availability and relatively low cost, ceCT is the technique most commonly used for staging colorectal cancer, however, supplementary imaging is often needed to provide sufficient information to inform surgical decisions, such as MRI imaging in rectal cancer. PET/CT is recommended for use in some cancers, but its advantages in staging for CRC are less clear. Several studies have shown PET/CT to be more accurate than diagnostic ceCT for staging colorectal cancer (98;99), however, PET/CT scanning is considerably more time consuming than either imaging device alone, and for CRC it is recommended as an addition to routine imaging devices (97), and therefore as an 'add-on' technology it carries

a heavy cost burden. Additionally, the diagnostic accuracy and therapeutic impacts of PET/CT in colorectal cancer are varied and unclear.

Given the unclear evidence as to the effectiveness and cost of PET/CT in colorectal cancer the HTA programme commissioned research to explore and synthesise the existing evidence. The brief specified a systematic review with the requirement for an economic component which incorporated patient management and quality of life, to explore potential cost-effectiveness of PET/CT as an addition to current practice in the UK. Clearly the HTA recognised the importance of exploring the current evidence (regarding effectiveness with an accompanying economic model to determine cost-effectiveness given the current evidence) before commissioning primary research for PET/CT in colorectal cancer.

In response to this call, a multi-disciplinary research team consisting of clinical experts (CRC surgeons, radiographers etc), systematic reviewers, statisticians and health economists was formed. A proposal was developed to undertake a systematic review and meta-analysis, which would be used to inform and build a DAM to assess potential cost-effectiveness. Probabilistic modelling was specified in the economic analysis to enable the research to explore uncertainty in the model parameter estimates. The proposal also included a value of information analysis component for the DAM in order to adequately answer the research questions. If the purpose of the research is a precursor to commissioning primary research, what we want to know from reviewing the evidence is "should more research be commissioned?" The research brief only specified a requirement for a cost-effectiveness decision model, not any VOI analyses, however, VOI techniques are necessary to help explore whether research is potentially worthwhile, and if so EVPPI can be undertaken to help define the type of research that is required. Rather than using lack of evidence as a justification for further research, VOI allows the potential value of further research to be formally assessed and understood through exploring uncertainty.

Chapter 3 detailed the development of the economic models and reported deterministic outcomes. The probabilistic outcomes are now reported,

exploring uncertainty in the parameter estimates through the cost-effectiveness planes, and then uncertainty in the decision using cost-effectiveness acceptability curves.

4.2 Probabilistic results

The three economic models (primary, recurrent and metastatic) were analysed probabilistically, using Monte Carlo simulation (2000 iterations) to determine the expected cost, expected outcomes and the expected cost-effectiveness. The costs represent the cost of the imaging scans in each strategy and the resultant (accurate and inaccurate) treatments received; while outcomes are represented in terms of QALYs gained. The cost-effectiveness of PET/CT as an add-on imaging device in pre-operative staging was assessed for: (i) primary rectal cancer, (ii) primary colon cancer, (iii) recurrent rectal cancer, (iv) recurrent colon cancer and (v) metastatic disease. The incremental costs and QALYs with their 95% confidence intervals are now reported and illustrated on cost-effectiveness planes to demonstrate the impact of parameter uncertainty on the cost and QALY outcomes for each of the analyses.

4.2.1 Primary CRC results

Primary rectal cancer

In primary rectal cancer the intervention strategy (ceCT, MRI and PET/CT) cost an additional £2,267 (95% CI £932 to £3602) per person with a QALY gain of 0.005 (95% CI -0.02 to 0.03) in comparison to the conventional strategy. This represents an incremental cost-effectiveness ratio of £431,691 per QALY gained and would not be considered cost-effective in the UK, given the NICE threshold of £20,000 to £30,000 per QALY (24). The confidence intervals for incremental costs and QALYs were calculated using the percentile method, using the 2.5 and 97.5 percentiles (4). It is important to note that the lower 95% confidence interval for incremental QALYs is negative, while the mean

value and upper confidence interval are positive. This represents uncertainty as to the existence of any QALY gains.

Figure 4.1 plots the probabilistic outcomes from the Monte Carlo simulation, illustrating the uncertainty surrounding the expected incremental costs and incremental QALYs for primary rectal cancer.
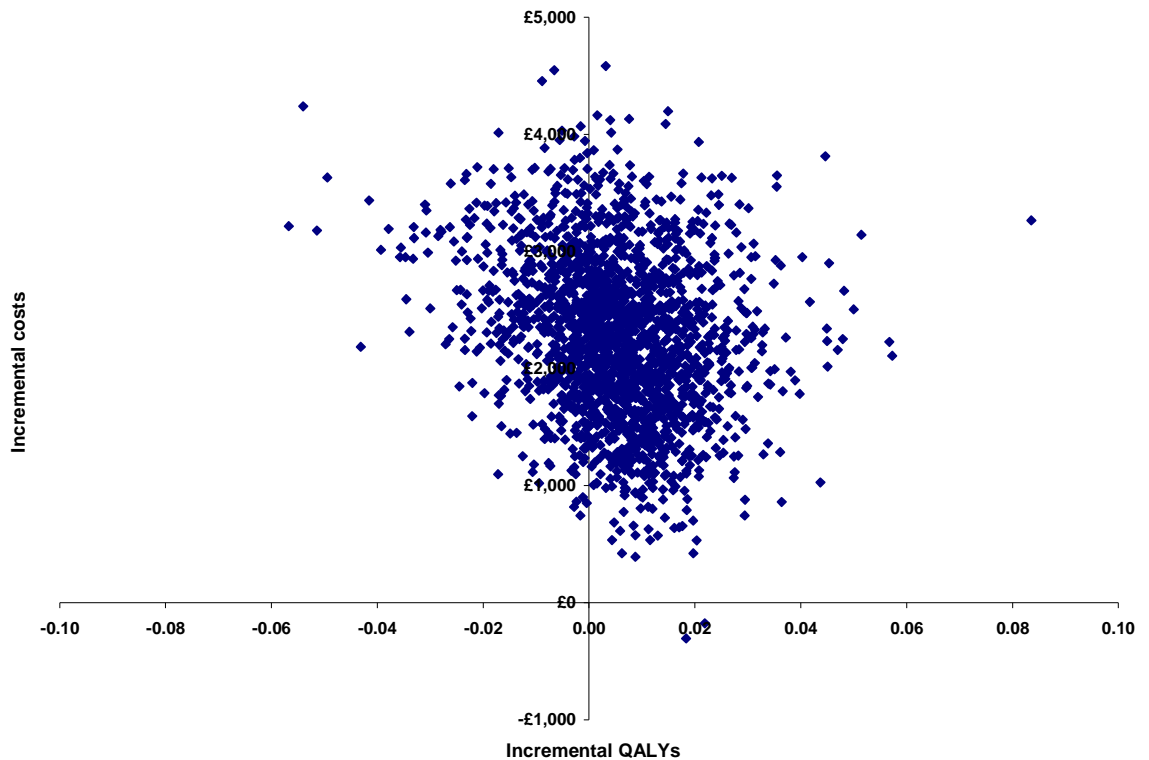

Figure 4-1: The CE plane for PET/CT in primary rectal cancer

The joint distribution of the costs and effects from the Monte Carlo simulation are plotted on the cost-effectiveness plane to demonstrate the impact of uncertainty in the model parameters on the model outcomes (expected incremental cost and effects). Some guidance exists regarding appropriate interpretation of cost-effectiveness results on the cost-effectiveness plane, and on the cost-effectiveness acceptability curves (49;70;71). Uncertainty in incremental outcomes is demonstrated when the results spread across the y-axis, representing both QALY gains (in the eastern quadrants) and QALY losses (in the western quadrants). Likewise a spread through the origin passing through the horizontal axis represents uncertainty in the incremental cost of

the intervention. The extent of the spread also indicates the extent of uncertainty.

In Figure 4-1 the horizontal plane illustrates the incremental QALYs and shows that there is considerable uncertainty regarding the existence of additional QALYs, as the spread of points pass through the origin. Negative QALYs represent outcomes in which the conventional strategy had a greater QALY gain that the intervention strategy, and therefore we are uncertain as to the existence of any QALY gains, and the extent of any gains, represented by the wide spread. The vertical plane represents incremental costs. Figure 4-1 illustrates that there is considerable uncertainty about the extent, but not the existence, of the additional expected costs. The additional costs range from approximately -£300 to £4500, however, there are only two points below the origin representing a negative incremental cost (where the conventional strategy is cheaper than the intervention strategy). Therefore we can be confident that the intervention will cost more than the conventional strategy, however we are less certain about what the additional cost will be, ranging from £900 up to £3,600 within the 95% CI.

Primary colon cancer

In primary colon cancer the intervention strategy (ceCT and PET/CT) cost an additional £2,253 (95% CI £1195 to £3310) per person with a QALY gain of 0.01 (95% CI -0.02 to 0.05) in comparison to the conventional strategy. The 95% confidence interval for the costs is similar to that in the primary rectal analysis, however the 95% CI for incremental QALYs is a little wider and again includes negative values in the lower interval, indicating considerable uncertainty in both the existence of any QALY gain, and in the extent of gain when there is a gain. These probabilistic outcomes are plotted on the cost-effectiveness plane in Figure 4-2.
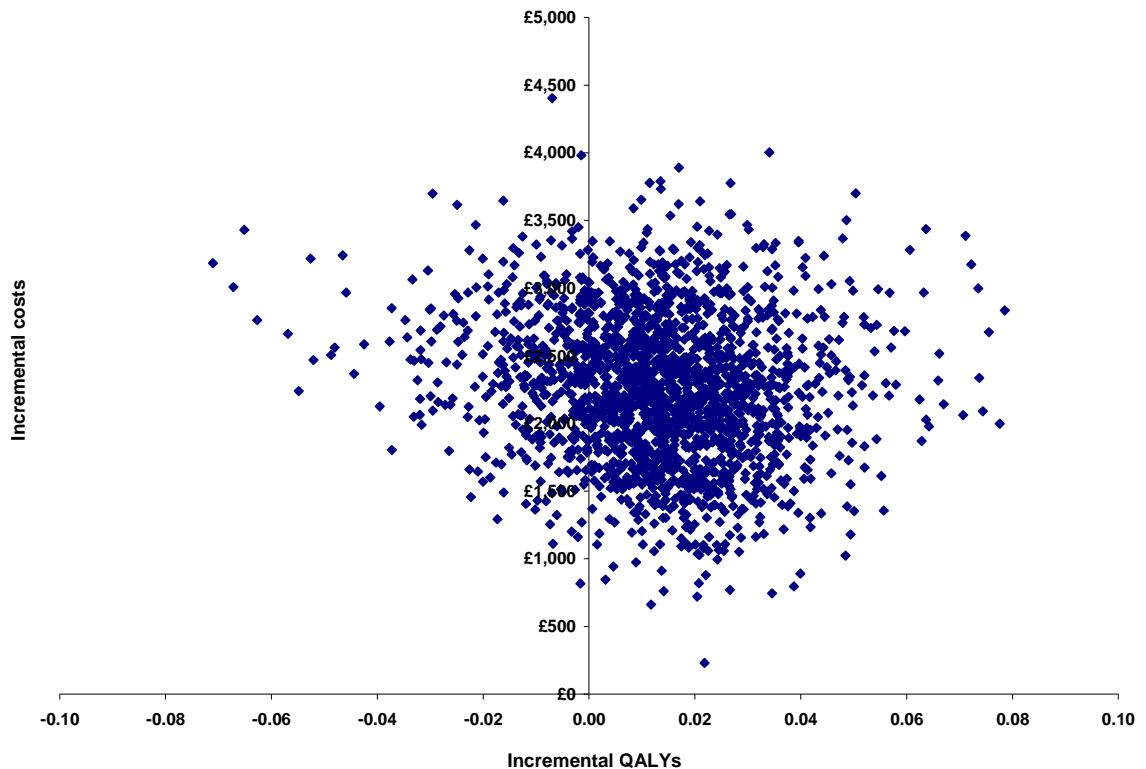
Figure 4-2: The CE plane for PET/CT in primary colon cancer

Figure 4-2 shows that none of the values for incremental cost fall below zero, indicating that we can be certain the intervention strategy will cost more than the conventional strategy, however, the extent of additional cost ranges between £220 and £4400, with 95% falling within the interval £1195 and £3310. The cost-effectiveness plane also illustrates the uncertainty in incremental QALYs which was evident from the 95% CI. The incremental QALY points pass through the origin on the horizontal plane and in general are quite spread out, illustrating the wide 95% confidence interval and lower negative value.

The cost-effectiveness plane represents the impact of the uncertainty in the model parameters on the incremental cost and QALY outcomes; however it does not tell us about uncertainty in the decision regarding cost-effectiveness. Figure 4-2 shows that some of the time the conventional strategy will dominate the intervention, when the ICER points are in the NW quadrant, however, the majority of the points fall into the NE quadrant where cost-effectiveness will be dependent on societies' willingness to pay. The cost-effectiveness acceptability curve can help explore uncertainty in this decision.

Primary Rectal Scenario: cePET/CT as a lone technology

The primary rectal scenario was undertaken to explore the potential cost-effectiveness of an advanced cePET/CT technology as a replacement for standard contrast enhanced CT, in the event that a cePET/CT technology becomes available in the UK. However, in primary rectal cancer an MRI scan will still be undertaken and therefore even the cePET/CT would be an add-on technology in rectal cancer.

The probabilistic results indicate an expected incremental cost of £975 (95% CI -£322 to £2271) and an incremental QALY gain of 0.01 (95% CI -0.01 to 0.03). The lower confidence intervals for both cost and QALYs are negative, while the mean value and upper intervals are positive, indicating considerable uncertainty in these outcomes.   Figure 4-3 illustrates these probabilistic outcomes, showing the spread of ICER outcomes from the Monte Carlo simulation. The incremental costs illustrated on the vertical axis have a wide range, from -£733 at the lowest value up to £3250 and even the lower 95% confidence interval is a negative value indicating uncertainty as to what the additional costs, if any, will be. There is also uncertainty as to the existence of additional QALYs, as these values also cross through the origin; however the range is narrow and quite tight around the mean value of 0.01.
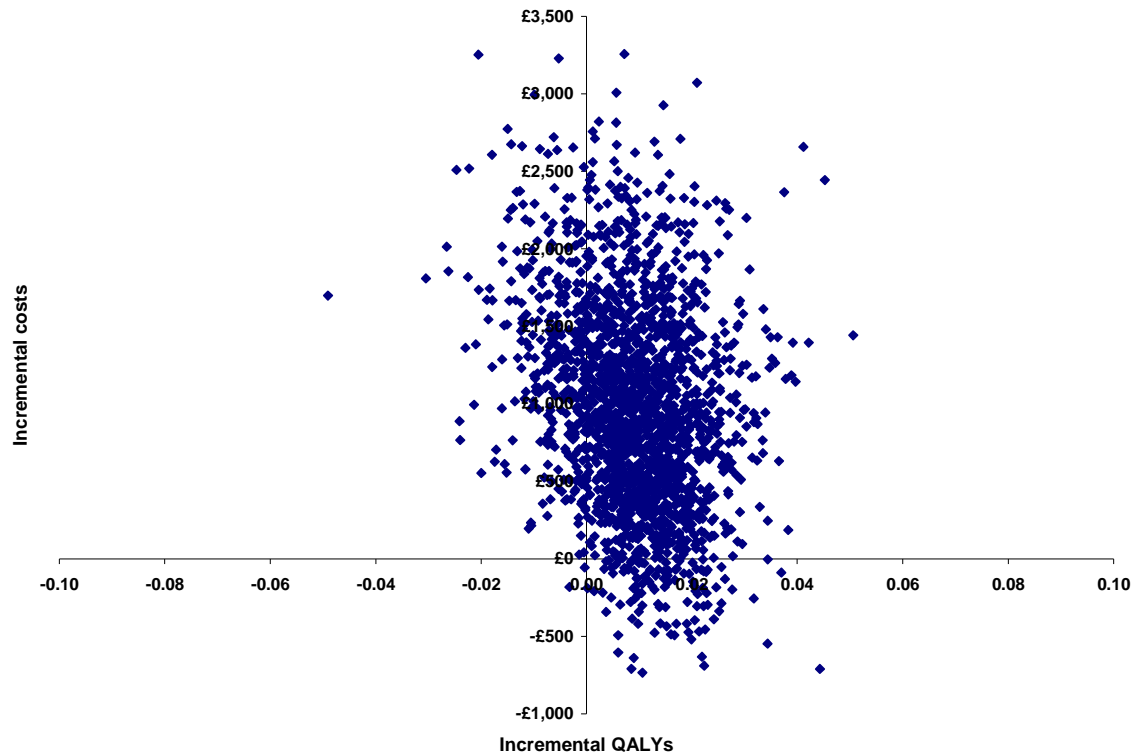
Figure 4-3: The CE plane for cePET/CT in primary rectal cancer scenario

These results indicate uncertainty in the cost and QALY outcomes from the primary rectal scenario analysis, which is due to parameter uncertainty in the model.

Primary Colon Scenario: cePET/CT as a lone technology

The scenario analysis was also undertaken for primary colon cancer where an advanced cePET/CT technology could be used as a replacement the conventional ceCT scan, in the event that a cePET/CT technology becomes available in the UK. In this scenario the ceCT is the only scan in the conventional strategy and therefore cePET/CT would be a replacement technology. Figure 4-4 plots the probabilistic outcomes on the cost-effectiveness plane.

The probabilistic results indicate an expected incremental cost of £206 (95% CI -£1476 to £1887) and an incremental QALY gain of 0.02 (95% CI -0.0024 to 0.03).
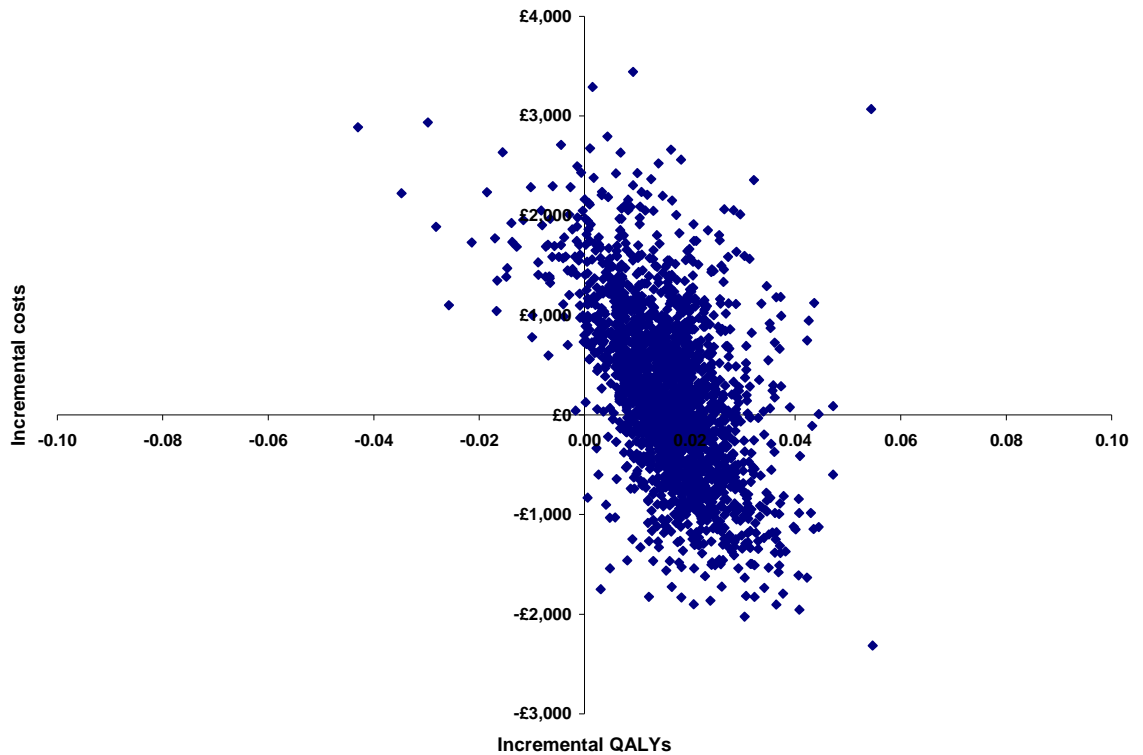
Figure 4-4: The CE plane for cePET/CT in primary colon cancer scenario

There is considerable uncertainty in whether a cePET/CT strategy would cost more or less than the conventional ceCT strategy, as shown in Figure 4-4 where the incremental cost ranges from -£2300 up to an additional £3400. The mean incremental cost of £206 is therefore a very uncertain expectation. Likewise there is uncertainty as to the existence and extent of any QALY gains, ranging from -0.04 to 0.05 QALYs, however, the 95% confidence interval is narrower around the mean, but still incorporates negative values, crossing through the origin on the horizontal axis in Figure 4-4. In this scenario, there is considerable uncertainty as the ICER values spread into three of the quadrants of the cost-effectiveness plane. The cePET/CT intervention dominates ceCT when ICER values fall into the South East quadrant representing a reduction in cost and a QALY gain, however, the North West quadrant represent ICER values where ceCT dominates PET/CT. In the North East quadrant of the cost-effectiveness plane neither technology dominates, therefore societies' willingness to pay will need to be considered to determine whether the ICER values are cost-effective in this quadrant.

4.2.2 Recurrent CRC results

Recurrent rectal cancer

In recurrent rectal cancer the intervention strategy (ceCT, MRI and PET/CT) cost an additional £712 (95% CI £185 to £1239) per person with a QALY gain of 0.03 (95% CI -0.04 to 0.11) in comparison to the conventional strategy. This represents an incremental cost-effectiveness ratio of £21,409 per QALY gained which would be considered cost-effective in the UK (24). However, it is of interest to consider the impacts of parameter uncertainty on uncertainty in the incremental cost and QALY outcomes. The confidence intervals for incremental costs and QALYs were calculated using the 2.5 and 97.5 percentiles from the Monte Carlo iterations.   The 95% CI for incremental cost ranges from approximately £200 to £1200 which is a substantial variation, however they are all positive. Figure 4-5 illustrates that only a few of the incremental cost outcomes fall below zero, with a minimum value of -£1150, however, these negative incremental costs values are outside the 95% CI.
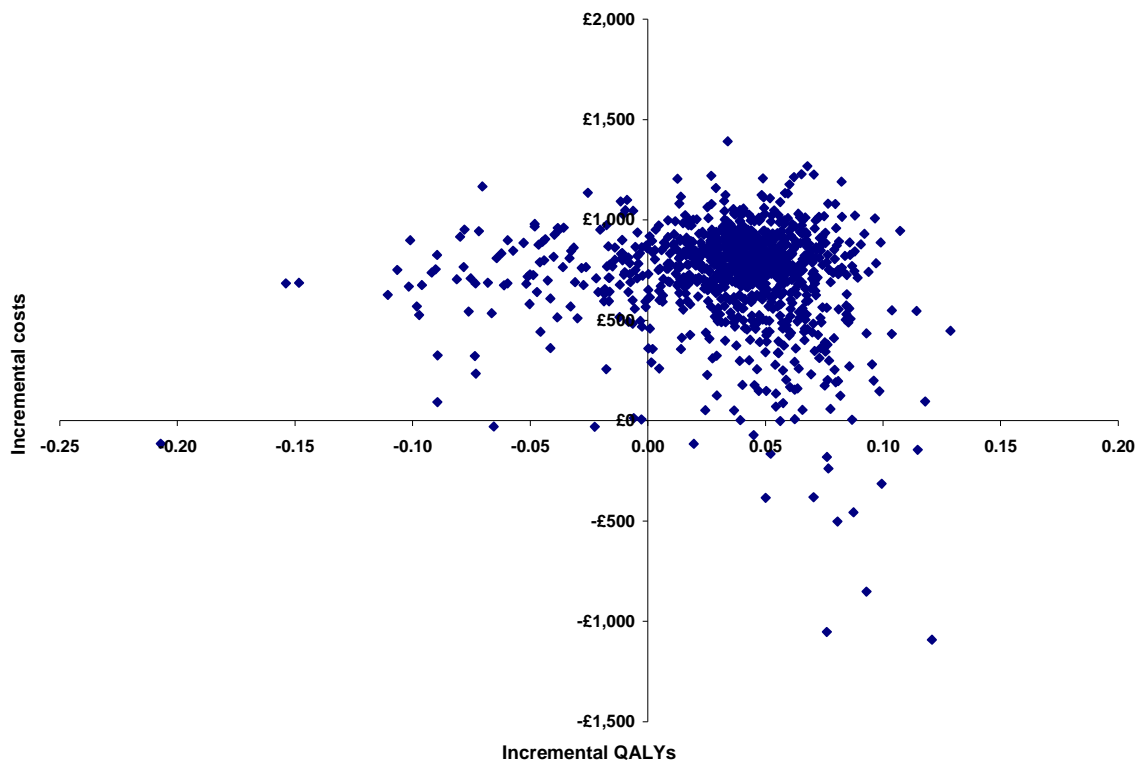

Figure 4-5: The CE plane for PET/CT in recurrent rectal cancer

The uncertainty surrounding the existence of additional QALY gains is substantial. There is wide variation around the mean point estimate of 0.03 and Figure 4-5 shows that the majority of incremental QALY values are positive, to the right of the origin on the horizontal axis, however the lower 95% CI incorporates negative QALY values up to -0.04 QALYs and there are a few outlying negative up to -0.17 QALYs. There is considerable uncertainty regarding both the existence of and the extent of any expected QALY gains, as highlighted by the spread of points along the horizontal plane crossing through the origin.

Recurrent colon

The results from the recurrent colon probabilistic analysis indicate an expected incremental cost of £866 (95% CI £562 to £1170) per person with a QALY gain of 0.14 (95% CI -0.08 to 0.36) in comparison to the conventional strategy. This translates to an ICER of approximately £6,000 per QALY gain and would be considered highly cost-effective under the typical UK definition (24).   Figure 4-6 illustrates the cost-effectiveness plane.
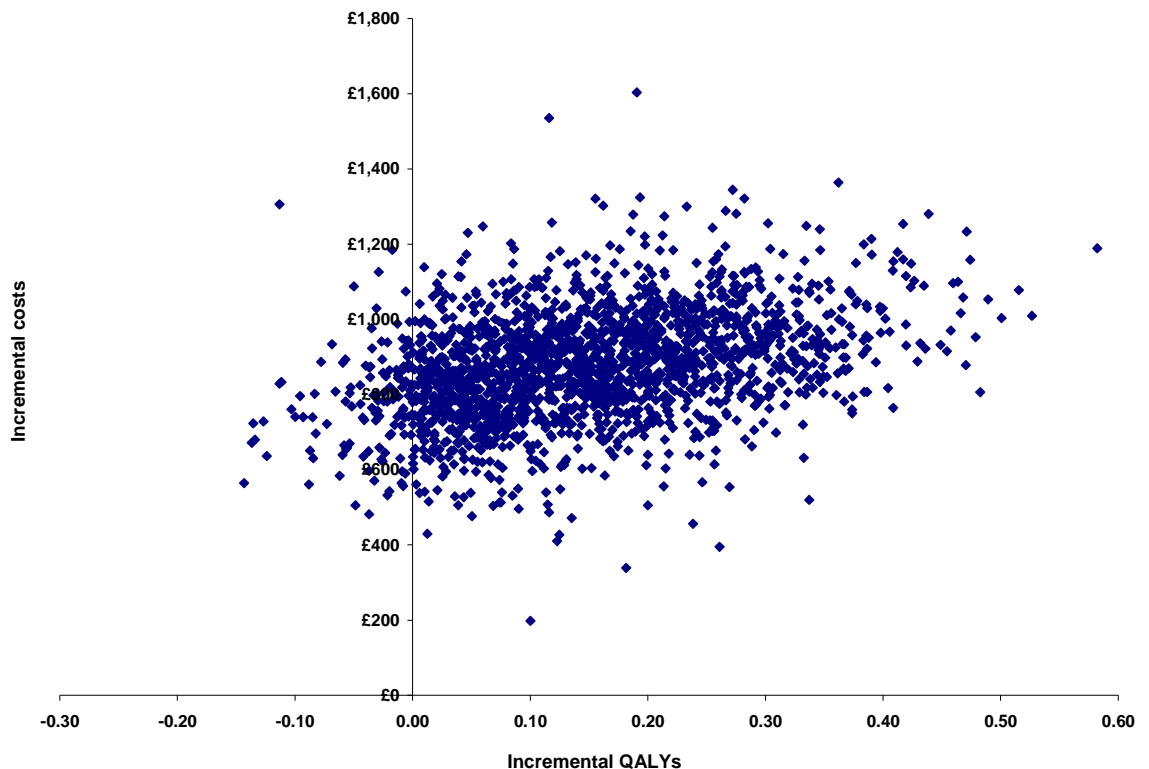
Figure 4-6: The CE plane for PET/CT in recurrent colon cancer

The expected incremental costs are all greater than zero, and the mean value has a reasonably tight 95% CI as illustrated on the vertical axis in Figure 4-6. Therefore, we can be confident that the intervention will incur additional costs and there is a little uncertainty regarding the extent of the additional costs. Figure 4-6 shows that there is greater uncertainty, as to both the existence and extent of any additional QALY gains. The 95% CI for incremental QALYs is very wide and the lower interval is negative passing through the origin on the horizontal axis, however, the majority of incremental QALYs are positive and even though the upper 95% interval is at 0.36 QALY gains, the maximum gain is as high as 0.59 QALYs. Given these values we are uncertain as to the existence of and extent of any additional QALY gains.

4.2.3 Metastatic results

With regards to the metastatic model, the intervention strategy (ceCT plus PET/CT) has an expected additional cost of £276 (95% CI -£4384 to £4937) per person with a QALY gain of 0.01 (95% CI -0.08 to 0.10) in comparison to the conventional strategy. This represents an incremental cost-effectiveness ratio of £21,434 per QALY gained which would be considered cost-effective in the UK; however, it is of importance to consider the confidence intervals and note that there is considerable uncertainty in both the incremental cost and QALY outcomes.

The results from the probabilistic analysis indicate that we are extremely uncertain as to the existence and extent of any additional cost in the intervention strategy. The 95% CI around the incremental cost expectation is very wide, crossing through the origin and incorporating negative values as low as -£4300, and at the upper level nearly £5000. Figure 4-7 illustrates these outcomes on the vertical axis of the cost-effectiveness plane.
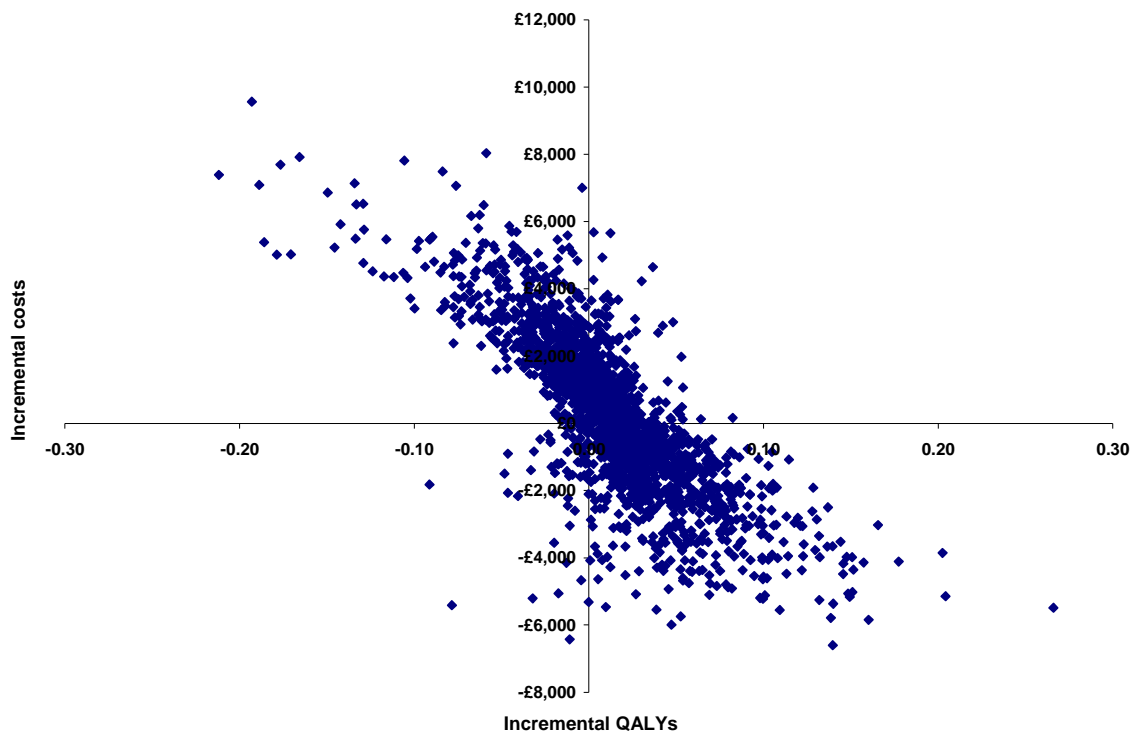
Figure 4-7: The CE plane for metastatic cancer

It can be seen that at a maximum there would be an additional cost of nearly £8000, while at the minimum the intervention strategy is approximately £7000 cheaper than the conventional strategy. Likewise, the confidence interval for QALYs is wide and the spread of values passes through the origin including a wide range of both additional and negative QALY gains.

Looking at the scatter of incremental values, about a third fall in the South Eastern quadrant, where the add-on PET/CT intervention is dominant over the conventional strategy, but another third of the incremental values fall in the North Western quadrant where it is the conventional strategy that dominates. The remaining values fall in the North East and South West quadrants, where ICERs will be present; therefore, even though the overall ICER can be considered cost-effective, the probabilistic results paint a very unclear picture. These results illustrate that there must be considerable uncertainty in the model parameters, which has impacted on the expected costs and QALY outcomes.

4.3 Decision Uncertainty

The incremental costs and QALYs and their 95% confidence intervals (2.5 & 97.5 percentiles from the distribution of Monte Carlo results) from the probabilistic sensitivity analyses were illustrated on cost-effectiveness planes to demonstrate the impact of uncertainty in the model parameters on uncertainty in the model outcomes (expected incremental cost and QALYs). The results showed that in all of the analyses, the spread of incremental QALYs passed through the origin on the horizontal axis, representing a situation where in some instances there would be a positive ICER value but in other instances the conventional strategy would dominated the intervention strategy (in the North West quadrant). If the spread of ICER values had all been contained to one quadrant, then a decision regarding cost-effectiveness would be more straight forward, yet still dependent on a willingness to pay value (λ) if ICER points fall in the North East or South West quadrants. As the spread of results involved two or more quadrants (for all of the analyses), cost-effectiveness is more complicated to determine.

As all of the analyses incorporated instances where the spread of results entered two or more quadrants of the cost-effectiveness plane, the 95% confidence intervals for the ICERs involve negative values due to some of results being dominated by one of the strategies. In such cases, the 95% CI for the ICER may appear back to front with the lower confidence interval representing a higher ICER value than the upper confidence interval (40;168). Therefore 95% confidence intervals were not presented for the ICER values. In such cases careful consideration must be given to the spread of ICER values and the interpretation of the confidence intervals (40); and decision uncertainty (as to whether the intervention is cost-effective or not) is better represented on cost-effectiveness acceptability curves (CEACs) than in confidence intervals (4).

Having presented the impact of uncertainty in the model parameters on the incremental cost and incremental QALY outcomes, cost-effectiveness acceptability curves will now be used to present decision uncertainty for each

of the analyses, illustrating the probability of each intervention being cost-effective at different willingness to pay thresholds.

4.3.1 Primary CRC decision uncertainty

The aim of calculating a CEAC is to summarise and illustrate graphically the probability that a particular intervention is the optimal choice over a wide range of values for the ceiling ratio (λ) (7). Fenwick and colleagues (11) demonstrate this process and provide a guide to interpreting the CEAC (7;71).

Primary rectal cancer

The cost-effectiveness acceptability curve (CEAC) for primary rectal cancer is illustrated in Figure 4-8, representing the uncertainty in the cost-effectiveness estimate for primary rectal cancer. The CEAC shows the probability that PET/CT is cost-effective as an add-on imaging device in comparison to ceCT and MRI at different values for the maximum acceptable cost-effectiveness ratio (λ).
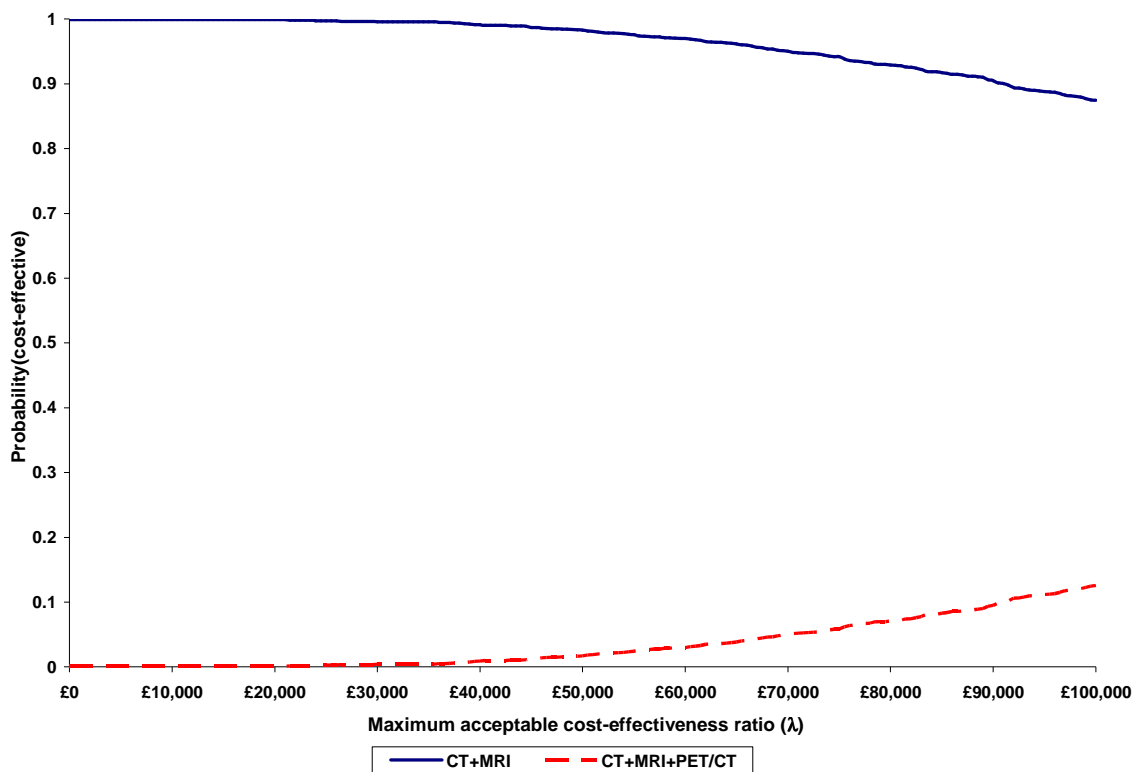
Figure 4-8: The CEAC for primary rectal cancer

At a monetary threshold of £100,000 the probability that the PET/CT intervention will be cost-effective is less than 20%. Within the usual UK range of values for the maximum acceptable cost-effectiveness ratio (λ) (£20,000 to £30,000 per QALY) the CEAC illustrates that the conventional strategy (CT and MRI) has approximately 100% probability of being cost-effective and the PET/CT intervention has a 0% probability of being cost-effective.

The CEAC illustrates that there is little uncertainty in the cost-effectiveness decision. The cost-effectiveness plane for primary rectal cancer in Figure 4-1 illustrated considerable uncertainty in the existence and extent of expected incremental QALYs and some uncertainty in expected incremental costs, however, the CEAC illustrates that this parameter uncertainty does not translate into decision uncertainty.

Primary colon cancer

The uncertainty in the cost-effectiveness estimate for primary colon cancer is illustrated on a cost-effectiveness acceptability curve in Figure 4-9.
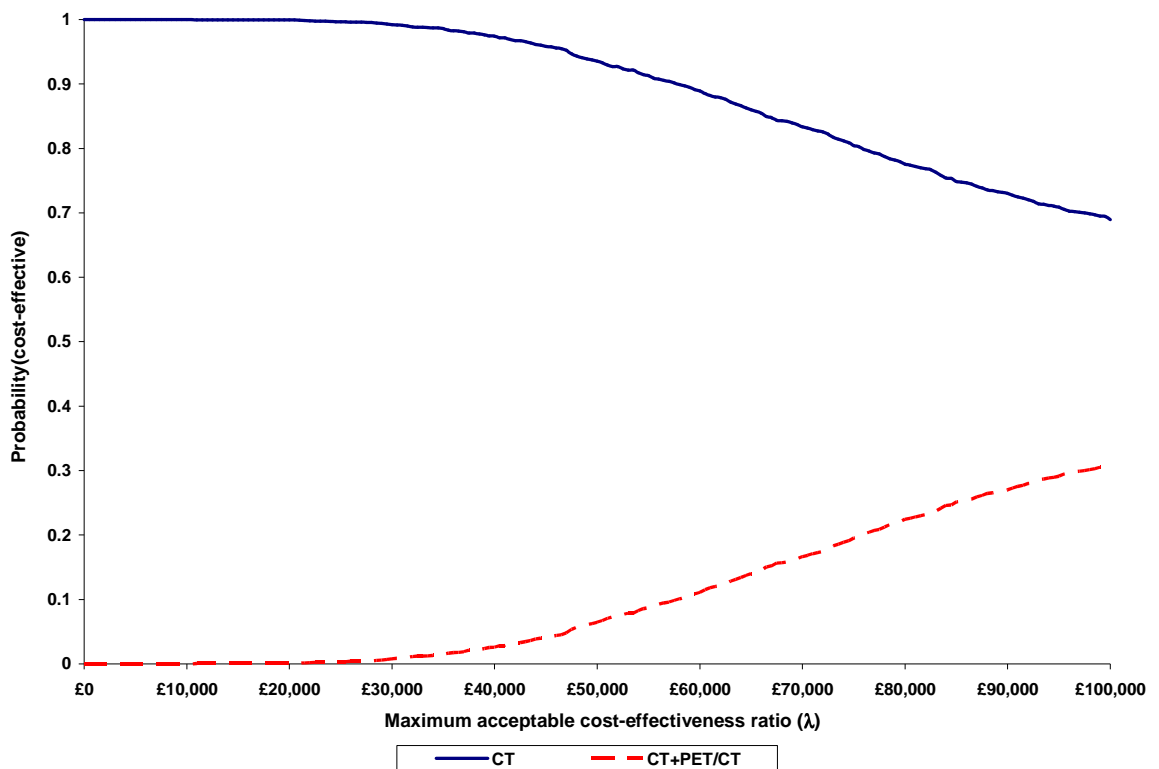
Figure 4-9: The CEAC for primary colon cancer

The CEAC shows that at a monetary threshold of £100,000/QALY the probability that the PET/CT intervention will be cost-effective is approximately 30%. At a threshold of £30,000 per QALY, the CEAC illustrates that the probability that PET/CT will be cost-effective is approximately 1%. At this threshold the probability that the conventional ceCT strategy will be cost-effective is approximately 99%.

The cost-effectiveness plane for the primary colon cancer PSA results in Figure 4-2 illustrated considerable uncertainty in the existence and extent of expected incremental QALYs and some uncertainty in the extent of additional costs; however, this did not translate to decision uncertainty. At a monetary threshold of £30,000 per QALY we can be 99% certain that the conventional strategy (ceCT) is cost-effective.

Primary Rectal Scenario: cePET/CT as a lone technology

The uncertainty around the cost-effectiveness estimate for the primary rectal cancer scenario is illustrated on the CEAC in Figure 4-10.
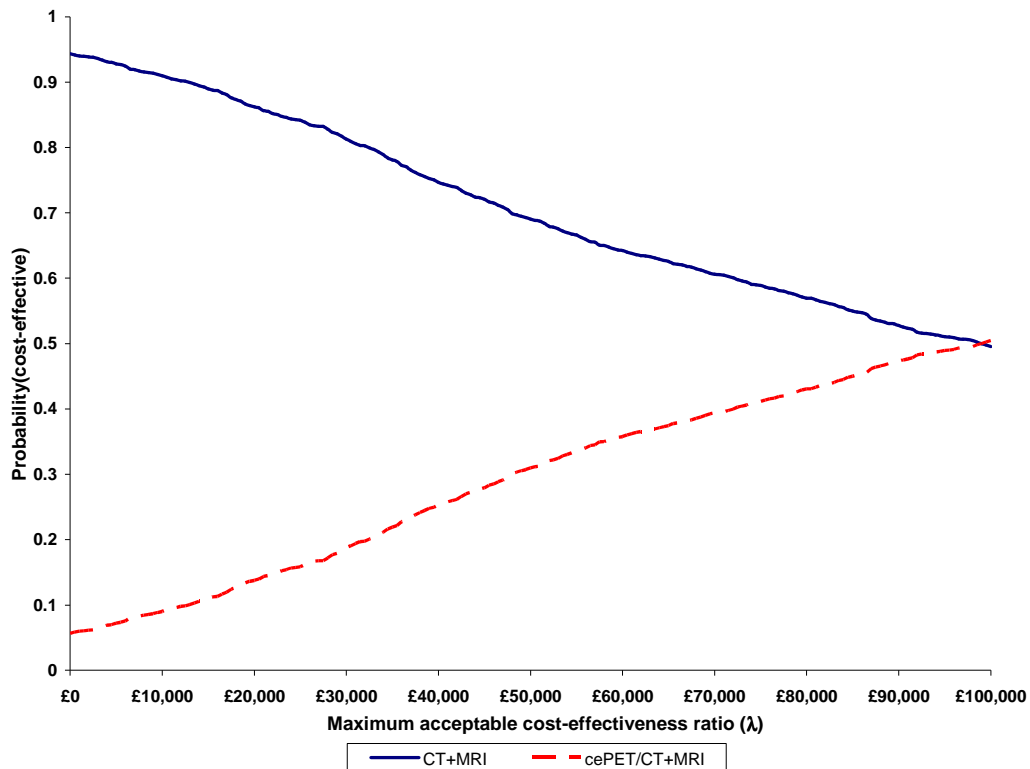

Figure 4-10: The CEAC for primary rectal scenario

The CEAC shows that at a monetary threshold of £30,000/QALY the probability of cePET/CT and MRI being cost-effective in comparison to ceCT and MRI is less than 20%.   As the monetary threshold increases, the probability of the intervention strategy (cePET/CT + MRI) being cost-effective increases, however even at a threshold of £60,000/QALY which is double the UK recommended threshold (24), the probability that the intervention strategy will be cost-effective is only 35%. The willingness to pay threshold would need to be greater than £100,000 per QALY for the cePET/CT and MRI strategy to be the optimal intervention.

The probabilistic sensitivity analysis for the primary rectal scenario resulted in a scatter of ICER vales over three quadrants of the cost-effectiveness plane, indicating uncertainty in the existence of both incremental QALYs and incremental costs. This parameter uncertainty translates into some decision uncertainty, as we can be reasonably certain (over 80%) that at a monetary threshold of £30,000 per QALY the conventional strategy would cost-effective.

Primary Colon Scenario: cePET/CT as a lone technology

The cost-effectiveness acceptability curve for the primary colon cancer scenario is illustrated in Figure 4-11. The figure shows that at a monetary threshold of £30,000/QALY there is a 63% probability of cePET/CT being cost-effective in comparison to ceCT. Note that the two curves cross at a willingness to pay threshold of approximately £12,800, which corresponds to the ICER point estimate value. If society is willing to pay greater than £12,800 per QALY then cePET/CT is likely to be the optimal strategy, given current information.
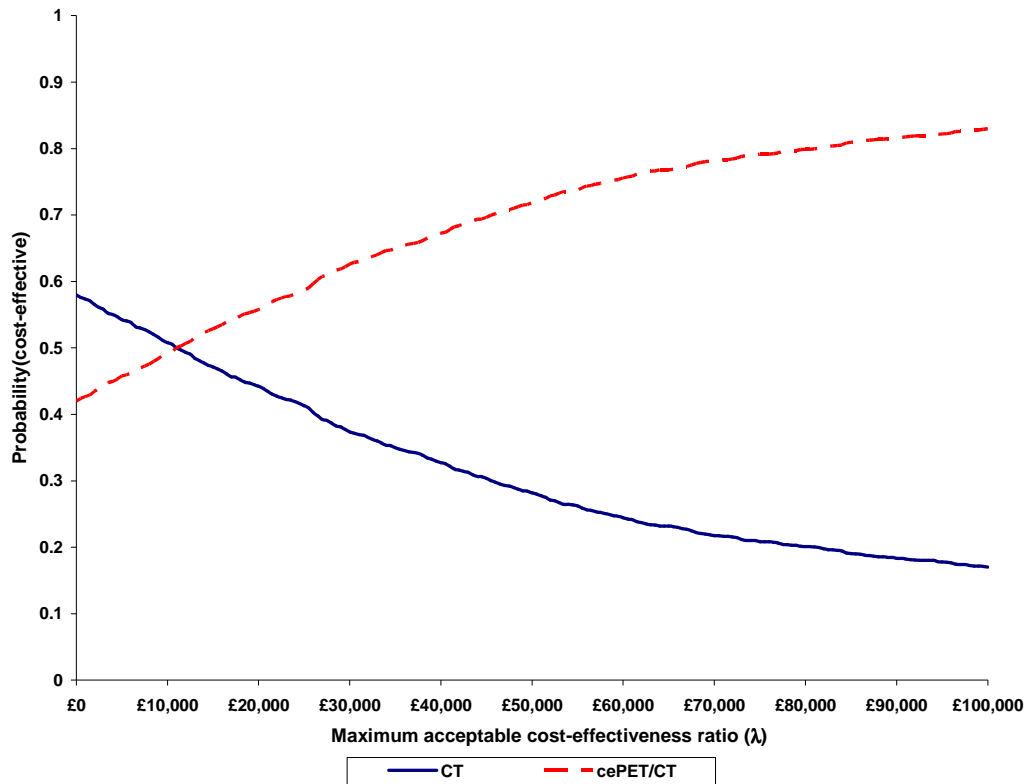
Figure 4-11: The CEAC for primary colon scenario

The cost-effectiveness plane in Figure 4-4 illustrated uncertainty in the existence and extent of both incremental costs and QALYs, with a spread of ICER values that crossed three quadrants of the plane. Some of this uncertainty translates into decision uncertainty as can be seen on the CEAC in Figure 4-11. If society is willing to pay greater than £12,800 per QALY then cePET/CT is likely to be the optimal strategy, however there is still considerable uncertainty surrounding this decision. At £30,000 per QALY the probability that the cePET/CT strategy is cost-effectiveness is 63%, and even at a threshold of £60,000/QALY the probability is only 75%. Therefore, there remains some decision uncertainty as to whether cePET/CT is the optimal strategy in the primary colon scenario.

4.3.2 Recurrent CRC decision uncertainty

Recurrent rectal

The cost-effectiveness acceptability curve, plotted in Figure 4-12, illustrates the uncertainty in the cost-effectiveness estimate for recurrent rectal cancer.
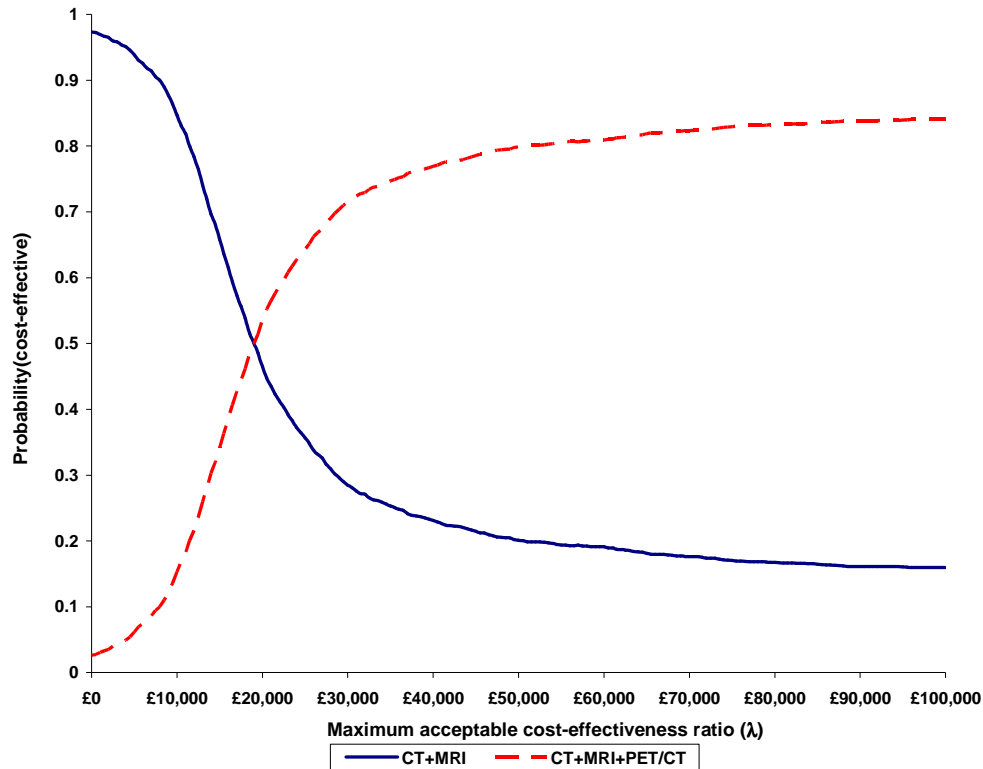
Figure 4-12: The CEAC for recurrent rectal cancer

The figure shows that at a monetary threshold of less than £21,000/QALY there is a greater probability that the conventional (ceCT and MRI) strategy is the most cost-effective, but at a monetary threshold of greater than £21,000/QALY the intervention (add-on PET/CT) strategy has a greater probability of being the most cost-effective strategy. The two curves cross at the mean ICER estimate of £21,000 per QALY, and therefore, at the £30,000 per QALY threshold recommended by NICE (24) the CEAC indicates approximately 70% probability that the intervention strategy will be cost-effective in comparison to the conventional strategy.

The cost-effectiveness plane in figure 4-5 illustrated considerable uncertainty regarding both the existence of and the extent of any expected QALY gains, and some uncertainty in the extent of additional costs. This parameter uncertainty has translated into some decision uncertainty.

Recurrent colon

Uncertainty in the cost-effectiveness estimate at different monetary thresholds is illustrated for recurrent colon cancer in Figure 4-13.
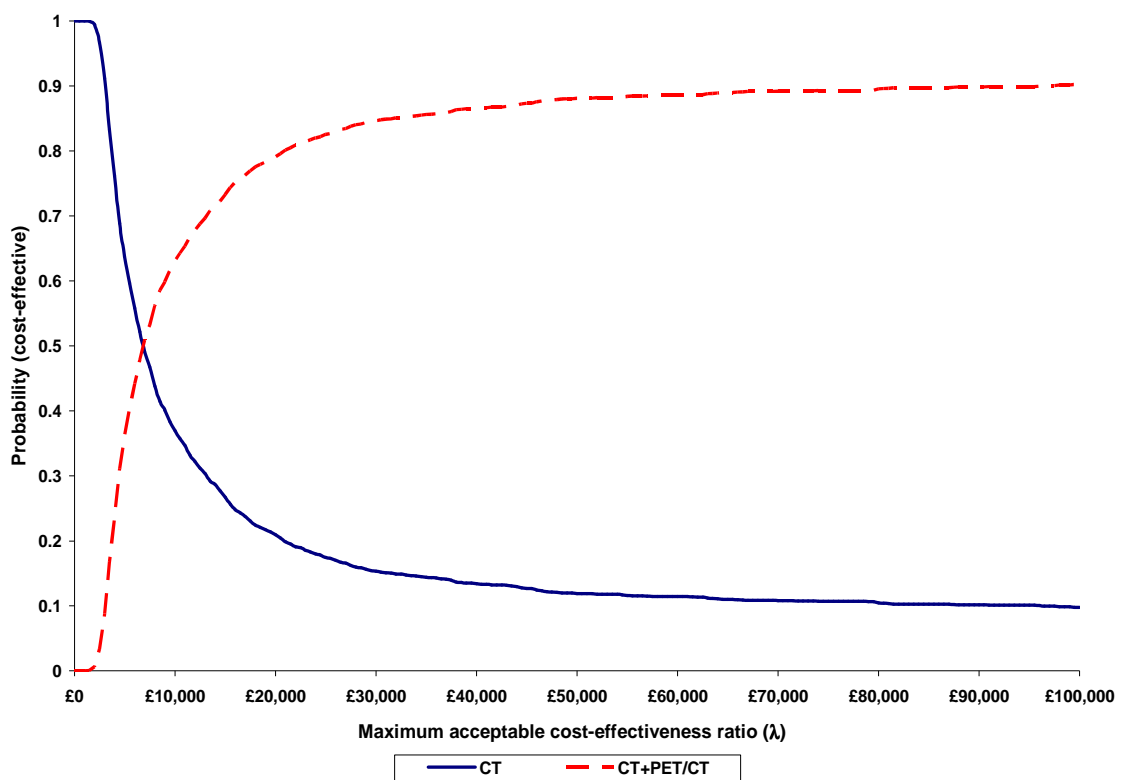
Figure 4-13: The CEAC for recurrent colon cancer

At a monetary threshold of less than £6,000 per QALY, the conventional strategy is the most cost-effective; while at a threshold above £6,000 per QALY the PET/CT strategy has the greatest probability of being cost-effective. At the monetary threshold of £30,000/QALY there is an 85% probability that the intervention will be cost-effective, versus a 15% probability that the conventional strategy will be cost-effective. Given that the cost-effectiveness plane in Figure 4-6 illustrated considerable uncertainty in the incremental

QALY estimate and some uncertainty in cost, this parameter uncertainty has translated into only a little uncertainty regarding cost-effectiveness. We can be confident that the PET/CT intervention is cost-effective at a monetary threshold of £30,000/QALy, but only with a probability of 85%.

4.3.3 Metastatic decision uncertainty

With regards to metastatic colorectal cancer, Figure 4-7 illustrated that the ICER values from the PSA fall into all four quadrant of the cost-effectiveness plane, indicating uncertain as to the existence and extent of any additional cost or QALYs in the intervention strategy. The cost-effectiveness acceptability curve, in Figure 4-14, can help provide some insight regarding the decision uncertainty.
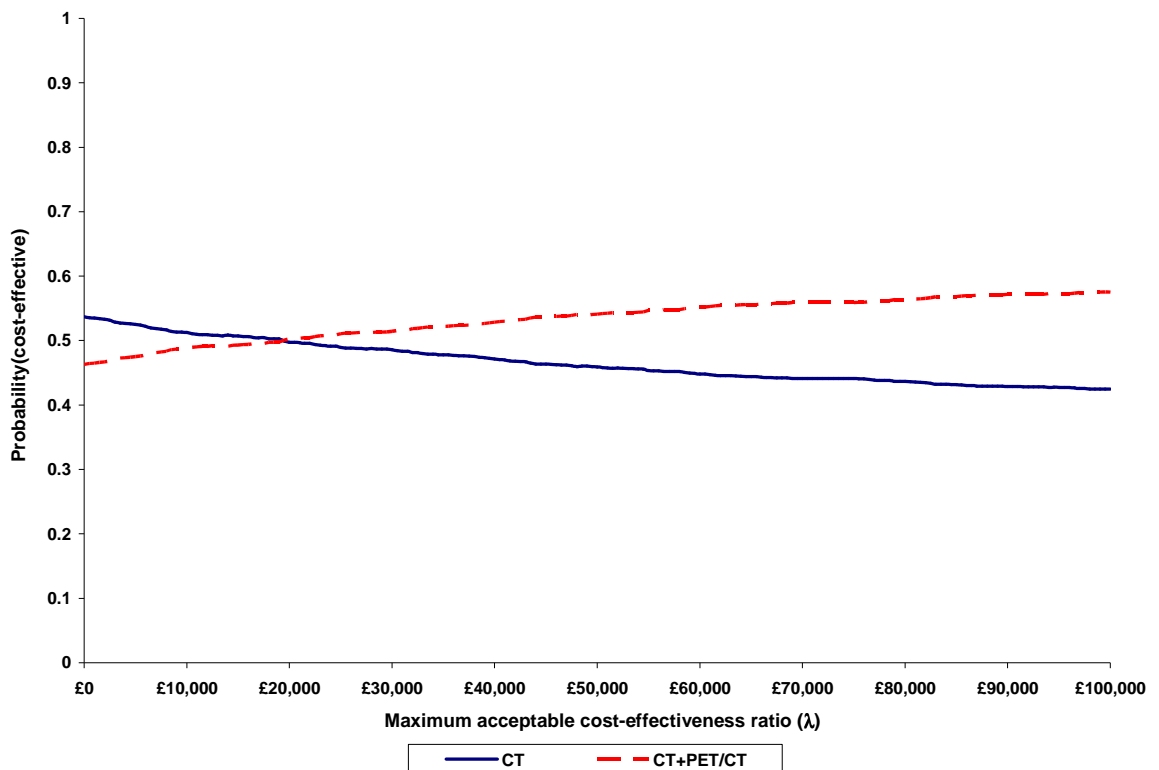
Figure 4-14: The CEAC for metastatic cancer

Figure 4-14 illustrates the considerable uncertainty surrounding the cost-effectiveness of the add-on PET/CT strategy. At a monetary threshold of £21,000 per QALY the probability that the PET/CT intervention will be cost-

effective equals approximately 50% as is the probability that CT will be cost-effective. Beyond this threshold the probability of cost-effectiveness for the intervention strategy increases slightly, but never reaches beyond 57%, even at a monetary threshold of £100,000/QALY. At the £30,000 per QALY threshold recommended in the UK (24) the CEAC indicates that the PET/CT intervention has a slightly greater probability of being cost-effective (52%) than the conventional strategy (48%). This confirms that there is considerable uncertainty as to which intervention is cost-effective in the metastatic analysis, regardless of societies' willingness to pay threshold.

4.4 Value in further research: EVPI & EVPPI

Having presented the PSA results and explored the decision uncertainty for each of the analyses, illustrating the probability of each intervention being cost-effective at different ceiling ratios, it is important to consider two questions: given the current evidence and decision uncertainty, should the technology be adopted (or not)?, and is further research required to help support this decision (4)? If these questions are not answered, decision makers may struggle to interpret the cost-effectiveness results, particularly regarding how to make the decision to adopt or reject the add-on PET/CT technology given the uncertainty. In answering these questions the researchers can help inform funders and decision makers on how to interpret the results for each CRC patient group and the appropriate next steps to take, based on the outcomes from the PSA.

In the PET/CT case study, the results from the probabilistic analysis indicate that given current evidence add-on PET/CT is not cost-effective in primary CRC and should not be adopted, but it may be cost-effective in recurrent and metastatic disease. This summary does not formally capture the decision uncertainty for each of the analyses. Therefore, value of information analysis can be used to formally evaluate whether further research is necessary to support the decision to adopt or reject the technology.

4.4.1 Primary CRC: Value of Information

Primary Rectal cancer

The expected value of perfect information (EVPI) was calculated using the probabilities of cost-effectiveness for each intervention which were generated in the CEAC calculation, over a range of monetary thresholds in increments of £500 from zero to £100,000 per QALY gain. The EVPI results showed that at a monetary threshold of £30,000/QALY the expected value of perfect information per decision in primary rectal cancer is £1.30.

This EVPI is very low; however, when the decision uncertainty demonstrated in the CEAC in Figure 4-8 is considered, it is not surprising. At the monetary threshold of £30,000/QALY the probability that PET/CT would be cost-effective was only 0.4% and therefore the conventional strategy (CT+MRI) was the optimal strategy, with a 99.6% probability. The EVPI is equivalent to the opportunity loss from choosing the optimal (conventional) strategy in the instances that the optimal strategy would have been 'wrong' given perfect information. Given that there was only a 0.4% probability that the conventional strategy is the wrong decision, it makes logical sense that the opportunity loss of choosing the conventional strategy is low. Even at a greater 'willingness to pay' threshold of £100,000/QALY, the probability that PET/CT is cost effective was only 20%. Therefore at this higher threshold there is a 20% probability that choosing the conventional strategy is the wrong decision, and this wrong decision resulted in an average opportunity loss, or value of perfect information, of £60 per decision.

It is important to represent what this EVPI per decision represents in terms of the relevant patient population who would benefit from the additional information.   In order to determine the population value of EVPI, the patient population over the lifetime of the technology must be considered.   The annual incidence of primary rectal cancer in the UK was estimated to be 13,315 (94) and a technology timeframe of two years was applied (and discounted at 3.5% (24), i.e. PET/CT in its current form will be considered as an add-on for imaging for two years. This (relatively short) timeframe was been determined in part by the continual development and upgrading of PET/CT, such that the estimates that we have for DTA are likely to change outside of this timeframe.   Therefore the effective population of primary rectal cancer patients who would potentially benefit from additional information was estimated to be 26,180. Figure 4-15 details the results from the expected value of perfect information analysis (EVPI) at a population level.

Figure 4-15: EVPI for primary rectal cancer – population level

At a willingness to pay threshold of £30,000/QALY the expected value of perfect information for the population is approximately £34,000. If a monetary threshold of £100,000 per QALY were applied, then the population value of further research would be £2.3 million, however, the UK recommended threshold is £20,000 to £30,000 per QALY (24). This value can be considered to be the maximum society should be willing to pay for further research, and therefore, it is not worthwhile seeking additional information for PET/CT for primary rectal cancer. The cost of undertaking further research activity, would out way this population value of £34,000 and therefore it is not worthwhile.

Primary colon cancer

Similar to the primary rectal results, the EVPI per decision for primary colon cancer was found to be very low. At a willingness to pay threshold of £30,000/QALY the expected value of perfect information per decision was £1.60. This low EVPI value can be explained by considering the (lack of) decision uncertainty demonstrated by the CEAC in Figure 4-9. The CEAC for primary colon cancer showed that at a monetary threshold of £30,000 per

QALY, there was only a 1 % probability that the PET/CT intervention would be cost-effective, and therefore the conventional strategy was the optimal choice with a 99% probability of being cost-effective.   Given that there was only a 1% probability that the conventional strategy is the wrong decision, it makes logical sense that the opportunity loss of choosing the conventional strategy under perfect information is low, and therefore the value of further research is low.

In order to determine the overall population value of EVPI for primary colon cancer the annual incidence of primary colon cancer was determined to be 21,574 (94) and the PET/CT technology timeframe was estimated to be two years, discounted at 3.5% (24) to give an effective population of 42,418 primary colon patients. As with the primary rectal evaluation, the short technology timeframe was determined to reflect the continual development and upgrading of PET/CT technology, such that the current estimates for DTA are likely to change outside of this timeframe. Figure 4-16 details the results from the expected value of perfect information analysis (EVPI) at a population level.
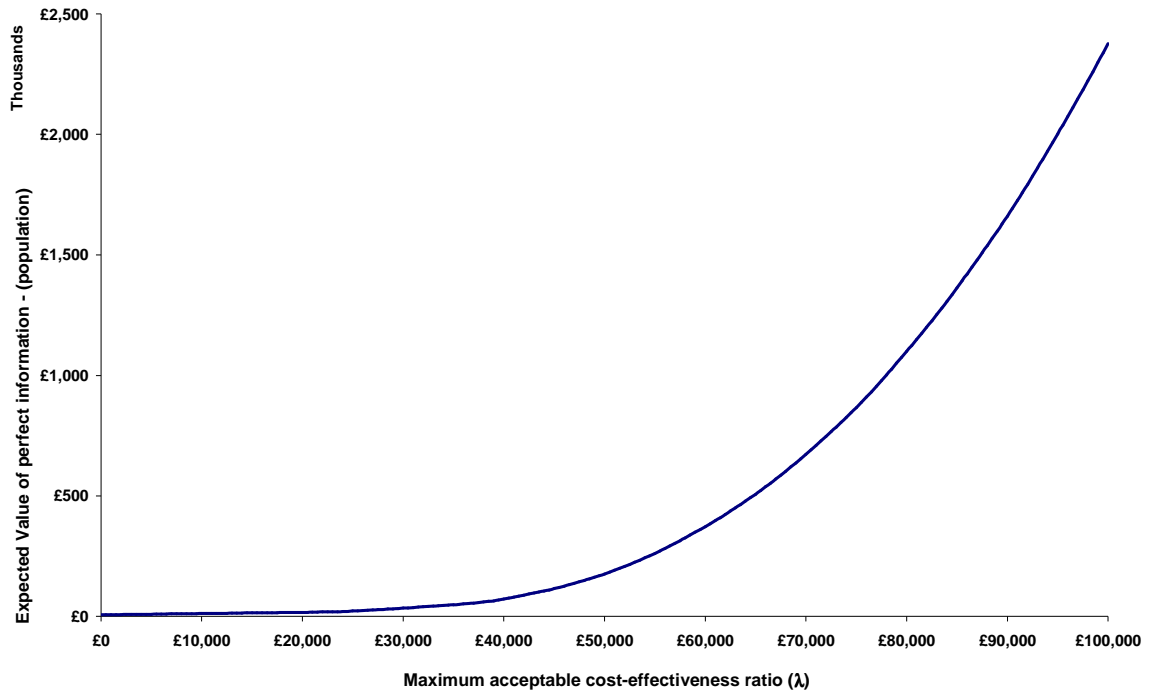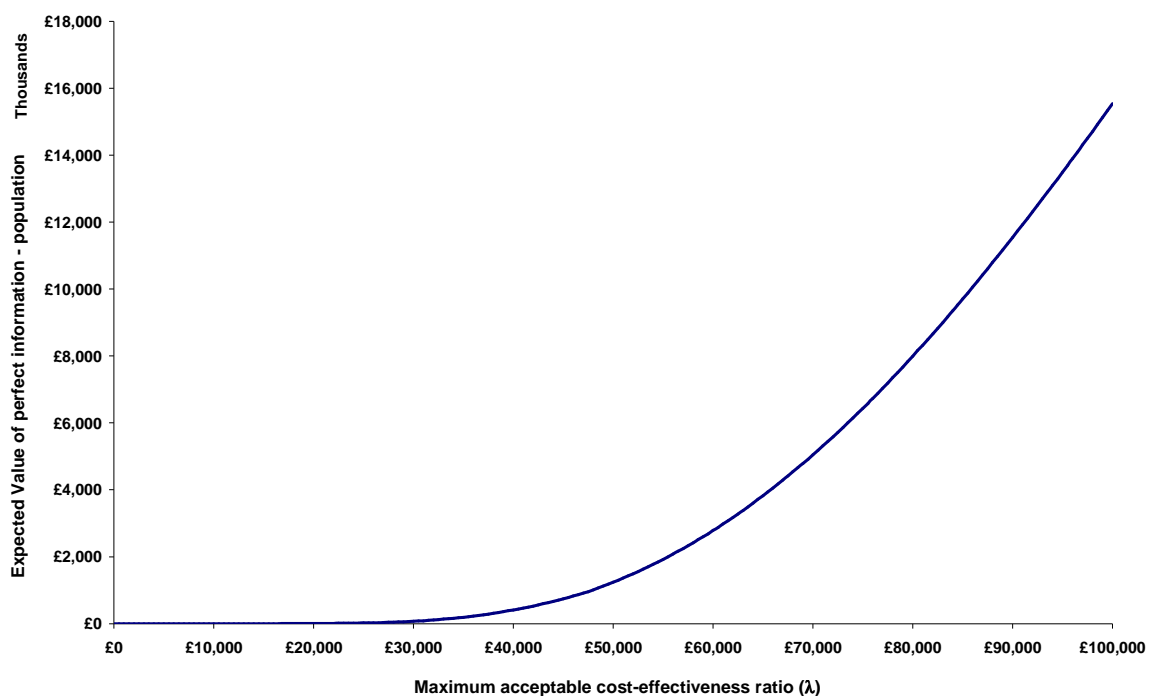

Figure 4-16: EVPI for primary colon cancer – population level

At a willingness to pay threshold of £30,000/QALY the expected value of perfect information for the population is approximately £70,000. Thus, it is unlikely that research will cost less than this value, and therefore it is not worthwhile seeking additional information for PET/CT for primary colon cancer.

Scenario analysis: cePET/CT in primary rectal cancer

The scenario analysis for primary rectal cancer was undertaken to explore the potential cost-effectiveness of an advanced cePET/CT technology along with MRI as an alternative to the conventional contrast enhanced CT and MRI scans. The CEAC in Figure 4-10 indicated that while the conventional strategy was the cost-effective option, there was some decision uncertainty around this choice.

The results from the population level expected value of perfect information analysis (EVPI) are presented in Figure 4-17, based on the same patient population (26,180) and technology timeframe as that specified for the baseline primary rectal evaluation.
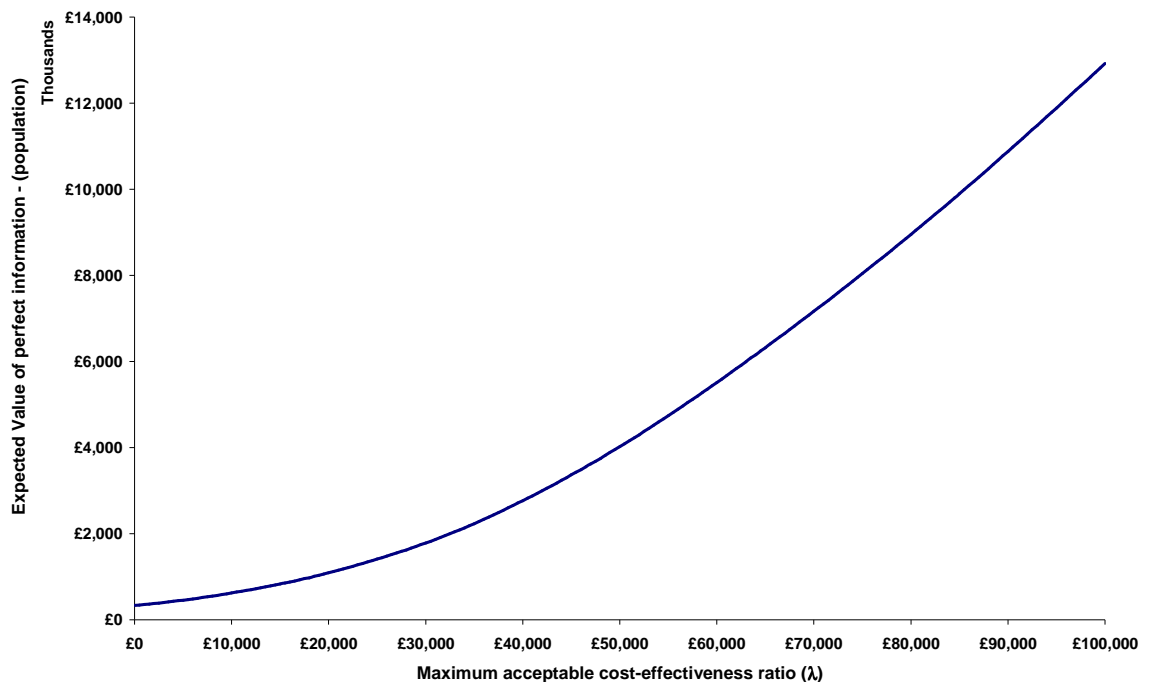
Figure 4-17: EVPI for primary rectal cancer scenario – population level

The EVPI results indicate that at a willingness to pay threshold of £30,000 per QALY, the expected value of perfect information per decision is £68, which translates to a population EVPI of £1.7million. Therefore it is potentially worthwhile undertaking further research to explore whether cePET/CT can (and is available to) be used as a replacement for ceCT in primary rectal cancer, if the research cost was less than £1.7 million.

Scenario analysis: cePET/CT in primary colon cancer

The EVPI analysis for the primary colon cancer scenario was undertaken based on the same patient population (42,418) and technology timeframe as that specified for the baseline primary colon evaluation. The results are presented in Figure 4-18.
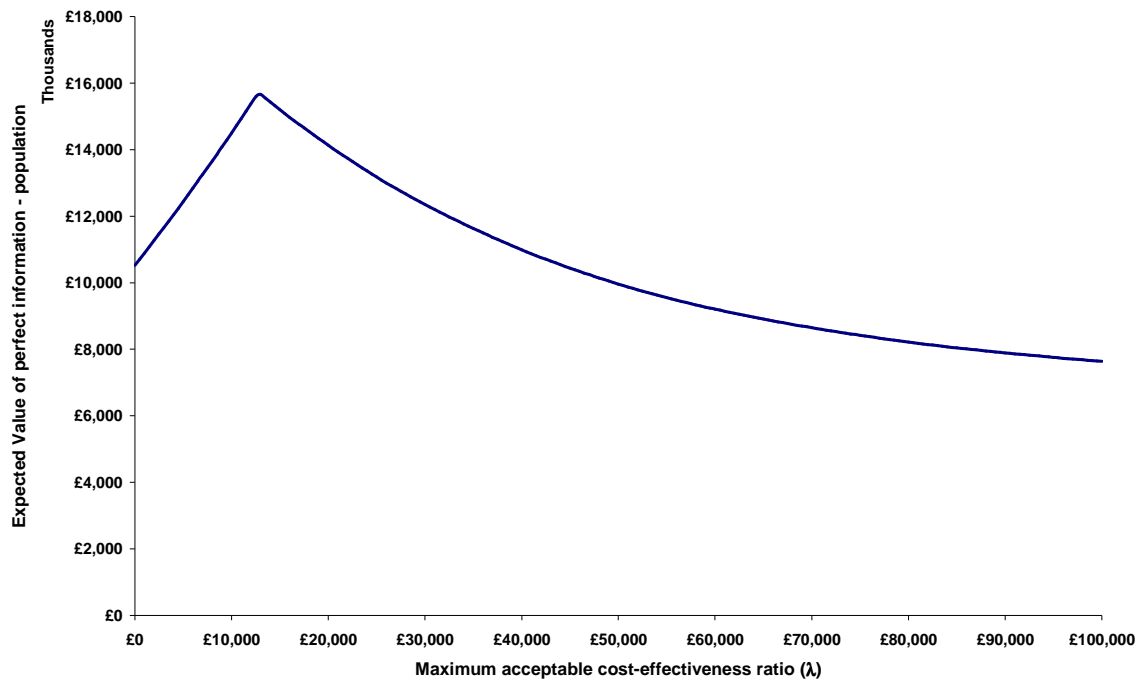
Figure 4-18: EVPI for primary colon cancer scenario – population level

The CEAC for the primary colon scenario found that if society is willing to pay greater than £12,800 per QALY then cePET/CT is likely to be the optimal strategy. However, there was still considerable uncertainty surrounding this

decision, and at £30,000 per QALY the probability that the cePET/CT strategy would be cost-effectiveness was 63%. This decision uncertainty is reflected in the value of further research. The EVPI analysis resulted in an expected value of perfect information per decision of £290 at a willingness to pay threshold of £30,000 per QALY. When scaled up to reflected this value to the patient population, further research is worth £12.3million as detailed in Figure 4-18.

Figure 4-18 shows the kink in the EVPI curve, which occurs at the point where the ceiling ratio is equivalent to the ICER. For the primary colon scenario this value is £12,800. Beyond this point the EVPI is still high but diminishing. Given the high population EVPI, it can be concluded that it is potentially worthwhile undertaking further research to reduce the uncertainty surrounding the question of whether cePET/CT can be used as a replacement for ceCT in primary colon cancer.

4.4.2 Recurrent CRC

Recurrent Rectal cancer

As with the Primary model analyses, the EVPI for recurrent CRC was calculated using the probabilities of cost-effectiveness for each intervention which were generated in the CEAC calculation, over a range of monetary thresholds from zero to £100,000 per QALY gain.

The EVPI results for recurrent rectal cancer report that at a monetary threshold of £30,000/QALY the expected value of perfect information per decision is £316. To understand this value it is useful to consider the probabilistic results and the decision uncertainty presented for recurrent rectal cancer in Figure 4-12. The results showed that at a monetary threshold above £21,000 per QALY it is most likely that the intervention (PET/CT) strategy will be cost-effective, but at a monetary threshold of £30,000/QALY there is only a 70% probability of cost-effectiveness. Given that there is a 30% probability that the optimal (intervention) strategy will be the wrong decision, the EVPI is equivalent to the opportunity loss from choosing the optimal (intervention) strategy in the 30% of instances that the optimal strategy would have been 'wrong' given, perfect information. At a ceiling ratio of £30,000/QALY the expected value of further research is £316 per decision.

It is important to represent the EVPI per decision in terms of the relevant patient population who would benefit from the additional information.   In order to determine the effective recurrent rectal cancer population an annual incidence of recurrent rectal cancer of 9,054 cases was estimated (derived using the annual incidence of rectal cancer (94) and applying a 70% probability of recurrence (150) and a 2.8% death rate prior to recurrence) along with a technology timeframe of two years, to derive an effective population of 17,802. As noted previously, the two year timeframe was determined in part by the continual development and upgrading of PET/CT, such that the estimates for DTA are likely to change outside of this timeframe. The population level EVPI results are illustrated in Figure 4-19.
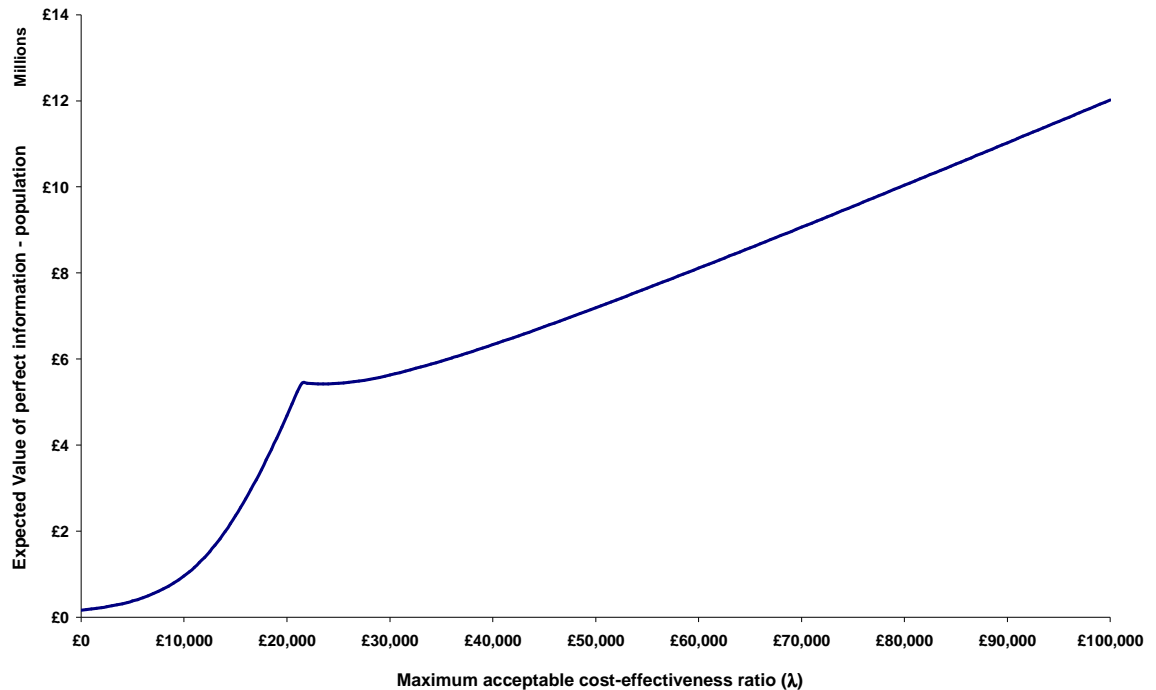
Figure 4-19: EVPI for recurrent rectal cancer – population level

The EVPI per decision of £316 translated to population level EVPI of approximately £5.6 million, at a monetary threshold of £30,000/QALY. It is likely that further research would cost less than £5million, and therefore, the EVPI results indicate that it is potentially worthwhile collecting more information about the use of PET/CT for staging in recurrent rectal cancer.

Having established that further research is likely to be worthwhile in recurrent rectal cancer, it is useful to consider what type of research is required. 'Further research' does not necessarily mean that a large scale, randomised controlled trial is required. Rather, the type of research will depend on the different parameters that require further information. For example, utility values can be collected alongside a large scale randomised trial, but if they are the sole source of uncertainty, then it would be much more efficient (and cheaper) to determine utility values from an observational study. Large scale clinical trials are only necessary for clinical information. Likewise further information regarding disease prevalence for a model could be derived by accessing and analysing routine data sources rather than undertaking a prospective clinical trial. To define the type, and possibly the scale, of the further research that is required (to reduce uncertainty in the cost-

effectiveness decision), researchers need to consider what is driving the uncertainty. What parameters are driving the uncertainty in the cost-effectiveness model and would therefore add the most value through further information? Expected value of perfect parameter information (EVPPI) is used to identify parameters for which more precise estimates would be most valuable.

In the case of recurrent rectal cancer, EVPPI was undertaken to explore which groups of parameters would add most value through further research. The various parameter inputs to the model were considered and six parameter groups were deemed to be of potential value in gaining further information: Prevalence, Diagnostic test accuracy, Five year overall survival, Cost of PET/CT, Cost of treating metastases and Utility values for the disease states.

The parameter groups are summarised in Table 4-1. With regards to prevalence the relevant parameters were those for establishing the probability of no recurrence, recurrence (local and metastatic), and for those who have recurrence: curable and incurable recurrence. Likewise the DTA group included the sensitivity and specificity parameters for the relevant tests.

With regards to the cost group, PET/CT is a relatively new technology in colorectal cancer and there is no UK reference cost information for PET/CT. Additionally there is uncertainty as to the resource use and professional time involved and therefore it may well be worthwhile collecting further information on the cost of PET/CT in colorectal cancer. Likewise, the cost of metastatic surgery and palliative care were considered to have some uncertainty in terms of resource use and were therefore included to explore whether or not there is any expected value of perfect information in these parameters.

Table 4-1: EVPPI parameter groups and parameters

Parameter Groups	Parameters
Prevalence	no recurrence
	recurrence
	recurrence curable
	recurrence incurable
Diagnostic Test Accuracy	ceCT sensitivity & specificity
	MRI sensitivity & specificity
	PET/CT sensitivity & specificity
5 year overall survival	no recurrence
	recurrence curable
	recurrence incurable
Cost PET/CT	cost PET/CT
Cost metastases	cost metastatic surgery
	cost palliative care
Utilities (disease state)	no recurrence
	recurrence curable
	recurrence incurable
Trial information	Diagnostic Test Accuracy group
	cost PET/CT
	Utilities group

Given these six parameter groups there is only one group which would require a randomised trial to gain further information, and that is the DTA group. If a randomised trial was undertaken to determine DTA, the trial could also gather information on quality of life for the disease states and the cost of PET/CT without requiring additional duration for longer term follow-up. Therefore, a seventh 'Trial information' group was established which considered the DTA, disease state utilities and cost of PET/CT parameters as a group of their own for the EVPPI. The EVPPI was run using 1000 x 1000 iterations for each of the seven parameter groups, using a monetary threshold (λ) of £30,000/QALY.

The EVPPI analysis reports outcomes in terms of the value per decision, but it is important to consider the EVPPI in terms of the relevant patient population who would benefit from the additional information. The population EVPPI was

based on the same patient population (17,802) and technology timeframe as that specified for the recurrent rectal EVPI calculation, and the results are presented in Figure 4-20.
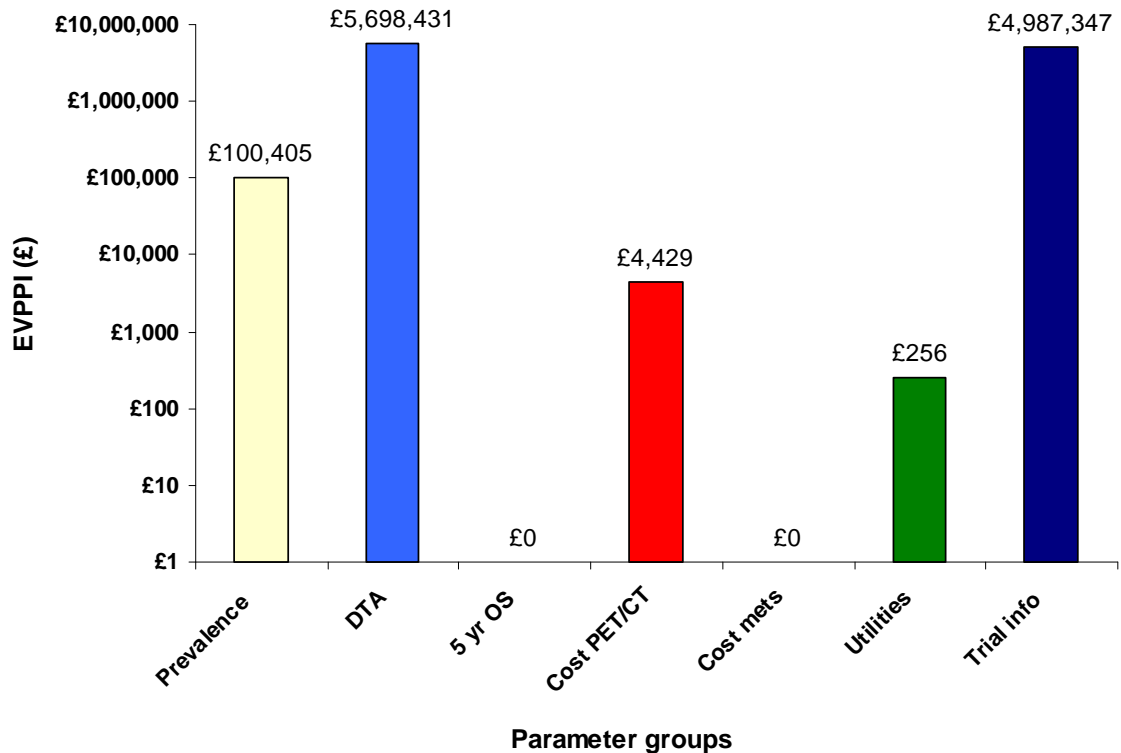
Figure 4-20: EVPPI for recurrent rectal cancer – population level (λ=£30,000)

The population EVPPI values differed substantially between the seven parameter groups, and therefore they are presented on a log scale in Figure 4-20 so as to adequately illustrate the different values. The analysis found that there was zero value in undertaking further research in the five years overall survival and cost of metastases parameter groups. In the utilities group the value of additional information per decision was very low at 1 pence (£0.0144), so even when this non-zero value is scaled up by the population to give a population EVPPI of £256, the cost of undertaking research would exceed this value.   Likewise, the value of further research for the cost PET/CT parameter was very low at 25 pence (£0.25) per decision and £4,429 at the population level. The value of perfect parameter information for the prevalence group was worth £5.64 per decision, translating to £100,000 at the population level.

It may or may not be worthwhile undertaking further research on prevalence parameters, depending on if the cost of research was lower than £100,000. The only parameter groups of any real value in the analysis were the DTA and the Trial information groups. Clearly it is the diagnostic test accuracy parameters which are driving uncertainty in the model, leading to uncertainty in the cost-effectiveness decision.

The trial information group (which combined the DTA, utilities and cost of PET/CT parameters) had an EVPPI similar to the DTA group, and considering the value for the cost and utilities groups, the EVPPI is clearly being driven by the DTA parameters in this group. It is interesting to note that the cost of PET/CT and utilities parameters appear to have a reducing impact on the value, lowering the combined Trial information EVPPI (£4.9 million) to less than that of DTA alone (£5.6 million). As described by Briggs and colleagues (4), individual parameters when considered in isolation do not resolve the same way as they do when they are joined as a group. The EVPPI for the individual parameters do not sum to the decision EVPI, and equally, the EVPPI for a group of parameters is not the sum of the individual EVPPIs (4). Therefore, there is no reason why the various parameter groups should act in an additive fashion when combined. This is evident in the Trial Info group. There is no particular reasoning as to why this group of parameters would have an additive or negative effect when combined, however, in this case the utilities and cost of PET/CT parameters have an effect of reducing uncertainty, resulting in a lower EVPPI value than when the DTA parameters are valued alone. Also, it is perfectly reasonable for EVPPI on all the individual parameters to be zero, but as a group they may be substantial.

In general the EVPPI for the various parameter groups are in line with expectation; all being very low with the exception of the DTA parameters. In populating the decision model with DTA evidence from the systematic review, the evidence was found to be limited and what did exist was found to be inadequate and subject to reporting bias for all stages of CRC disease. Therefore, it is unsurprising that it is this set of parameters that is driving uncertainty in the model, and driving the expected value in further research to help reduce that uncertainty. Given the results of the EVPPI, the type of

further research that would be required is likely to be a trial to evaluate the DTA characteristics for both PET/CT and MRI technologies. Such a trial would not necessarily be an RCT; however, the existing DTA studies (identified in the systematic review) derived their evidence mostly from retrospective studies in university or oncology hospital settings. Most of the studies used consecutive samples and only a few had comparators with no details provided regarding the use of blinding. Therefore an RCT would provide some stronger evidence regarding DTA especially if it had an appropriate comparator, was blinded etc. Additionally, any further research on DTA would not necessarily require a long-term duration, as the EVPPI showed that there is little value in further research for the survival parameters. A short enough duration to determine unbiased DTA characteristics would suffice, and could possibly also be used to derive mean utilities for the disease states.

Recurrent Colon Cancer

The EVPI results for recurrent colon cancer are now reported. The expected value of information analysis was undertaken to explore whether it is potentially worthwhile collecting more information about the use of PET/CT for recurrent colon cancer.

 At a willingness to pay threshold of £30,000/QALY the expected value of perfect information per decision is £178. The recurrent colon cancer analysis determined an ICER of £6000 per QALY, however despite this being considerably below the £30,000/QALY ceiling ratio, the CEAC in Figure 4-13 illustrated that this intervention strategy (ceCT plus PET/CT) was only cost-effective with an 85% probability. Therefore, at the £30,000/QALY threshold there remains some decision uncertainty, and this is reflected in the EVPI value of £178 per decision. Figure 4-21 displays the population level EVPI results.
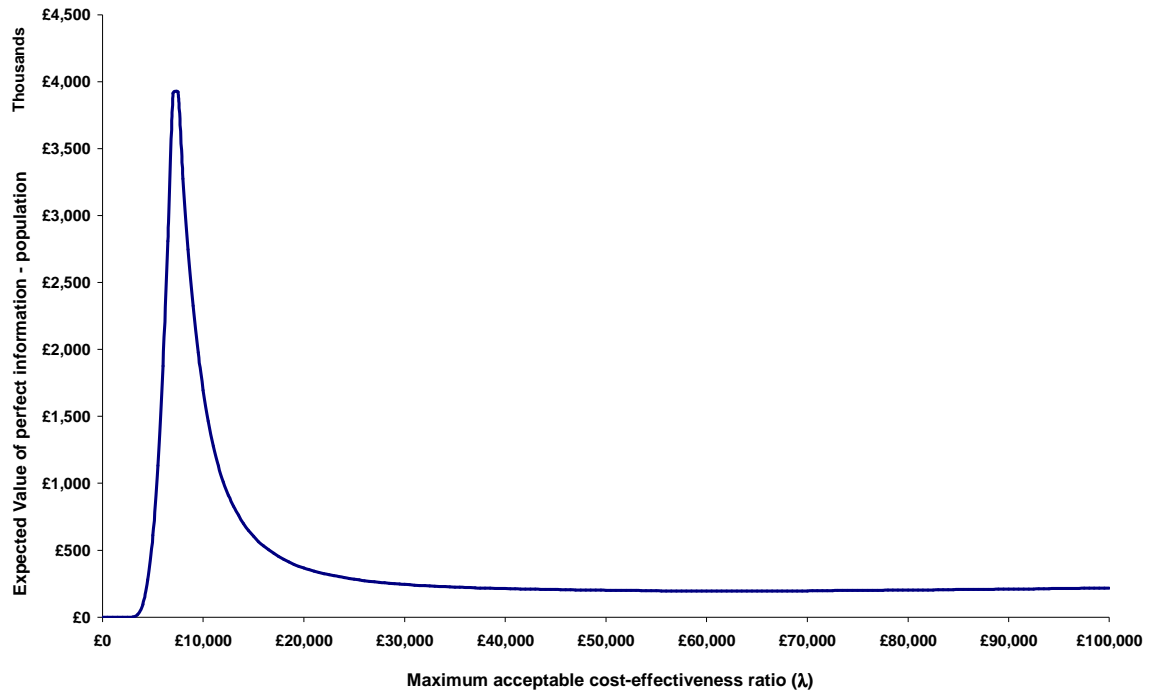
Figure 4-21: EVPI for recurrent colon cancer – population level

The overall population value of EVPI was based on an annual incidence of 14,670 cases of recurrent colon cancer, derived in the same manner as for the recurrent rectal evaluation (using the annual incidence of colon cancer (94) and applying a 70% probability of recurrence (150) and a 2.8% death rate prior to recurrence). As with the previous EVPI analyses, a timeframe of two years was applied to the PET/CT technology to represent its current lifespan, out with which the DTA estimates are likely to change reflecting the continual development and upgrading of PET/CT technology. The effective population of recurrent colon patients for which further information may be required, was determined to be 28,845.

Figure 4-21 illustrates that at a willingness to pay threshold of £30,000/QALY the expected value of perfect information for the population is approximately £5.1 million. It is likely that further research would cost less than £5million, and thus it is potentially worthwhile seeking additional information for PET/CT in recurrent colon cancer.

Having established that further research is likely to be worthwhile in recurrent colon cancer, EVPPI is undertaken to explore what parameters are driving the

uncertainty in the cost-effectiveness analysis. Establishing what parameter information would add the most value will help determine what type of 'further research' is required.   The same parameter groups which were determined to be possible sources of uncertainty for the recurrent rectal analysis were used for the recurrent colon EVPPI analysis. The seven parameter groups identified were: Prevalence, Diagnostic test accuracy, Five year overall survival, Cost of PET/CT, Cost of treating metastases, Utility values for the disease states and a Trial Information group. Table 4-1 details the specific parameters within each group, and these are the same for colon recurrence, except for the DTA group which only includes ceCT and PET/CT, as MRI scans are not used to stage colon patients.

The EVPPI analysis reported results in terms of the value per decision, and these are presented in terms of the relevant patient population who would benefit from the additional information. The population EVPPI was based on the same patient population (28,845) and technology timeframe as that specified for the recurrent colon EVPI calculation, and the results are presented in Figure 4-22. The results are presented on the log scale, for consistency with the recurrent rectal EVPPI results in Figure 4-20.

The analysis found that there was zero value in undertaking further research in the majority of the parameter groups, except for DTA and Trial information. In consideration of the EVPPI results for these parameter groups from the recurrent rectal analysis, this is not surprising. The recurrent rectal EVPPI also found zero value in cost of treating metastases and overall survival parameter groups, and in the Utilities, cost of PET/CT and Prevalence groups the EVPPI was extremely low. It is clear that in the recurrent CRC model, these parameters may have some uncertainty surrounding them, but not enough to warrant any further information. The only parameter groups of any value in the colon analysis were the DTA and the Trial information groups, with per decision values of £155 and £150 respectively. Clearly it is the diagnostic test accuracy parameters which are driving uncertainty in the model, leading to uncertainty in the cost-effectiveness decision.
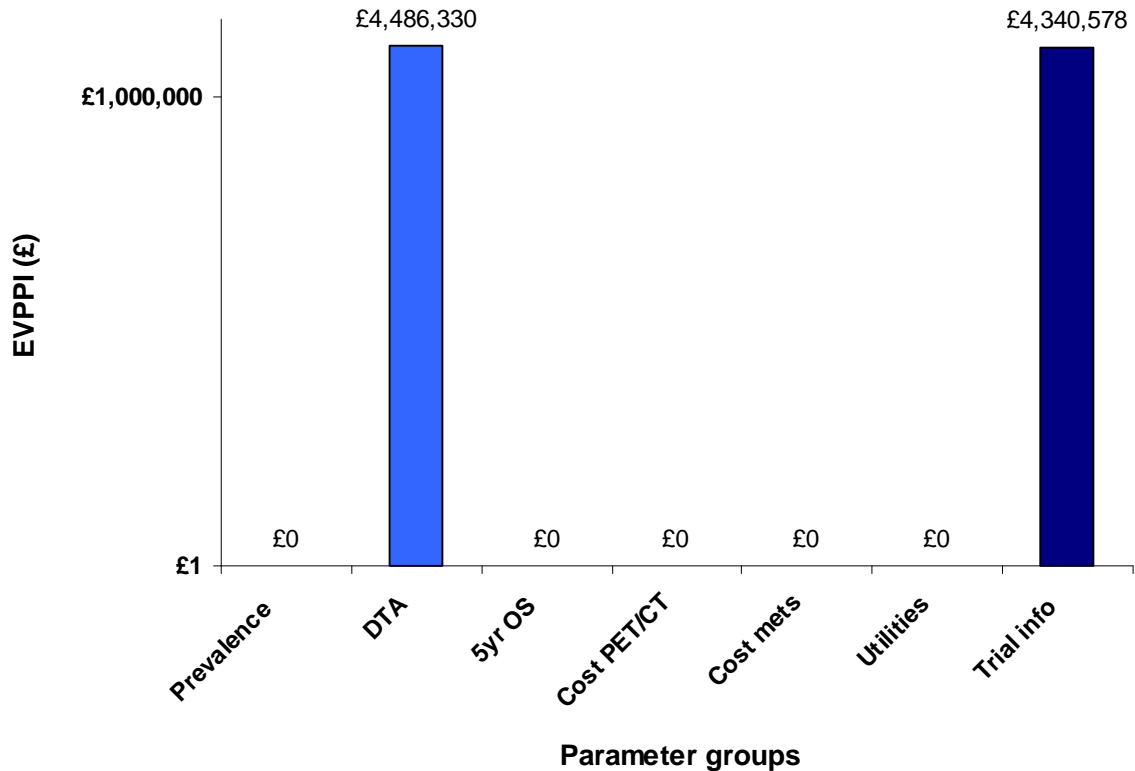
Figure 4-22: EVPPI for recurrent colon cancer – population level (λ=£30,000)

As in the rectal analysis, the Trial information group (which combined the DTA, utilities and cost of PET/CT parameters) had an EVPPI similar to the DTA group, and again the cost of PET/CT and utilities parameters appear to have a reducing impact on the EVPPI (even though they had no value in their own separate group evaluations), lowering the combined Trial information EVPPI (£4.3million) to less than that of DTA alone (£4.4million). As described by Briggs et al. (4), it is perfectly reasonable for EVPPI on all the individual parameters to be zero, but as a group they may be substantial. In the case of recurrent (colon and rectal) cancer the utilities and cost of PET/CT parameters have zero individual EVPPI, but when combined they have an effect of reducing uncertainty, resulting in a lower Trial Info EVPPI value than when the DTA parameters are valued alone. The results of the EVPPI for recurrent colon cancer indicate that a short term DTA study could help evaluate the diagnostic test characteristics for both PET/CT and ceCT technologies, and could possibly also gather information on patient quality of life for the disease states and establish the routine cost of PET/CT.

4.4.3 Metastatic

As with the Primary and Recurrent model analyses, the EVPI for metastatic CRC was calculated over a range of monetary thresholds from zero to £100,000 per QALY gain.

The EVPI analysis was undertaken to determine whether there is value in further research, based on the uncertainty surrounding the cost-effectiveness decision. At the £30,000 per QALY threshold the CEAC in Figure 4-14 indicated that even though the PET/CT intervention had an ICER of £21,400 the probability of this being cost-effective was only 52%, with the conventional strategy having a 48% probability of being the optimal strategy. This confirms that there is considerable uncertainty as to which intervention is cost-effective in the metastatic analysis, and it is therefore unsurprising that the EVPI results for metastatic CRC report that at a monetary threshold of £30,000/QALY the expected value of perfect information per decision is £1328. This is very high in comparison to the EVPI values for the primary and recurrent CRC models, yet it is unsurprising given that the metastatic model had the most uncertain outcomes, as illustrated on the CEAC.

In order to determine the overall population value of EVPI an annual incidence of 4000 cases of metastatic colorectal cancer was assumed (derived from the annual incidence of colorectal cancer (94), the probability of metastases (150) and the probability of death prior to metastatic diagnoses) and a technology timeframe of two years was assumed, in line with the previous EVPI calculations. The effective population of metastatic CRC patients who could potentially benefit from further evidence was determined to be 7872. The EVPI results are presented in Figure 4-23 at the population level.
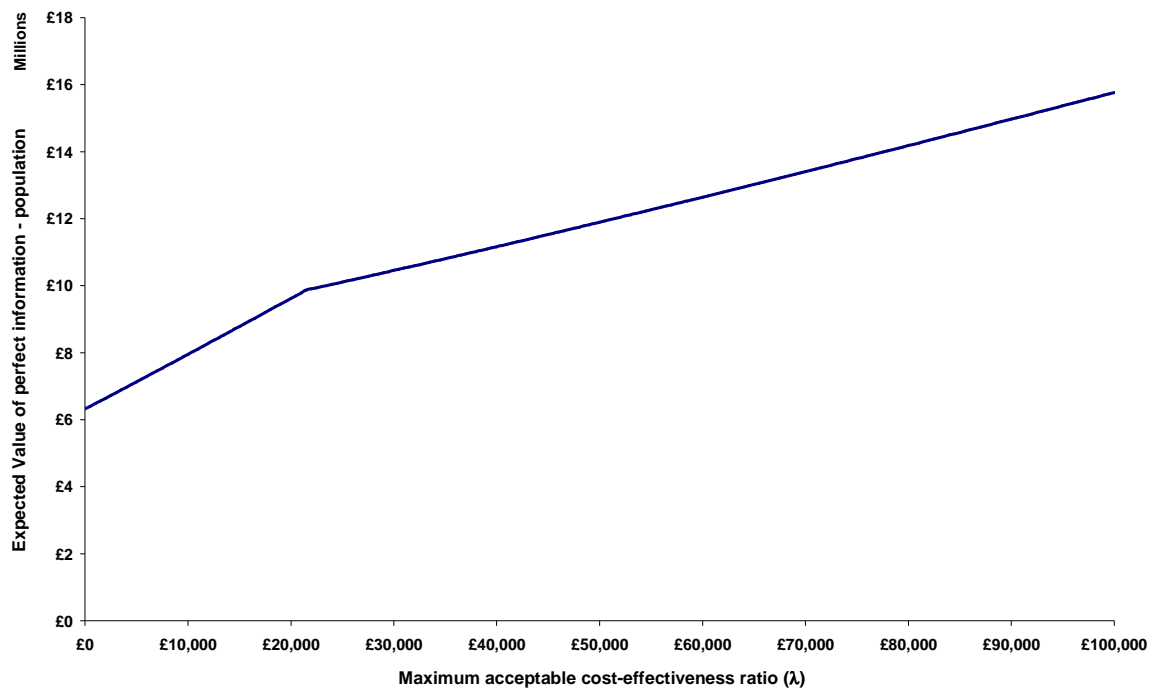
Figure 4-23: EVPI for metastatic colorectal cancer - population

The EVPI results show that it is worthwhile collecting more information about the use of PET/CT for recurrent rectal cancer. At a willingness to pay threshold of £30,000/QALY the expected value of perfect information per decision is £1,328, which translates to approximately £10.5 million at the population level. Even if societies' willingness to pay was zero, the value of additional information would still be extremely high at £6.3 million. This high population EVPI is mostly due to the high degree of uncertainty given current information, and therefore it is worthwhile seeking additional information to help inform the decision regarding the cost-effectiveness of a PET/CT strategy for metastatic colorectal cancer.

Given that further research is worthwhile in metastatic cancer, the analysis can be extended by using EVPPI to establishing what type of further research is required. In this way the model outcomes can be meaningful to decision makers. The same parameter groups which were determined to be possible sources of uncertainty for the recurrent model were used for the metastatic EVPPI analysis: Prevalence, Diagnostic test accuracy (ceCT and PET/CT), Five

year overall survival, Cost of PET/CT, Cost of treating metastases, Utility
values for the disease states and a Trial Information group. Table 4-2 details
the seven parameter groups and the parameters specific to the metastatic
model.

Table 4-2: EVPPI parameter groups and parameters

Parameter Groups	Parameters
Prevalence	no metastases
	metastases
	metastases at one site
	extra metastases
Diagnostic Test Accuracy	ceCT sensitivity & specificity
	PET/CT sensitivity & specificity
5 year overall survival	no metastases
	metastases at one site
	extra metastases - surgery
	extra metastases - palliative
Cost PET/CT	cost PET/CT
Cost metastases	cost metastatic surgery
	cost palliative care
Utilities - disease state	no metastases
	metastases
	extra metastases - surgery
	extra metastases - palliative
Trial information	Diagnostic Test Accuracy group
	cost PET/CT
	Utilities group

The population EVPPI was based on the same patient population (7,872) and
technology timeframe as that specified for the metastatic EVPI calculation.
The EVPPI analysis reported results in terms of the value per decision, and
these are presented in terms of the relevant patient population who would
benefit from the additional information in Figure 4-24. The results are
presented on the log scale, so that the lowest and highest EVPPI values can be
adequately illustrated and for consistency with the previous EVPPI analyses.
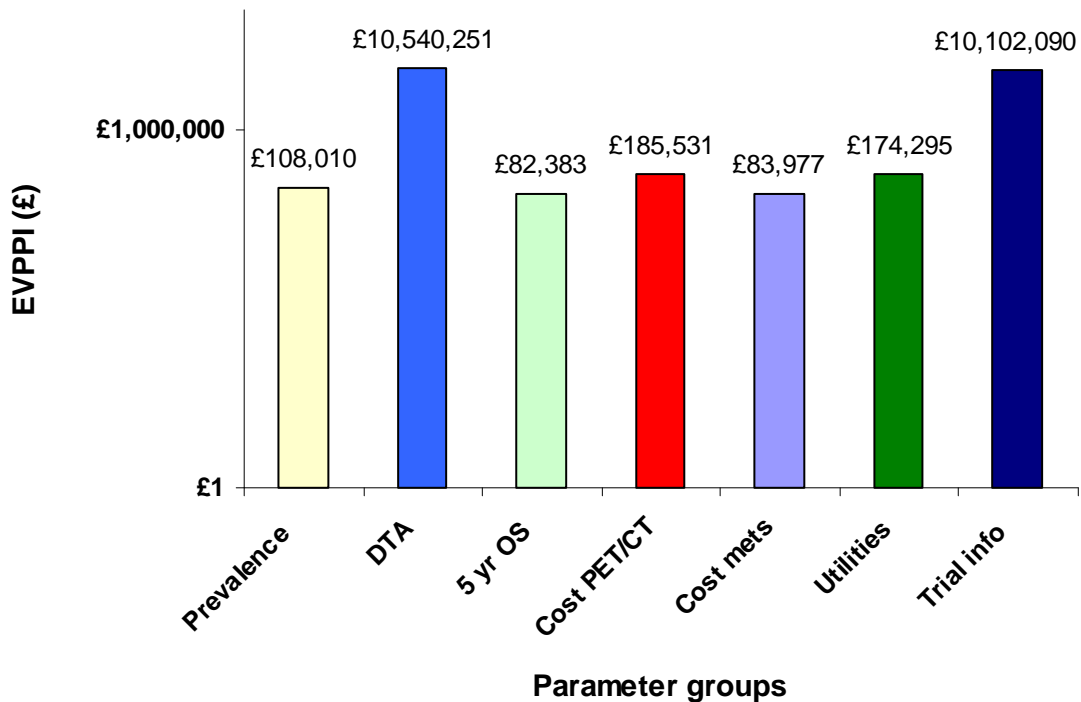
Figure 4-24: EVPPI for metastatic CRC – population level (λ=£30,000)

The analysis found that there was value in undertaking further research in all of the parameters, but that the DTA and Trial information parameter groups were driving the majority of uncertainty in the model.

The five years overall survival parameter group had an EVPPI per decision of £10 which was similar to the cost of metastases parameter groups at £11 per decision. These translated to population values of approximately £80,000. Gathering additional evidence on overall survival would not necessarily require a trial (if for example a routine CRC data set could be accessed), and therefore it may be possible to undertaken research on these parameters at a cost lower than £80,000. The per decision value of further information on the prevalence, utilities and cost of PET/CT groups were £14, £22 and £24 respectively. This translates into population level values of greater than £100,000. These results show that there is a greater uncertainty surrounding the cost of PET/CT than there is the cost of metastatic surgery and palliative care. Also, it is likely to be worthwhile collecting information on the cost of

PET/CT and the utility parameters. While the cost of PET/CT and utilities information can be collected in a clinical trial, one is not necessary, and therefore they could be collected for less than £174,000. Further information on these five parameter groups is of some value; however, the major driver of uncertainty is clearly the uncertainty in DTA parameters. At a per decision value of £1339 and a population value of £10.5 million it is clear that further information would be beneficial and help reduce uncertainty in the cost-effectiveness decision.

Given that these results indicate that a study would be of value to collect information on the DTA characteristics for ceCT and PET/CT in metastatic CRC, the Trial data group indicates that including data collection on utilities and cost of PET/CT would also be worthwhile. However, the cost of PET/CT and utilities parameters appear to have a reducing impact on the overall value, lowering the combined Trial information EVPPI (£10.1 million) to less than that of DTA alone (£10.5 million). This effect was also apparent, and in the same direction, in the recurrent CRC model. The utilities and cost of PET/CT parameters have an effect of reducing uncertainty, resulting in a lower EVPPI value than when the DTA parameters are valued alone. However, it can be concluded that at a value of £10.1 million if a trial was undertaken to gather evidence on DTA it would also be worthwhile collecting data on utility for the different disease groups and the cost of PET/CT.

4.5 Discussion

This chapter used the case study of PET/CT to demonstrate an example of good practice where evidence synthesis and early decision analytic modelling was funded prior to primary research. The chapter explored uncertainty in the results for the each of different CRC patient groups and demonstrated that in using value of information techniques the results can be examined in terms of decision uncertainty to give meaningful recommendations to funders and decision making bodies.

The results for each of the analyses will now be summarised followed by a discussion of using value of information analyses with early decision analytic modelling.

4.5.1 Primary results

The cost-effectiveness planes in the primary analyses showed that there was a lot of uncertainty surrounding both the incremental costs and incremental effects associated with PET/CT; however the cost-effectiveness acceptability curves show that this parameter uncertainty which impacted on the model outcomes does not translate into decision uncertainty. For both the primary rectal and primary colon analyses the probability that PET/CT as an add-on imaging device is cost-effective, is zero over the recommended QALY threshold range of £20k-£30k per QALY. Therefore, we can be certain (99%) that the conventional strategy is cost-effective. In considering the value of further research, the EVPI value was very small, reflecting the low decision uncertainty. Therefore, given current evidence we can be certain that the use of PET/CT as an add-on imaging device for staging primary colorectal cancers is not cost-effective and that there is no value associated with the collection of further information.

As PET/CT technology develops, there will be an increased potential in the future for this improved technology to be used as a lone device, replacing contrast enhanced CT, as opposed to being utilised as an add-on imaging

device. In primary rectal cancer, cePET/CT could potentially replace ceCT with the addition of an MRI scan, and in primary colon cancer, cePET/CT could be used alone as a replacement for ceCT. The two scenario analyses undertaken to explore this in primary CRC indicated that such an improved cePET/CT device is unlikely to be cost-effective for use in primary rectal cancer, but is likely to be very cost-effective for use in colon cancer.

In primary rectal cancer, an improved cePET/CT technology will not negate the necessity for an MRI scan, and therefore the potential incremental value of improved cePET/CT is limited by the strong DTA achievable with MRI scanning. The primary rectal cancer scenario CEAC indicated that at a monetary threshold of £30,000/QALY the probability of cePET/CT technology being cost-effective in comparison to the conventional ceCT and MRI is less than 20%. The colon cancer scenario analysis indicated substantial improvement in DTA from cePET/CT compared to ceCT and improved efficiency through eliminating the need for an add-on test, thereby giving a highly cost-effective outcome. However, there remains considerable parameter uncertainty and decision uncertainty, which is highlighted in the value of information analyses, indicating potential value in further research with a population EVPI of £1.7 million for the primary rectal population, and a value of £12.3 million for the primary colon population.

The scope of the research undertaken was set by the funding body and focussed on PET/CT as an add-on device, however, given the information from the systematic literature review it was evident that as an add-on technology PET/CT was not efficient and would have limited benefit in primary CRC; while an advanced cePET/CT technology (which is currently available in Japan and may become available in the UK in the near future) could be used as a replacement technology for conventional ceCT, rather than as an add-on device. The analysis was limited by the pre-specified research question, and therefore a scenario analysis was undertaken to demonstrate the potential for this alternative to the funders and decision-makers. However, if the funding body had undertaken some explorative research in defining the research question (at stage 1 of the iterative process) they might have been able to specify a more appropriate research question for primary CRC. By applying the

formal process of evidence synthesis and early decision modelling (stage 2 of the iterative economic approach) we realise that the decision problem identified by the funding body was not quite appropriate. The original scope was not fit for purpose with regards to primary CRC, and if a truly iterative process had been adopted, it could have allowed for a re-specification of the research question to explore the potential for cePET/CT in primary colorectal cancer.

4.5.2 Recurrent results

The ICER for the recurrent colon cancer model is considerably lower than that for the recurrent rectal cancer model, indicating that PET/CT is more cost-effective in the assessment of colon recurrence than in rectal recurrence. This difference is likely to be due to the sensitivity estimate for the ceCT diagnostic test parameter, which has considerably lower sensitivity than the PET/CT sensitivity estimate. The wide difference favours the accuracy of PET/CT, and even though uncertainty around both these estimates was incorporated into the model, the strong influence of the DTA parameters on model outcomes is evident. The PET/CT intervention does not have the same DTA sensitivity advantage in the recurrent rectal model, as the MRI scan DTA estimates are also incorporated. The MRI DTA was superior to ceCT and therefore in the recurrent rectal model, the conventional imaging DTA estimates are closer to those of PET/CT, limiting the incremental value of PET/CT.

At a cost per QALY threshold of £30,000 the probability that the PET/CT intervention will be cost-effective for recurrent rectal cancer is 70%, this is greater for recurrent colon cancer (85%). Compared with the primary model analyses (which showed 99% and 100% certainty that PET/CT was not cost-effective) the recurrent model exhibits greater levels of decision uncertainty in both the rectal and colon analyses leading to non-zero values for the EVPI analyses. At a population level the EVPI is £5.6 million for recurrent rectal cancer and £5.1 million for recurrent colon cancer, and therefore there is potential worth in collecting further information to inform the decision regarding the use of PET/CT in the future.

Given that further research is potentially worthwhile, EVPPI analyses were then undertaken for rectal and colon cancer to determine what type of 'further information' would add most value. Seven parameter groups were identified as potential sources of uncertainty and these were used in the EVPPI to explore whether any or all of them were responsible for driving the decision uncertainty. The EVPPI outcomes for the various parameter groups were in line with general expectations; all being very low (or zero in the colon analysis) with the exception of the DTA parameters which had population values of £5.6 million and £4.5 million in the rectal and colon analyses respectively. Given the limited and poor-quality DTA evidence found in the systematic review, it is unsurprising that it is this set of parameters that is driving uncertainty in the model. For both rectal and colon recurrence the EVPPI for the DTA parameter group is very close to the overall EVPI value and therefore it is clear that the DTA parameters are driving the expected value in further research to help reduce the decision uncertainty.

As it is the DTA parameters that require additional evidence, the type of further research that would be required is a study to evaluate the DTA characteristics for the PET/CT, ceCT and MRI technologies. Such a trial would not necessarily be an RCT, however, the existing DTA studies (identified in the systematic review) derived their evidence mostly from retrospective studies in university or oncology hospital settings. Most of the studies used consecutive samples and only a few had comparators with no details provided regarding the use of blinding. Therefore an RCT would provide some stronger evidence regarding DTA especially if it had an appropriate comparator, was blinded etc. Such a trial would not necessarily be a long term trial (as the EVPPI showed that there is little or no value in further research for the survival parameters), but of a short enough duration to determine unbiased DTA characteristics, and could possibly also be used to derive more accurate data on patients' quality of life in the different disease states and the cost of PET/CT.

4.5.3 Metastatic results

The metastatic model found PET/CT as an add-on device to have an ICER of £21,434 per QALY gained. This ICER value is within the UK monetary threshold

range of £20-30,000 per QALY (24) and could therefore be considered cost-effective, however, the probabilistic sensitivity analysis and cost-effectiveness acceptability curve highlighted that there is considerable uncertainty regarding the cost-effectiveness decision. The CEAC (in Figure 4-14) illustrated that at £30,000 per QALY the probability that the intervention strategy (add-on PET/CT) is cost-effective was only 52%. This level of uncertainty leads to an EVPI of £10.5million for the population. Thus there is considerable decision uncertainty regarding PET/CT in metastatic CRC and therefore it is potentially worthwhile collecting further information to inform the decision regarding the cost-effectiveness.

Having established that further research is worthwhile, it is necessary to explore the EVPPI in order to provide more meaningful results and recommendations to decision makers; particularly regarding what type (and possibly adequate duration) of 'further research' is required. The EVPPI analysis found that there was potential value in undertaking research on all seven parameter groups identified, but that it is the diagnostic test accuracy parameters that drive the decision uncertainty.   With a population value of £10.5 million it is clear that further diagnostic information on the ceCT and PET/CT parameters would be beneficial and help reduce uncertainty in the cost-effectiveness decision.

The EVPPI results indicate that a study (possibly, but not necessarily an RCT) would be of most value to collect unbiased evidence on the DTA characteristics for ceCT and PET/CT in metastatic CRC. If such a trial were being undertaken it would be advantageous to also collect data regarding the utility for the different disease groups and the cost of PET/CT as indicated in the combined Trial information group which had an EVPPI of £10.1million.

4.5.4 Model conclusions

The economic literature and DTA systematic review undertaken for this research found only a small amount of evidence to support the use of PET/CT in the pre-operative staging of primary, recurrent and metastatic CRC and the data were generally divergent and the quality of research poor.   The economic

models were therefore designed, developed and populated based on a variety of information from published data sources, literature and in consultation with clinical experts.

The economic evaluations reveal that given the high degree of uncertainty in the models and decision uncertainty in the results, PET/CT as an add-on imaging device is cost-effective in recurrent colon, recurrent rectal and metastatic colorectal disease. There is value in undertaking further research in these disease areas, particular for generating evidence on diagnostic test accuracy. The evaluations also found that add-on PET/CT is not cost-effective in primary colon or rectal cancer and further research is not worthwhile. However, the results of a scenario analysis suggest that future developments in PET/CT technology to enhance the CT element, making it equivalent to regular ceCT, might make cePET/CT cost-effective as a replacement rather than an add-on imaging device in primary colon cancer, although not in primary rectal cancer. Under this scenario further research is potentially worthwhile for both primary rectal and primary colon cancer.

4.5.5 Early DAM & VOI

This chapter has demonstrated the advantages of developing early stage probabilistic models which can be used to inform future research needs, as opposed to a more traditional approach whereby a research problem or lack of evidence is identified and used to support a case for primary research, without involving any decision analytic modelling. In this case study, an exploration of existing evidence was undertaken, and separate analyses were undertaken for each of the appropriate patient groups (primary rectal cancer, primary colon cancer, recurrent rectal cancer, recurrent colon cancer, and metastatic disease); which each had different outcomes in terms of cost-effectiveness and the research needs.

The decision problem was identified (at stage 1 of the iterative approach to economic evaluation) by the funding body and therefore the overall research question was pre-specified. This is commonly the case with research funding bodies which define an overall research question and put out a call for bids to

undertake the research. Research bodies are, arguably, well placed to do this as they are likely to have formal criteria for setting research priorities, as well as the ability to engage with various stakeholders, however, a call for research from a funding body does not necessarily mean that the research question will adequately reflect the correct decision problem. The call may be too prescriptive or alternatively too broad to adequately define the problem and, additionally, the research problem itself may not have been sufficiently explored, resulting in a sub-optimal research question.

In this case of the PET/CT study the original research question was broad, stipulating an economic model to capture PET/CT as an add-on technology in pre-operative staging of colorectal cancer; however, staging pre-operative colorectal cancer is complicated. PET/CT and conventional imaging devices have different diagnostic test accuracies (DTA) for staging primary, recurrent and metastatic colorectal cancer, and as such, in order to address the initial research question, three separate economic models were designed. Patient management routes also differ between colon and rectal cancer, and therefore the primary and recurrent models were adapted to incorporate the specifics of rectal and colon cancer separately. Therefore, five analyses were undertaken, which was considerably more time consuming than originally envisaged. Additionally, the original scope of the question was not fit for purpose with regards to primary colorectal cancer. Early into the research project it became evident that as an add-on technology PET/CT was not efficient and would have limited benefit in primary CRC; while an advanced cePET/CT technology (which may become available in the UK in the near future) has potential value as a replacement technology in primary CRC.

It is interesting to consider that had the funding body undertaken some explorative research and possibly a DAM in defining the research question (at stage 1 of an iterative process) an iterative process could have been applied, allowing for a re-specification of the research question. In the context of the iterative approach methodology, Stage 1 in itself may involve an iterative process where exploration of the literature leads to updating and re-defining of the research question for evidence synthesis and modelling in stage 2. A pre-prescribed question set by a funding body may have less room for such

updating at an early stage and can lead to a research scope that is not fit for purpose or an inappropriate research question.

Having developed a decision analytic model and undertaken probabilistic sensitivity analysis, the results can be used to undertake value of information analyses. VOI is a methodological approach which uses the uncertainty surrounding the cost-effectiveness decision based on current evidence, the size of the patient population, the length of the technology lifespan and societies willingness to pay in order to place a monetary value on the worth of further research. VOI techniques can be used to make informed decisions as to whether a new intervention should be adopted (or rejected) based on current evidence, or whether further information is required to help make the decision, as opposed to making decisions subjectively.

In some cases, such as with the primary colorectal evaluations, further research is found to be not worthwhile and if undertaken would be a waste of resources which could be used elsewhere, i.e. funding other, more promising research or used to fund clinical practice. This case study highlighted that a lack of evidence (or poor quality evidence) does not necessarily mean that further research is required. The primary CRC results illustrated considerable uncertainty in the cost and QALY outcomes from the probabilistic analysis (as detailed in Figures 4-1 and 4-2), however this did not translate into decision uncertainty. If the analysis had not been extended beyond the PSA to consider decision uncertainty in the CEAC followed by EVPI, the cost-effectiveness outcomes may have been interpreted differently. Limited DTA evidence in primary CRC along with uncertainty in the cost and QALY outcomes could have been considered to mean that further evidence in this area would be beneficial. However, the EVPI showed that perfect information would not change the cost-effectiveness decision, and therefore there is no value in conducting further research. This highlights the advantage of employing early decision modelling and using it to formally explore further research needs rather than deciding research needs subjectively on the basis of a lack of evidence.

The EVPI analyses for the recurrent CRC and metastatic patient populations demonstrated that further research is potentially worthwhile, with values in the region of £5 million and £10 million respectively. As a research recommendation these EVPI outcomes are useful in terms of setting an upper limit on the cost of further research. The EVPI sets an upper bound on the value of further research, so that research must cost less than the EVPI to be considered potentially worthwhile. EVPPI analysis can be undertaken to help explore what type and scale of research is appropriate. EVPPI analysis helps consider what parameters are driving the uncertainty in the cost-effectiveness decision, and it is through identifying the relevant parameters, that the appropriate type of research can be determined. Further research does not necessarily mean that a large scale, randomised controlled trial is required, but rather, the type of research will depend on the different parameters that require further information. Likewise, the cost of further research is also dependent upon the type, for example a large scale RCT with a 4 year follow-up period may cost millions of pounds, which would be substantially greater than a smaller scale observational study, or an analysis of a routine dataset. Therefore, using value of information analyses to undertake EVPI, and then an EVPPI if it is potentially worthwhile is necessary to make the most use of outcomes and give informed recommendations to funders and decision-makers.

The EVPPI results for the seven parameter groups identified were in line with expectations. In both the recurrent and metastatic models it was the DTA parameters that had the highest EVPPI value and were driving uncertainty, which makes sense as these were the parameters with most uncertainty identified from the literature. If uncertainty had been restricted to parameters such as the utility or prevalence parameters then a clinical trial would not have been necessary, however for evidence on the diagnostic test accuracy of PET/CT, ceCT and MRI a trial would be required to determine unbiased test characteristics (but not necessarily a randomised controlled trial). The duration of such a study could be relatively short term as the EVPPI determined that there would be little value in additional information regarding five year overall survival. Both the recurrent and metastatic analyses indicated that the Trial information group (which included DTA, utility and cost

of PET/CT parameters) would be of value, however, the utilities and cost of PET/CT parameters have an effect of reducing uncertainty (in both the recurrent and metastatic analyses), resulting in a lower EVPPI value than when the DTA parameters are valued alone. However, it can be concluded that at a value of £5 million and £10 million for recurrent and metastatic CRC respectively, if a trial was undertaken to gather evidence on DTA it would also be worthwhile collecting data on utility for the different disease groups and the cost of PET/CT.

In summary, this chapter has demonstrated that even though early decision analytic modelling is not often funded and value of information analyses are even less frequently utilised in practice, it can be relatively straightforward and can have substantial benefits in terms of understanding outcomes and determining future research priorities when adequate funding and time is allocated to these activities. Just because current evidence is limited or of poor quality, does not necessarily mean that further research is required, and therefore, applying early DAM and VOI techniques allows decision makers to make informed decisions as to whether a new intervention should be adopted (or rejected) based on current evidence, or whether further information is required to help make the decision, as opposed to making decisions based on subjective reasoning.

5 Designing clinical trials with economic evaluations

Economic analyses are commonly undertaken alongside randomised controlled trials (RCTs) however; such trials are rarely designed with the economic outcomes in mind, and rarely involve the use of an early decision analytic model to aid in the trial design. RCTs are typically designed to show evidence surrounding the clinical effectiveness of a new treatment or health technology in comparison to the current standard of care, and the economic component is often seen as an add-on to the trial (20;28). Design considerations are different for clinical and economic analyses (44) and consequently when economic evaluations are conducted alongside clinical trials which were designed with only the effectiveness outcome in mind, they may be inadequately powered and designed for the cost-effectiveness analysis and therefore may represent a partial or limited form of economic analysis.

If the purpose of the analysis is to inform decision making based on cost-effectiveness outcomes then sample size and power calculations should be directly related to the cost-effectiveness result rather than to the effectiveness outcome alone. Over the last decade there has been much discussion in the health economics community over alternative methods for undertaking such calculations (40;46;169-171) and the methodology for such an approach is now established. Therefore considerations of economic assessments alongside clinical trials can and should be used to guide conventional trial design (39). Despite the established methodology, in practice, health economists are rarely given the opportunity to contribute to trial design (46).

This chapter reports on an atypical real life case in which a clinical and cost-effectiveness trial was designed and powered to determine the sample size with regards to cost-effectiveness. The Health Technology Assessment (HTA) programme (50) had issued a commissioning call for research to undertake an RCT and therefore the trial had already been planned without prior economic involvement, i.e. neglecting the first two stages of the iterative economic

approach. In principal there could/should have been a formal exploration of the decision problem followed by a literature review and evidence synthesis process involving a comprehensive decision analytic model built on the existing evidence. If such an approach had been undertaken prior to the call for an RCT, probabilistic sensitivity analysis could have been undertaken to determine the potential cost-effectiveness of the new treatment and to formally explore uncertainty in the current evidence with regards to the cost-effectiveness decision. Value of information analyses (EVPI and EVPPI) could also have been undertaken to determine whether and what further evidence was potentially worthwhile with an EVSI analysis (as described in Chapter 2) undertaken to determine an appropriate sample size for any such trial. EVSI is a technique used to quantify the expected value to the decision maker of obtaining sample information before making a decision (84). EVSI and ENBS techniques are used to establish the efficiency of the proposed research design, ensuring a cost-effective use of research funding.

As is often the case, an iterative process such as this had not been followed, and an RCT was already planned, however, the opportunity arose to power the trial for economic endpoints.   Therefore a simple decision model was constructed as the proposal was being prepared and used to help inform the design of the trial. Even when there has been no prior economic modelling, involvement in the design of a trial from an economic perspective is still desirable and made possible by constructing a simple model within a short time frame.

This chapter demonstrates the feasibility of sample size calculations from an economic perspective in the context of a standard frequentist trial, while simultaneously highlighting that a decision model can be constructed quickly and simply alongside the preparation of a trial proposal.

The chapter discusses the proposed RCT of fetal fibronectin screening (fFN) for women with threatened pre-term labour, followed by the design of a preliminary model which was developed as the proposal was being prepared to inform the trial design calculation. The predictions from this early 'pre-trial' model indicate potential cost-savings, but with a marginal detrimental impact

on the effectiveness endpoint, neonatal morbidity. The atypical context of this situation, i.e. a trade-off between cost savings and a reduction in effectiveness, meant it may be appropriate to design the RCT as a non-inferiority trial. The aim of a non-inferiority trial is to demonstrate that the fFN test is not worse than the comparator (in terms of the outcome neonatal morbidity) by more than a pre-specified small amount known as the non-inferiority margin (172). Therefore this chapter explores an appropriate design and sample size for the RCT using the outcomes from the pre-trial decision model. The model predictions are discussed and used to explore sample size calculations based on each endpoint to determine their importance. Following this, the net monetary benefits (NMB) approach for cost-effectiveness is discussed and used to calculate the required sample sizes at different powers. A non-inferiority approach, which is likely to be the best design for the trial, is then explored and a suitable non-inferiority margin determined. The sample size calculations are then recalculated using the non-inferiority margin and compared with the NMB sample size, to select an optimal sample size for the trial which is sufficient for the cost-effectiveness endpoint and also to demonstrate non-inferiority for the effectiveness endpoint.

5.1 Case study: fFN testing for pre-term labour

Pre-term births (before 37 weeks) occur in nearly 8% of pregnancies in the UK (over 50,000 deliveries per annum); they are the leading cause of neonatal death and are also associated with both short and longer term morbidities (173). Neonatal morbidities, including respiratory disease syndrome, are experienced by approximately 24% of pre-term infants (174), impacting on the infants' future quality of life, but also affecting the parents and families quality of life. There is currently no effective way of preventing pre-term labour; however, timely interventions such as administering steroids (175) to pregnant women who have symptoms of pre-term labour can help reduce morbidity and mortality experienced by preterm infants.

In the UK approximately 15% of all pregnant women will experience symptoms of pre-term labour (120,000 annually) and under current practice these women

are hospitalised and receive steroids to lower the risk of infant morbidity and mortality. Accurate diagnosis of preterm labour is challenging, and only 20% of women who are clinically diagnosed with symptoms actually experience premature births (174). This clinical diagnostic uncertainty results in a large proportion of women with symptoms of pre-term labour being treated unnecessarily to ensure benefits to the small proportion of babies that do actually deliver pre-term. While this 'treat-all' approach is clinically understandable, it is also detrimental, both for the women who are unnecessarily hospitalised and frequently transferred between maternity units causing considerable distress to them and their families (176), and also to the NHS which incurs a substantial cost burden.

The fetal fibronectin (fFN) test is a diagnostic test which is easily performed and is potentially an effective means of diagnosing symptoms of pre-term labour. A research study was proposed to explore whether using fetal fibronectin testing in women threatened with pre-term labour in the UK would lead to an improvement in patient management and resource use through reducing hospital admissions, transfers and intervention rates without increasing the risk of morbidities and mortality experienced by preterm infants. The test was to be compared with standard practice in which diagnosis of pre-term labour amongst presenting women is based on clinical judgement and an 'admit all to hospital' approach. This could be considered as an equivalence trial, assessing any improvements in patient management, with an expectation of similar endpoints in terms of actual pre-term births, neonatal mortality and neonatal morbidity. There is, however, the potential for a slight increase in neonatal morbidity due to false negative test results, which needs to be considered. Therefore the trial was designed based on the hypothesis that introduction of the fFN test would decrease unnecessary antenatal interventions, leading to substantial cost savings for health services, but that the existence of false negative test results could potentially increase neonatal morbidity. Given this hypothesis, it is likely that a non-inferiority trial would be the most appropriate design, aiming to demonstrate that the fFN test is not worse than standard practice, (the admit all approach) in terms of neonatal morbidities, by more than a pre-specified small amount.

While infant death from pre-term labour is rare (177) pre-term labour is very common, with a resulting morbidity impact on approximately 9 in 1000 pregnancies in the UK (178-180). Therefore the RCT was designed in order to undertake a cost-effectiveness analysis of the fFN diagnostic test, in which cost savings that accrue through reduced hospitalisations could be accounted for, along with any potential impact on neonatal morbidities.

5.1.1 A pre-trial model for fFN testing

Prior to the commission of the RCT in fFN, there had been no economic involvement in the project, and therefore no funding or time allocated to the development of a comprehensive decision analytic model. Such a situation is very common in the 'real world', where an economic component of the trial is not considered until the desire for and possibly the development of the trial is underway. However, such a situation does not necessarily negate economic input. A simple economic model can be constructed within a relatively short time frame and therefore it is still possible to have some economic input while the trial is being designed. Decision analytic models are simplified versions of reality and they can be as simple or complex as required without losing credibility (30). Therefore, in the fFN case study, a simple 'pre-trial' decision model was developed with the intent of informing the trial sample size calculations.

The model was populated with available data to assess the potential impact of introducing the fFN test on hospitalisations, neonatal morbidity and neonatal mortality in comparison to no test. These model outcomes could then be used to aid the protocol design and undertake sample size calculations. The model was developed in order to inform the trial design within the constraints of a grant submission timeframe utilising the most readily available evidence on the costs and effectiveness of fFN testing. This model is therefore a simple, preliminary estimation, and like all models, subject to uncertainty in the parameter estimates.

The two interventions of interest in the trial are (i) practice based on the results of the fFN test and (ii) standard practice in which there is no test and

diagnosis is based upon clinical judgement alone. The pre-trial model was developed using a decision tree. It was structured over a time horizon of months, representing the maximum time frame for women presenting with threatened pre-term labour (24+ weeks gestation) through to neonatal discharge (or until 3 months post birth). Those women with a positive diagnosis (either based on clinical judgement or fFN test results) are hospitalised and receive steroids to lower the risk of pre-term related infant morbidity and mortality, while those diagnosed as negative are sent home and do not receive steroids. The model timeframe ends at 3 months post-birth in order to capture the resultant neonatal morbidity and mortality impacts. The decision tree detailed the patient pathway from a clinical diagnosis of threatened pre-term labour through to test outcome and actual birth term status to distinguish between accurate and inaccurate diagnoses and the resultant impact on hospitalisation, morbidity and mortality. The details of the decision tree are explained in the following section. Once the patient pathway was depicted, model parameters (such as the costs, prevalence of pre-term delivery, diagnostic test accuracy and risk of morbidity) were assigned to the appropriate branches in the tree, to calculate the proportion of hospitalisations, pre-term neonatal morbidities and neonatal mortalities in each arm.

5.1.2 Decision Tree

The decision tree (Figure 5-1) was developed to illustrate the alternative pathways in the proposed trial. The tree begins with the population of women in the UK who have been clinically diagnosed with pre-term labour (threatened pre-term labour) who will then be randomised to either the intervention arm (fFN Test) or the control arm (No test). The tree has been designed based on the probability of actual term and pre-term births within the threatened pre-term labour population. The tree splits the patient population according to pre-term and term prevalence prior to the test outcomes, so that accurate and inaccurate test diagnosis can be identified based on the sensitivity and specificity of the fFN test.
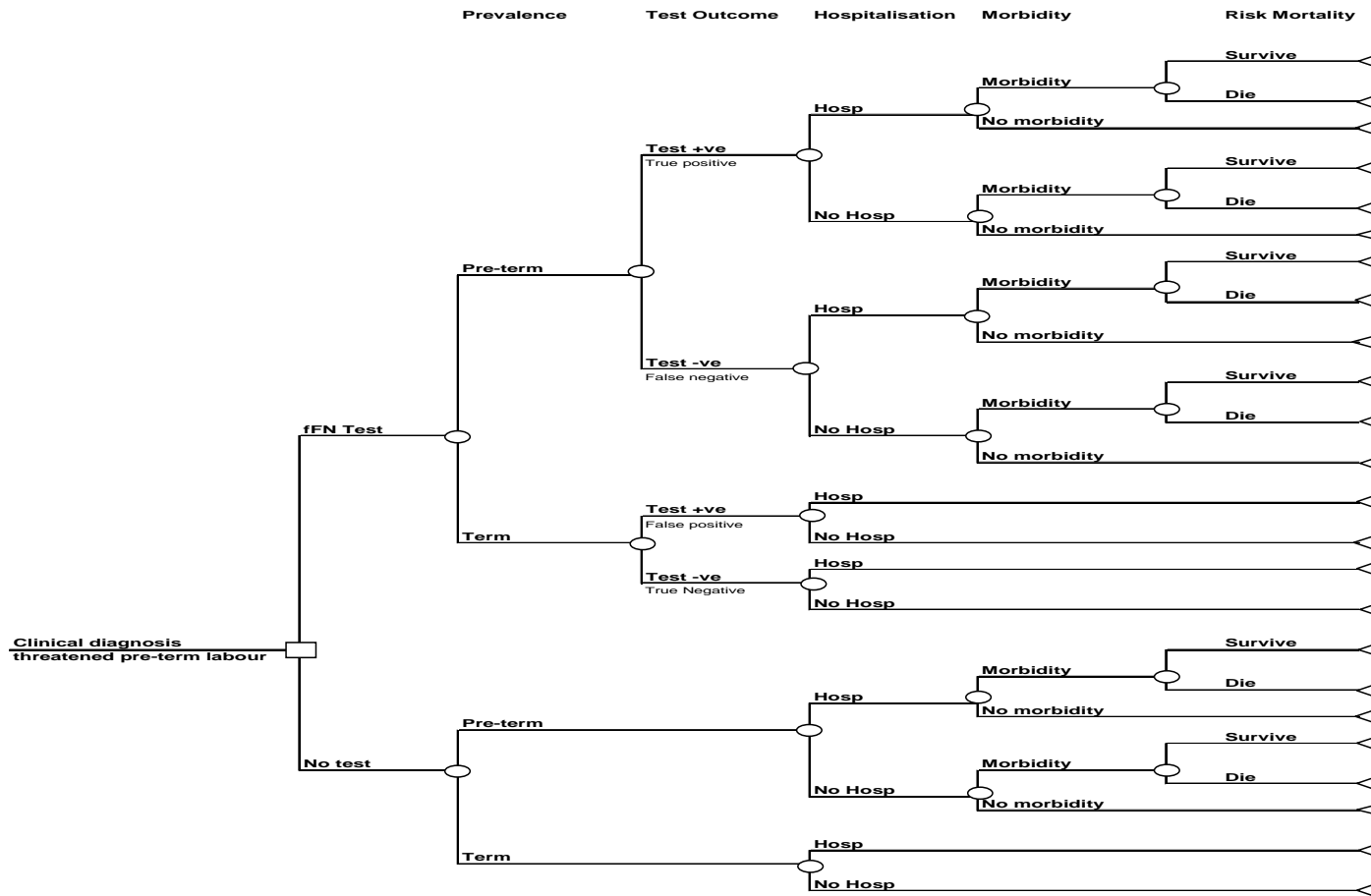
Figure 5-1: Decision tree for diagnosing pre-term labour

In the intervention arm (the top branch of Figure 5-1), those identified as "Test positive" by the test are diagnosed as pre-term labour (accurately and inaccurately) and are hospitalised and receive steroids which reduce the risk of neonatal morbidity. Those identified as "Test negative" by the fFN test are diagnosed as a term labour (accurately and inaccurately) and are not hospitalised and therefore do not receive the morbidity reducing steroids.

It is unrealistic to assume that the introduction of a diagnostic test will completely overrule clinical judgement and be adhered to fully, and therefore the structure of the model allows for the possibility of hospitalisation with negative test results and also for the possibility of no hospitalisation in some cases that had positive test results. These assumptions were based on clinical advice and probabilities were determined from relevant audit data (181;182). By overruling the diagnostic test results with their clinical judgement the clinicians could be considered to be behaving irrationally, however, as clinical judgement has been the long standing current practice, this is more of an issue of professional trust in a new diagnostic test. By overruling the test and 'admitting anyway' the clinicians may in fact be (implicitly) addressing potential false negative results, possibly avoiding complications through failing to administer steroids and any potential resultant litigation proceedings for malpractice/incorrect diagnosis.

The control arm (No test) is represented in the bottom half of the decision tree. As in the top half of the tree, the population is split based on actual birth term status. There is no test in this arm of the trial, and therefore all women clinically diagnosed with threatened pre-term labour would be hospitalised and receive steroids. While current clinical practice follows an 'admit all' approach, a 100% admittance assumption was deemed to be unrealistic in consideration of obstetrician's clinical judgement and therefore the model structure allows for the possibility that some women who were clinically diagnosed as pre-term would not be admitted to hospital.

The model assumes that all pre-term infants are subject to a risk of neonatal morbidity, however those whose mothers were hospitalised in the model (and therefore received steroids) have a lower risk of morbidity, than those whose

mothers were not hospitalised, and did not receive steroids. Therefore, women who experience false negative test results and so do not receive the risk reducing steroids may consequently have infants who experience greater neonatal morbidity than those in the control arm (under an 'admit all' approach). The model does not attribute any negative effects of receiving steroids unnecessarily, as clinicians advised that any adverse effects would be experienced over the longer term (183;184) out with the time horizon of the model. The clinical experts on the project team advised that in current practice (where there is an 'admit all' approach based on clinical judgement) the potential risk of future negative consequences from receiving steroids unnecessarily are likely to be outweighed by the short term risk to mother and child of not receiving steroids that are needed. The model assumes that only pre-term infants subject to neonatal morbidity are at risk of neonatal mortality. Term infants are assumed not to be subject to either pre-term morbidity or mortality.

The parameter estimates applied to the model were based on available evidence and are detailed in Table 5-1. The sensitivity and specificity of the fFN test, probabilities of delivering pre-term, experiencing pre-term neonatal morbidity, mortality and the morbidity risk reduction from steroids were based on published evidence.

The probabilities of hospitalisation in the intervention arm for test positive (93%) and test negative (32%), were based on recent UK audit data (181;182), and in the control arm an author assumption (based on clinician advice) was applied for probability of hospitalisation (90%), rather than assuming a protocol adherence of 100%.

A scenario analysis was also undertaken, where the clinical judgement assumptions were relaxed and there was 100% adherence to protocol. Therefore 100% of women in the control (no test) arm would be hospitalised, and in the intervention arm 100% of positive diagnoses would hospitalised and 100% of negative diagnoses would be sent home.

Table 5-1: Model parameters

Item	Unit	Description	Source
fFN Sensitivity	0.822	Average sensitivity from 7, 14 & 21 days sensitivity outcomes	(185)
fFN Specificity	0.787	Average specificity from 7, 14 & 21 days sensitivity outcomes	(185)
Prob pre-term	0.2	Probability of delivering pre-term amongst threatened pre-term population	(174)
Prob pre-term morbidity	0.244	Probability of morbidity in pre-term births (ICU admissions only)	(174)
Steroid risk reduction	0.54	Relative risk pre-term morbidity reduction with steroids	(175)
Prob mortality	0.0257	Probability of mortality in pre-term births (average ≤36 weeks)	(177)
Prob hosp admit fFN +ive	0.93	Clinical judgement: Probability of hospital admission when fFN positive	(181;182)
Prob hosp admit fFN -ive	0.32	Clinical judgement: Probability of hospital admission when fFN negative	(181;182)
Prob hosp admit NoTest	0.9	Clinical judgement: Probability of hospital admission when No Test	Author assumption
Risk hospital transfer	0.321	Risk of hospitalised patients being transferred to another hospital	(186)
Cost fFN Test	£50	Cost of administering test	(185)
Cost hospital admission	£1,068	Maternity inpatient cost per stay (average 2.2days) including treatment	(187)
Cost hospital transfer	£1,000	Estimate to reflect NHS cost of transfers between different hospitals	Author & clinical expert assumption

ICU=intensive care unit, prob=probability

5.1.3 Pre-trial model predictions

The pre-trial model was used to calculate the proportion of women that were hospitalised and the proportion of infants that experienced pre-term neonatal morbidity and mortality, for the intervention (test) and control (no test) arms. These model outcomes are predictions based on the decision tree pathways, assumptions and parameter estimates. Table 5-2 details the model predictions, showing that there are nearly 37.6% fewer hospitalisations in the test arm, while there is a minor increase in pre-term morbidity and mortality of 0.18% and 0.005% respectively. So, in comparison to the current standard of care where there is no test administered, the introduction of the fFN test is likely to substantially reduce hospitalisations but at the expense of an additional 18 neonatal morbidities and 0.5 neonatal mortalities per 10,000 threatened pre-term births.

Table 5-2: Model predictions

Model Arm	Proportion Hospitalised	Proportion Mortality	Proportion Morbidity
Test (intervention)	0.52423	0.00078	0.03036
No test (control)	0.90000	0.00073	0.02860
Difference	-0.37577	0.00005	0.00176

The average cost of hospitalisation was found to be £1439. This is based upon the unit costs detailed in Table 5-1 for admissions, test and transfer and the associated probabilities of incurring these costs. Therefore, a reduction in hospitalisations by 0.375 is equivalent to a cost saving of £541 per women with threatened pre-term labour. Considering that it is estimated approximately 120,000 women are clinically diagnosed with threatened pre-term labour in the UK each year (173), an fFN screening test which reduced pre-term hospitalisation of these women by 37.5% at an average cost of £1439 per hospitalisation, could potentially save the NHS approximately £65 million per year through avoiding unnecessary hospitalisations, however, this strategy is likely to present an additional 211 cases of pre-term neonatal morbidity and 6 additional cases of neonatal mortality annually.

In this case the model predictions would fall into the South West quadrant of the cost-effectiveness plane. This is demonstrated in Figure 5-2, showing the model predictions of a reduction in effectiveness (an increase in probability of morbidity by 0.0018) accompanied by cost savings of £541 per women with threatened pre-term labour.

fFN test versus No test
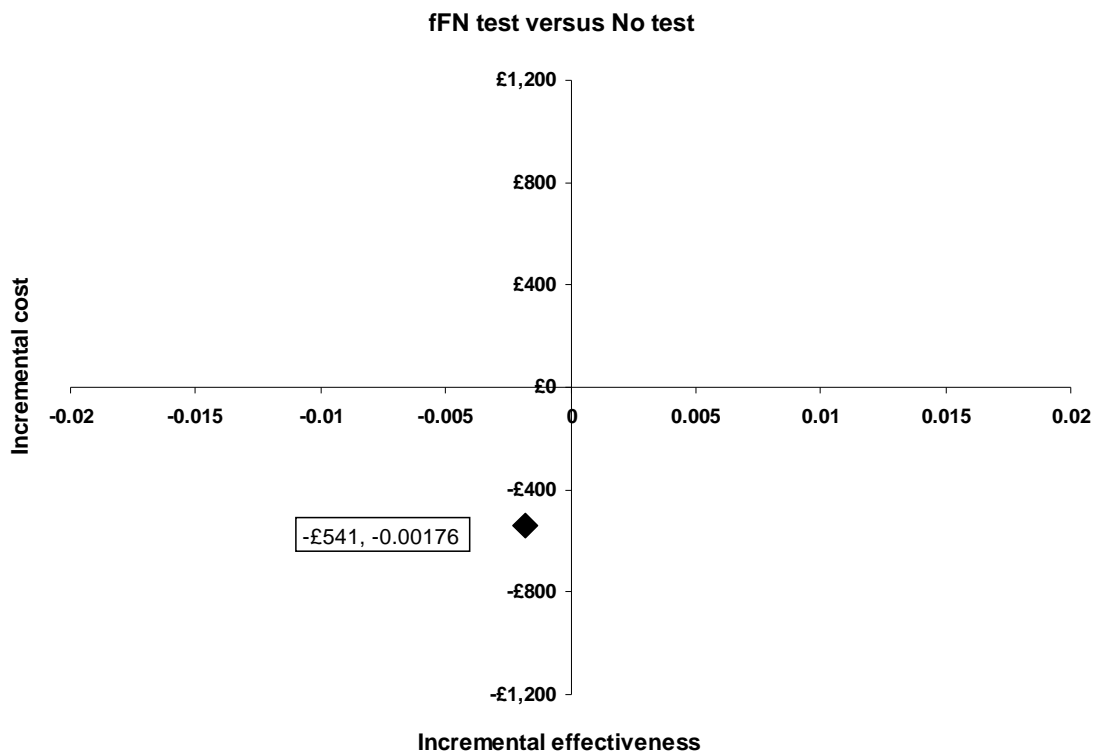
Figure 5-2: Cost-effectiveness plane for fFN model prediction

Interventions that fall in the North East and South West quadrants require an estimate or 'value judgement' regarding societies willingness to pay (or willingness to accept savings) per unit of effectiveness gained (lost). This value is the maximum acceptable cost-effectiveness ratio or ceiling ratio (λ), as discussed in Chapter 2 (section 2.4.3). To determine whether the fFN model outcomes are cost-effective or not, the ceiling ratio needs to be determined[7].

---

[7] A detailed discussion of the ceiling ratio and how it was derived for the fFN trial is provided in section 5.2.2

Most commonly society is happy to accept a monetary threshold of willing to pay per unit of effectiveness gained, but many decision makers can become uncomfortable when considering the reverse, a willingness to accept a cost saving per unit of effectiveness lost. Logically, if we are willing to accept a specified ceiling ratio for gaining effectiveness, we should be willing to accept the reverse. However, in the South West quadrant ethical issues may come in to play and some decision makers may be less willing to accept the same ceiling ratio for this quadrant. O'Brien and colleagues make the case that consumers' willingness to accept a payment to forego a benefit is greater than their willingness to pay to get the benefit and therefore, such considerations should be applied to the maximum cost-effectiveness ratio (78). Under this approach decision makers may require a much greater cost saving in order to accept a reduction in effectiveness, resulting in a kinked ceiling ratio.

This concept of a kinked ceiling ratio (78) is however, irrational as it results in a situation where for example, the ceiling ratio for the North East (NE) quadrant of the cost-effectiveness plane is £30,000 per QALY gained, but for the south west (SW) quadrant is £50,000 per QALY lost. In such a situation society is not willing to accept an intervention which saves £30,000 but is accompanied by a reduction of 1 QALY; the intervention would have to save at least £50,000 per QALY lost. However, this is irrational, and importantly it does not consider the opportunity cost of imposing a greater ceiling ratio in the SW quadrant. The £30,000 saved (at the expense of 1 QALY lost) could be spent more efficiently on another health technology in the NE quadrant which had for example an ICER of £15,000 per QALY, i.e. gaining 2 QALYs for £30,000. So by setting a greater ceiling ratio for willingness to accept a reduction in effectiveness (SW quadrant) society is imposing an even greater opportunity cost in terms of other highly cost-effective technologies which could have been funded on the basis of accepting a technology which gives a saving of £30,000 per QALY lost.

Scenario: 100% protocol adherence

Table 5-3 details alternative model predictions under a scenario where the clinical judgement assumptions are relaxed and patients in the intervention arm are only hospitalised when the fFN test is positive, and 100% of women in the control (no test) arm are hospitalised under the 'admit all' approach.

Table 5-3: Scenario model predictions

Model Arm	Proportion Hospitalised	Proportion Mortality	Proportion Morbidity
Test (intervention)	0.33480	0.00078	0.03035
No test (control)	1.00000	0.00068	0.02635
Difference	-0.66520	0.00010	0.00400

The scenario analysis predicts a 66% reduction in hospitalisations in the test arm compared to the control, with an increase in pre-term morbidity and mortality of 0.4% and 0.01% respectively. In comparison to the baseline model predictions in Table 5-2, the reduction in hospitalisations is improved (as there are more hospitalisations in the control arm, and fewer under the test arm); however the negative impact on effectiveness in terms of mortality and morbidity is greater. Under the assumption of no clinical judgement and 100% protocol adherence, the introduction of the fFN test is likely to substantially reduce hospitalisations by 66% but at the expense of an additional 40 neonatal morbidities and 1 neonatal mortality per 10,000 pre-term births in comparison to the current standard of care where all threatened pre-term labour women are admitted to hospital to receive risk reducing steroids. By relaxing the clinical judgement assumptions the model outcomes are exaggerated in both directions. The cost savings through reduced hospitalisations are even greater than in the baseline model; however, the cases of pre-term morbidity have more than doubled, increasing from 18 per 10,000 births in the baseline model predictions to 40 per 10,000 births in this extreme scenario. In reality clinical judgement will always come into play alongside the results of a diagnostic test, and in the control arm, where only clinical judgement is used it is unrealistic

to assume that 100% of the threatened pre-term labour women would be admitted to hospital.

The pre-trial model predicted that the new intervention (fFN test) is likely to be cost saving, with a small but potentially detrimental impact on effectiveness in terms of both a marginal increase in neonatal morbidity and also in neonatal mortality. Therefore it is important to investigate both the morbidity and mortality endpoints when designing the fFN trial in order to explore what can be considered an acceptable reduction in effectiveness, which in turn will impact on the sample size calculations.

5.2 Sample Size calculations

5.2.1 Superiority Sample Size

The sample size for a trial is calculated at the design stage to ensure the study will have adequate power to show a significant difference between the two arms, this is known as a superiority trial design i.e. it is designed to detect a significant difference between treatments (188). The endpoints of interest from the fFN trial are the differences in hospitalisations, neonatal morbidity and neonatal mortality, reported as proportions. In this study hospitalisations represent the cost endpoint, but there are two effectiveness endpoints of potential interest, morbidity and mortality. Superiority sample sizes were initially calculated for each of these three endpoints, in order to determine which of the two effectiveness endpoints was likely to be the most meaningful for the cost-effectiveness sample size calculation.

The superiority sample size requirements were calculated using standard methods for proportions (72). Typically sample sizes are calculated by comparing the means of two populations which have the same known variance, however; when detecting a difference between two proportions, the approach is slightly different. A test of the null hypothesis is specified, that each proportion is equal to the pooled value of the two proportions and the variance of the difference between the proportions can be different for the null and

alternative hypotheses. As this case study uses proportions from the pre-trial, for simplicity sake a normal distribution was assumed for the proportions (as n approaches infinity) to allow equal variance for the difference in effect under the null and alternative. Table 5-4 details the parameters used for these calculations and their values, where M represents the effectiveness outcome for the morbidity calculation, D represents the effectiveness outcome for mortality and H represents the hospitalisation outcome.

Table 5-4: Sample size calculation parameters

Abbreviation	Value	Description
H	0.712114	Hospitalised proportion pooled
H_T	0.524228	Hospitalised proportion Treatment
H_C	0.9	Hospitalised proportion Control
M	0.0294788	Morbidity proportion pooled
M_T	0.0303608	Morbidity proportion Treatment
M_C	0.0285968	Morbidity proportion Control
D	0.0007576	Mortality proportion pooled
D_T	0.0007803	Mortality proportion Treatment
D_C	0.0007349	Mortality proportion Control
λ	£25,700	Value of avoiding morbidity
C	£1,439	Cost Hospitalisation
Power	0.9	Beta (β)
Significance	0.05	Alpha (α)
$Z\alpha/_2$	1.96	Critical value of significance test
Zβ	1.28	Critical value for desired power
ρ	1	correlation of the difference in costs & effects

The sample size (n) is required to show a significant difference in effect (ΔE) between the two arms. This is calculated using the algorithm detailed in Equation 5.1. The sample size problem is expressed as one of determining the sample size given the power and magnitude of the specified effect to be detected (72). Therefore, following the convention for statistical significance, alpha (α) was set to show a significant difference at the 5% level and the power was set at 90%. z_α is the 100 percentile for the standard normal distribution, so that $z_{\alpha/2}$ is the standard normal deviation with a two sided significance level, and given a specified power of 90%, z_β is the critical value. The null hypothesis (ΔE_N) assumes there is no difference between the pre-term neonatal

morbidity in the two arms ($E_T = E_C$), i.e. the pooled proportion is used; and the alternative hypothesis (ΔE_A) assumes that there is a difference between the two arms ($E_T \neq E_C$). $z_{\alpha/2}$ is multiplied by the variance (v) of the change in morbidity under the null hypothesis, and z_B is multiplied by the variance (v) of the change in morbidity under the alternative hypothesis.

$$n > \left[\frac{z_{\alpha/2} * v(\Delta E_N) + z_B * v(\Delta E_A)}{\Delta E} \right]^2 \qquad (5.1)$$

Equation 5.2 (72) provides a more detailed breakdown of this calculation, specifying the variance of the difference in morbidity under each hypothesis. As the null assumes no difference ($\Delta E = 0$) between the treatment and control arms, a pooled value for effect (E) is used, while the alternative hypothesis calculates the difference in effect ($\Delta E = E_T - E_C$) using the proportion of effectiveness in the intervention arm (E_T) and the proportion of effectiveness in the control arm (E_C).

$$n > \left[\frac{\left(z_{\alpha/2} * \sqrt{2E(1-E)}\right) + \left(z_B * \sqrt{E_T(1-E_T) + E_C(1-E_C)}\right)}{E_T - E_C} \right]^2 \qquad (5.2)$$

This equation was solved for pre-term neonatal morbidity using the M values (M, M_T and M_C) detailed in Table 5-4 to represent E, E_T and E_C which are the pooled, treatment and control predictions for morbidity from the pre-trial model. The calculation was then repeated solving for hospitalisations (using the H proportions in Table 5-4 to represent E, E_T and E_C) and for mortality (using the D proportions in Table 5-4 to represent E, E_T and E_C). The sample sizes derived from each of these calculations are detailed in Table 5-5.

Table 5-5: Superiority sample sizes for each endpoint

Outcome	Sample per arm	Total sample size
Hospitalisations (H)	28	56
Morbidity (M)	193,222	386,444
Mortality (D)	7,740,945	15,481,890

Using this superiority approach it can be seen that the mortality outcome measure predicts an enormous and unrealistic sample size requirement of over 7 million participants per arm to show a significant difference between the intervention and control. The sample size based on the morbidity measure is much smaller but also unrealistic at over 193,000 per arm; while the hospitalisation measure requires an extremely small sample size of only 28 participants per arm to show a significant difference between the treatment and control.

These outcomes illustrate that it is possible to base a superiority sample size calculation on the hospitalisation measure; however, this is only an indicator of cost and does not provide any information about safety or mortality outcomes. It is clear from the model predictions in Table 5-2 and the superiority sample size outcomes in Table 5-5 that the mortality outcome will only show a negligible difference between the intervention and control arms. The morbidity outcome measure predicted from the pre-trial model indicates a slight increase in morbidity (18 per 10,000 threatened pre-term labours), and therefore has some potential to be a meaningful measure of effectiveness. Therefore it was deemed appropriate to explore the relevance of a non-inferiority approach for morbidity as a measure of effectiveness, i.e. to show that the fFN test is not worse than the control, within a pre-specified margin of non-inferiority for morbidity (172;188).

Ultimately the trial was concerned with calculating the sample size based on the cost-effectiveness outcome rather than the cost or effectiveness outcomes individually. Therefore it was decided that the hospitalisation and morbidity predictions from the pre-trial model would be used in the cost-effectiveness sample size calculation using the net monetary benefit approach (NMB), and would then be double-checked against a non-inferiority calculation for the morbidity effect alone.

The next section discusses the NMB approach and details the sample size calculation that was undertaken, followed by a discussion of the non-inferiority approach that was used to ensure the sample was large enough to show that

the test was no less effective than the control within a pre-specified margin of non-inferiority.

5.2.2 Net Monetary Benefit Sample Size

The NMB approach

With regards to the fFN case study, neonatal morbidity is the effectiveness endpoint while hospitalisations are used to reflect any potential cost savings. As such the NMB approach will balance the monetarised value of the increase in neonatal morbidities against the cost savings resulting from reduced hospitalisations. Therefore the two elements of the cost-effectiveness outcome are the difference in hospitalisations (ΔH) and the difference in neonatal morbidity (ΔE) between the two arms. Equation 5.3 illustrates the net monetary benefit calculation, which requires a 'willingness to pay' value (λ) to reflect the monetary value of avoiding morbidity. This is multiplied with the difference in neonatal morbidity (ΔE), while the average cost of hospitalisation (C) is combined with the difference in hospitalisations (ΔH) and subtracted.

$$NMB = \lambda * \Delta E - C * \Delta H \tag{5.3}$$

The 'willingness to pay value' (λ) is a subjective value. In this case it is used to reflect the monetary value of avoiding pre-term neonatal morbidity and therefore some assumptions are used to derive an appropriate value. In order to derive a realistic value, the statistical value of a life was taken to be a useful monetary indicator for the value of preventing an infant mortality and was multiplied by the probability of experiencing pre-term mortality, given pre-term morbidity in the model, to determine a value of avoiding an infant morbidity. The European Union (2001) recommendations suggest that a value between €0.9 and €3.5 million (1€ ~ £0.6004, 2001) should be assigned to reflect the value of a life (189). The UK Department for Transport recommend a value of £1.4million (190) while Abelson et al. (191) also report similar ranges in a review of studies that have estimated the value of a life, however they note that some studies have indicated that the top end of these ranges

are overestimations. Given this evidence it was deemed to be reasonable to assign a value between £500,000 and £2.5million, and it was assumed that an estimate of £1million would be a realistic reflection of the statistical value of a newborn life which has not yet contributed to society. The statistical value of avoiding a pre-term neonatal morbidity was then derived (by multiplying the probability of mortality amongst pre-term births (177) with the statistical value of a life at birth), and estimated to be £25,700.

The NMB can then be calculated using the values from the model predictions, for the difference in neonatal morbidity (ΔE) (effectiveness), the difference in hospitalisations (ΔH) (cost) and the average cost of hospitalisation (C) and incorporating the willingness to pay estimate (λ) into the NMB formula.

Using NMB to calculate sample size

Over the last decade there has been much discussion on appropriate methods for calculating conventional trial sample sizes based on expected cost-effectiveness outcomes (40;46;169-171). Most typically this can be done following the net monetary benefit approach. By using the NMB approach, a sample size for a cost-effectiveness outcome can be calculated in the standard way as typically used for mean effectiveness (72), but based on the expected change in NMB. Briggs et al. (46), provides a thorough explanation of the theory and equations, which are briefly summarised here.

The calculation is based on a hypothesised difference in both costs and effects with associated hypothesised variances and covariance along with a set monetary willingness to pay value (λ). These are used to calculate a hypothesised net monetary benefit ($N\tilde{M}B$). The null hypothesis ($N\tilde{M}B_N$) assumes the net benefit is zero, i.e. there is no difference between the costs and effects of the interventions, while the alternative hypothesis ($N\tilde{M}B_A$) assumes that there is a difference between the costs and effects, i.e. the net benefit is different from zero.

Equation 5.4 illustrates the standard equation to show a hypothesised net monetary benefit ($N\tilde{M}B$) as different from zero. Let z_B represent the critical

value from the standard normal distribution corresponding to a required power of 1-ß and the variance expressions (v) for net-benefit are based on the hypothesised variance in cost and effect, and their covariance. $z_{\alpha/2}$ is multiplied by the variance of the net monetary benefit under the null hypothesis, and z_β is multiplied by the variance of the net monetary benefit under the alternative hypothesis. Note that this is the same as the standard approach for effectiveness, as illustrated in Equations 5.1 and 5.2, with the difference in NMB under the null ($N\tilde{M}B_N$) and alternative ($N\tilde{M}B_A$) hypotheses replacing the difference in effectiveness under the null and alternative hypotheses.

$$N\tilde{M}B > z_{\alpha/2} * v\left(\Delta N\tilde{M}B_N\right) + z_\beta * v\left(\Delta N\tilde{M}B_A\right) \qquad (5.4)$$

Where:

$$NMB = \lambda * \Delta E - C * \Delta H, \qquad (5.3)$$

and:

$$v\left(N\tilde{M}B\right) = \lambda^2 * v(\Delta E) + v(\Delta C) - 2\lambda cov(\Delta E, \Delta C) \qquad (5.5).$$

As the net monetary benefit calculation uses the cost and effect variables, which in this case are proportions, the standard methods are slightly modified for dealing with proportions (72), i.e. the null hypothesis uses the pooled proportions for cost and effect, and the variance for the NMB under the null and alternative hypotheses can differ. A normal distribution was assumed for simplicity, allowing equal variance expressions for the null and alternative hypotheses. Therefore, given Equations 5.4 and 5.5, the variance of the difference in NMB under the null hypothesis is:

$$v\left(\Delta N\tilde{M}B_N\right) = \lambda^2 * \left(\frac{2E(1-E)}{n}\right) + \left(C^2 * \frac{2H(1-H)}{n}\right) - 2\lambda C\rho * \sqrt{\frac{E(1-E)}{n} * \frac{H(1-H)}{n}} , \qquad (5.6),$$

and the variance of the difference in $N\tilde{M}B$ under the alternative hypothesis is:

$$v\left(\Delta N\tilde{M}B_A\right) = \lambda^2 * \left(\frac{E_T(1-E_T)}{n} + \frac{E_C(1-E_C)}{n}\right) + \left(C^2 * \left(\frac{H_T(1-H_T)}{n} + \frac{H_C(1-H_C)}{n}\right)\right) - 2\lambda C\rho * \sqrt{\left(\frac{E_T(1-E_T)}{n} + \frac{E_C(1-E_C)}{n}\right) * \left(\frac{H_T(1-H_T)}{n} + \frac{H_C(1-H_C)}{n}\right)}$$

(5.7).

By incorporating these variance expressions for the hypothesised null and alternative NM̃B (Equations 5.6 and 5.7) into the standard NM̃B significance expression (Equation 5.4) the variance expressions can then be divided by the required number of participants in the control and treatment arms (n), and rearranged to solve for the required sample size (n) to show a significant difference in hypothesised NM̃B. Equation 5.8 illustrates this and is fully expanded in Equation 5.9 to detail the variance expressions.

$$n > \left[\frac{z_{\alpha/2} * v\left(\Delta N\tilde{M}B_N\right) + z_\beta * v\left(\Delta N\tilde{M}B_A\right)}{\Delta N\tilde{M}B}\right]^2 \qquad (5.8)$$

$$n > \left[\frac{(z_{\alpha/2} * \sqrt{\lambda^2 v(E_N) + v(C_N) - 2\lambda * cov(E_N C_N)} + (z_\beta * \sqrt{\lambda^2(v(E_A) + v(C_A) - 2\lambda * cov(E_A, C_A)))}}{\lambda(\Delta E - \Delta C)}\right]^2 \quad (5.9)$$

This is the standard methodological approach for sample size calculations using the NMB approach, but in this case we have used proportions for the cost and effect variables, as opposed to continuous variables. This process was used to determine the sample size required for the fFN trial, in order to detect a net monetary benefit greater than zero. The calculation was based on conventional power and significance levels, taken to be 90% power at the 5% significance level. The inputs for this sample size calculation are the same as those used in the individual sample size calculations, (i.e. the pre-trial model prediction values for morbidity (M, M_T, M_C) (effectiveness), hospitalisations (H, H_T, H_C) (cost), and the average cost of hospitalisation in the model (C)), as detailed in Table 5-4, with the addition of the willingness to pay value (λ), and a correlation value for the covariance (ρ).

Table 5-6 shows the outcomes from this calculation, detailing the sample sizes required in the trial to detect a NMB greater than zero at different levels of power. As would be expected a stronger power for the study requires a larger

sample size per arm. At a 90% power a total sample of 1850 participants would be adequate to detect a NMB greater than zero.

Table 5-6: NMB sample sizes under different power specifications

Power	Sample per arm	Total sample size
90%	925	1850
80%	701	1402
70%	559	1118

Choice of ceiling ratio

As previously discussed the ceiling ratio or 'willingness to pay' value for the NMB expression is a subjective value; there is no universally accepted ceiling ratio which exists for decision making purposes. In the UK, the threshold of £20,000-30,000 per QALY, recommended by NICE (24), tends to be adopted.

The value of a statistical life was used in the fFN pre-trial model to reflect the value of preventing an infant mortality, and used to derive the value of preventing an infant morbidity. This value was set at £1million, however there is considerable debate regarding the approaches used to derive this value (189) and the resultant monetary amount that appropriately reflects this value (190;191). It was felt that a value of £1million would be a realistic reflection of the value of a statistical life for a new-born who has not yet contributed to society, for the fFN case. However, it may have been appropriate to use a value reflecting the cost of initial intensive care for neonates with pre-term morbidity to represent the willingness to pay value. Costs of caring for low birth weight and premature babies are extremely variable with some sources reporting between £720 - £3280 per case depending on severity (192), while others report a mean cost of £5063 (price year 1998), which again varies considerably depending on gestational age (193). Alternatively the monetary value of a QALY (194;195) could be used to reflect the willingness to pay value, rather than the monetary value of a statistical life or a hospitalisation cost of neonatal morbidity.   Using alternative methods to derive a willingness to pay value will generate different sample sizes. Briggs & Gray (170) suggest plotting

the sample size requirements as a function of the maximum cost-effectiveness ratio to explore the impact of the choice of ceiling ratio on the sample size.

In order to explore the structural uncertainty surrounding which approach and alternative values should be used; the sample size calculations were repeated using a variety of QALY thresholds for willingness to pay.

The baseline model developed was based on morbidities avoided, applying a willingness to pay societal value per morbidity avoided. Using this method the cost-effectiveness evaluation was directly related to the proposed trial primary outcome (pre-term morbidity). Alternatively, a QALY model could have been developed. This alternative was considered and a very rough assessment of QALY outcomes incorporated. The present value of a QALY at birth was calculated using 2007 life expectancies for both males and females (196) and UK population norm QALY values (124), discounted at a rate of 3.5% (24). This generated an estimated present value of a new life of 42.7 QALYs. This QALY outcome was used to re-calculate the NMB sample size under different monetary thresholds for a QALY (λ). Table 5-7 details the baseline sample size generated for the fFN RCT using the statistical value of avoiding morbidity (£25,700), alongside the sample sizes generated at alternative thresholds (λ) of £20,000, £30,000, £40,000 and £50,000 per QALY. All calculations were undertaken using a power of 90% and significance at 5%.

Table 5-7: NMB sample sizes under alternative monetary thresholds (λ)

Monetary threshold(λ)	Sample per arm	Total sample size
£25,700†	925	1850
£20,000 per QALY	662	1324
£30,000 per QALY	1537	3074
£40,000 per QALY	2683	5366
£50,000 per QALY	4045	8090

†baseline WTP value to avoid morbidity derived from statistical value of a life at birth (assumed to be £1million)

As the monetary threshold increases, the sample size requirements increase. Increasing the λ value reflects an increase in societies willingness to pay to avoid a neonatal morbidity (each neonatal morbidity has a greater cost to

society), and therefore the cost saving which accrue from fewer hospitalisations with the fFN test become outbalanced by the greater cost to society for the few false negative test cases who do experience a neonatal morbidity. Therefore as the willingness to pay value increases, the fFN test will have a lower NMB gain against current practice, becoming less cost-effective and requiring a greater sample size to show that the fFN test has a significantly positive NMB.

In light of these alternative outcomes, potential funders should consider what society/they consider to be an appropriate ceiling ratio/willingness to pay (accept) value per QALY gained (lost). The current recommended ceiling ratio in the UK is £30,000 per QALY, however, as the cost-effectiveness decision is in the south west quadrant of the cost-effectiveness plane, decision makers may require a much greater cost saving in order to accept a reduction in effectiveness (resulting in a kinked ceiling ratio), as previously discussed. If the willingness to accept a cost saving per QALY loss is greater than £30,000 per QALY, then funders should consider that a trial with a sample size of approximately 2000 is unlikely to show any conclusive results. For example, once funded, such a trial may well require an extension, or may well not show a significant difference. However, given the current ceiling ratio of £30,000 per QALY recommended by NICE in the UK, the sample size required to show a significant difference is feasible.

5.2.3 Non-inferiority approach

Standard approach to non-inferiority

A non-inferiority trial aims to demonstrate the intervention is not worse than the control by more than a pre-specified, small amount known as the non-inferiority margin (NI margin) (172). The NI margin is a criterion that is set so that the new interventions non-inferiority can be judged, i.e. that any difference in effect between the intervention and control is less than or equal to the NI margin, but not greater than this amount. Therefore, non-inferiority designs require the specification of the NI margin in advance of a trial. Setting an appropriate NI margin is essentially an arbitrary judgement; however,

guidelines on the choice of non-inferiority margin (172) are available. These highlight important considerations for the choice, which should be based upon a combination of statistical reasoning and clinical judgement.

In establishing an NI margin for a new intervention, it is common to base the margin on a proportion of the control effect size, for example less than a third of the established effect, and therefore, it is recommended to look historically at placebo-controlled trials for the active comparator and use statistical reasoning to determine an appropriate NI margin (197). The Committee for Medicinal Products for Human use (CHMP)(172) propose that a systematic review be undertaken to identify studies relevant to the comparison of the control treatment with placebo in the disease area being considered. Using such literature, indirect comparisons can then be made with studies comparing the control with a placebo using the lower bound of a 95% confidence interval to establish an efficacy advantage over a placebo.

The CHMP guidelines are concerned with establishing an NI margin based on an effectiveness endpoint, however, from an economic perspective we are interested in both the effect and cost endpoints and therefore acceptance of a new intervention is dependent on 'non-inferiority' in both clinical and cost terms.

Non-inferiority in the fFN trial

Due to the atypical nature of the fFN intervention (cost saving but with a potential marginal reduction in effectiveness), a non-inferiority design may be appropriate for the effectiveness endpoint, which aims to demonstrate that the fFN test is no worse in terms of the clinical outcomes than the control, by more than a pre-specified, small amount. However, in terms of the cost endpoint the intervention is likely to show superiority over the control. The pre-trial model predictions detailed in Table 5-2 and the superiority sample sizes in Table 5-5 illustrate that this is likely to be the case. The trial is likely to show a substantial reduction in the number of hospitalisations, which will have considerable cost saving implications, without having a significant detrimental effect on neonatal morbidity. Therefore, a non-inferiority sample

size calculation for morbidity (combined with a superiority outcome for costs) was undertaken in addition to the previously specified NMB sample size calculation. This 'additional' NI sample size calculation was undertaken to ensure that the sample for the trial was also large enough to show that the test was no less effective than the control within a given cost-effectiveness defined margin of non-inferiority.

The non-inferiority margin (NI_m) was determined based upon a combination of statistical reasoning and clinical judgement, as recommended by the CHMP, using an economic definition of non-inferiority which incorporated both the cost and effectiveness considerations. NI_m was determined from the cost-savings that might accrue through reduced hospitalisations in the intervention arm and the willingness to pay value to avoid neonatal morbidity (λ). The difference in hospitalisations (ΔH) between the intervention and control arms is multiplied with the average cost of hospitalisation (C), to give the cost savings which accrue through fewer unnecessary hospitalisations in the fFN test arm. This is then divided by the value of avoiding neonatal morbidity (λ) to provide the non-inferiority margin from an economic perspective, as illustrated in Equation 5.10:

$$NI_m = \frac{\Delta H * C}{\lambda}$$
(5.10),

i.e. where the net health benefit equals zero $(NHB = 0)$.

For this calculation the model prediction values for the difference in hospitalisations and the average cost per hospitalisation were used, as detailed in Table 5-4, giving an expected saving of £541 per woman tested. Populating Equation 5.10 with these values along with the willingness to pay to avoid morbidity value of £25,700 (λ), a non-inferiority margin of 0.02 was determined. On this basis, an increase in morbidity of up to 0.02 would be considered "acceptable" in the trial. Having specified this non-inferiority margin for morbidity, the standard sample size calculation can be used to determine the sample required to show that an expected difference of 0.00176

(pre-trial model prediction for morbidity, Table 5-2) has an upper confidence limit below 0.02.

This specified non-inferiority margin (NI$_m$) was incorporated into the basic sample size calculation previously detailed in Equations 5.1 and 5.2, utilising the morbidity outcomes (M probabilities: M, M$_T$ & M$_C$) from Table 5-4 as the measure of effectiveness (E, E$_T$ & E$_C$), using 90% power and a significance of 5%. Equation 5.11 details the sample size calculation based on the morbidity non-inferiority margin, and is expanded in Equation 5.12 to detail the variance under the null and alternative hypotheses.

$$n > \left[\frac{z_{\alpha/2} * v(\Delta E_N) + z_B * v(\Delta E_A)}{\Delta E + NI} \right]^2 \tag{5.11}$$

$$n > \left[\frac{(z_{\alpha/2} * \sqrt{2E(1-E)} + z_B * \sqrt{E_T(1-E_T) + E_C(1-E_C)}}{(E_T - E_C) + NI} \right]^2 \tag{5.12}$$

The resultant non-inferiority sample size for morbidity is detailed in Table 5-8, indicating the variation in sample size requirements to changes in the power specification.

Table 5-8: Non-inferiority sample size under different power specifications

Power	Sample per arm	Total sample size
90%	1156	2312
80%	864	1728
70%	679	1358

When the non-inferiority approach is used with the morbidity outcome measure, a trial with 90% power will require a sample size of 2312 participants. This is approximately 500 participants more than that required to show cost-effectiveness using the NMB approach (Table 5-6). Therefore, designing the trial with a total recruitment sample of 2312 participants has a 90% power, which is adequate to detect a NMB greater than zero and non-inferiority given a cost-effectiveness defined non-inferiority margin.

5.3 Discussion

The 'pre-trial' model developed to inform the fFN trial sample size calculations predicted that potential cost-savings were likely, but with a marginal detrimental impact on neonatal morbidity, indicating that a non-inferiority trial design may be appropriate. Therefore, the sample size calculations were undertaken using the NMB approach to demonstrate cost-effectiveness, and also using a non-inferiority design to ensure the trial sample size would be sufficient to demonstrate non-inferiority for the effectiveness endpoint, given a cost-effectiveness defined NI margin. Adopting the NMB approach, a total sample size of 1850 with a power of 90% is required to demonstrate a NMB greater than zero; while a total trial sample size of 2312 would be required to demonstrate non-inferiority, given the cost-effectiveness defined non-inferiority margin. Therefore, the trial was powered at 90% requiring a total sample size of 2312, in order to satisfy both approaches.

This real life application of a sample size calculation based on a cost-effectiveness outcome demonstrates the feasibility of this approach, leading to appropriately sampled trials with respect to the cost and cost-effectiveness outcomes. While calls for health economic involvement in trial design are abundant (39;44;169-171), very few trials which involve economic evaluation as a primary outcome actually undertake sample size calculations based on the economic outcome.

In terms of practicality, the skills required for building a simple, early model to inform trial design and a funding application are feasible to health economists who undertake economic evaluations. The skills required for calculating sample sizes, however, are slightly less generalisable. Many health economists may not normally become involved with sample size calculations, or indeed the wider trial design, leaving the sample size aspect up to trial statisticians. However, given that the methodology for such an approach is now well established, and that the role of the health economist is becoming ever more statistical (particularly for economic evaluations undertaken alongside clinical trials) (40) , economic led sample size calculations are not such a barrier for

health economists as they might once have been. As recommended in good practice guidelines, it is important that the economic requirements contribute to the study design to ensure that the structure of the trial will provide the necessary data for a high quality economic study (39).

It is encouraging that in this case study the health economics perspective was included in time to influence the sample size calculation; however, this was possibly only due to the atypical context of this situation, i.e. a trade-off between cost savings and a reduction in effectiveness, as opposed to the typical trade-off between additional costs and an improvement in effectiveness. Had this been a typical scenario, the cost-effectiveness perspective may well have been ignored until after the design of the trial. In trials where the cost-effectiveness trade-off is expected to be straightforward the design tends to focus solely on effectiveness, overlooking the importance of the cost-effectiveness outcome and not even considering whether a cost-effectiveness based sample size may be appropriate. It should be further promoted and recognised within a wider clinical audience that these techniques are valid methodology for calculating cost-effectiveness based sample sizes in both typical and atypical situations.

Diagnostic test accuracy evaluations present the possibility for litigation proceedings to the NHS, due to false negative test outcomes (failing to detect disease). Current NICE guidelines (24) recommend all relevant costs be incorporated into an evaluation, and in the case of DTA studies (particularly one such as fFN) it could be argued that litigation costs are likely to have a significant impact on cost the of introducing the fFN test to the NHS, given the probability for false negative outcomes and the increasing tendency in the UK for claims against the NHS. The potential importance of litigation costs for the DTA evaluation could be explored in a scenario, by making an assumption about the proportion of false negative cases which result in morbidity that will proceed to legal/court proceedings and include the additional litigation costs. It is indeed worthwhile exploring the importance of and potential impacts of this on the model outcomes in a scenario analyses.

The design and development of trial protocols are typically undertaken in short periods of time in order to meet tight submission deadlines; however this can impact on the quality of the trial design. In the fFN case study discussed in this chapter, a pre-trial model was developed over a relatively short time period utilising the most readily available secondary evidence in order to undertake sample size calculations. This demonstrates that even without prior funding or support to develop a comprehensive decision analytic model, a simple DAM can be constructed relatively quickly which can be used to help design the study.

A basic, deterministic model can still have considerable advantages (in terms of evaluating alternative endpoints of interest to select the most appropriate for the trial, and enabling sample size calculations based on economic endpoints) within the context of a standard frequentist trial design. Preferably uncertainty in the model and sample size estimates should be considered. Briggs & Gray (170) recommend increasing the sample size of the study beyond that suggested by the formulas in order to account for additional uncertainty to be quantified in sensitivity analysis. They also propose increasing upon the initial recruitment target to account for potential drop-out in the study, as is standard practice when sample size calculations are based on effectiveness alone. The role of uncertainty in sample size calculations has been given further attention in published literature and it has been proposed that probabilistic sensitivity analysis is a good way to deal with this uncertainty (198), however, in practical terms the time constraints imposed by protocol submission deadlines may interfere and dealing with this uncertainty may only be feasible after protocol submission.

If a non-inferiority design is appropriate for a study, as opposed to the typical superiority design, then there is a further role for economists (or the research team) in defining an appropriate non-inferiority (NI) margin. While there are general guidelines to adhere to in choosing a non-inferiority margin (172), based upon a combination of statistical reasoning and clinical judgement, the approach taken will also need to incorporate economic considerations (for both cost and clinical outcomes) of non-inferiority when the trial is undertaken to determine cost-effectiveness. The NI margin is subjective by nature and for

example, different values used to reflect the value of a statistical life will generate different sample sizes. The structural uncertainty surrounding which approach should be used to determine this value therefore has implications for the sample size estimate.

There is also scope for assessing how variations in required sample sizes will affect recruitment costs and feasibility for a trial. The greater the sample size, the more precise the estimates of the parameters and their differences, but the more expensive the trial will be. Exploring the specific trade-off involved can be useful in determining the appropriate balance between precision and availability of resources, however, this line of enquiry leads back to the methodological advantages of using value of information analyses. With regards to 'frequentist' type sample size calculations, uncertainty can be incorporated and there should be a trade-off considering the feasibility and cost of recruitment as well as the sample size to show a significant difference. However, when you are starting to formally consider the trade-offs in terms of the size, cost and feasibility of recruitment, the VOI approach is much more appropriate. Rather than implicitly considering the trade-offs between size, cost and feasibility – which is the current state of play in practice– research funders could more formally be incorporating these trade-offs, explicitly considering further research on the basis of how it addresses current uncertainty and valuing it on that basis. Expected value of sample information and ENBS as part of an iterative economic approach explicitly incorporates these considerations and trade-offs, rather than conventional trial design as explored in the case study for this chapter.

The feasibility or cost of recruitment may also become an issue if a regulatory or funding body dictate what the power endpoint of interest should be for a trial and it yields an unrealistic sample size. This was not the case in the fFN trial design as the researchers were given freedom to explore various endpoints (morbidity, mortality and hospitalisations) and also to determine the most appropriate method for the evaluation, i.e. a net monetary benefit approach or a non-inferiority design.

In the context of this thesis it would have been preferable to develop a comprehensive decision analytic model and undertake EVSI following an iterative approach to trial design; however, this was not feasible within the context of the case study research bid. This chapter demonstrated that even without prior economic involvement (or funding) to develop a comprehensive decision analytic model, a simple DAM can be constructed relatively quickly (and fairly simply) which can be used to help design the study. Given that a basic, deterministic model such as that developed for the fFN trial can help in the design and development of a standard frequentist trial, there is no reason for economists not to be involved in the design stage of a trial, when the purpose of that trial is to determine cost-effectiveness. The fFN model helped determine the power and sample size for the study; exploring both a NMB approach and a non-inferiority approach for sample size calculation and demonstrating their feasibility, while simultaneously highlighting limitations and demonstrating the role for economic considerations to guide non-inferiority. Considerations of economic assessments alongside clinical trials can and should be used to guide conventional trial design.

6 Designing trials following an iterative approach

Following on from the fFN case study in the previous chapter it is of interest to consider whether the same, or an alternative conclusion would have been drawn had the iterative approach to health technology assessment been adopted. Under an iterative approach, formal evidence synthesis and early decision analytic modelling would have been undertaken in advance of designing the trial, helping to explore and set the research priorities. Following the development of an early probabilistic DAM, expected value of perfect information (EVPI) analysis could have been undertaken to explore whether further research was potentially worthwhile, and if so, expected value of perfect parameter information (EVPPI) analysis could have indicated the appropriate type of any further research (i.e. an RCT or observational study) based on the parameters that were driving uncertainty in the model outcomes. If a study was deemed potentially worthwhile then expected value of sample information (EVSI) analysis could be used to determine an appropriate sample size for the study.

As discussed in Chapter 5, the UK Health Technology Assessment (HTA) programme (50) had issued a commissioning call for research to undertake a randomised controlled trial to explore the cost-effectiveness of fFN testing in women threatened with pre-term labour. Therefore the desire for a study and specification for an RCT had already been planned without prior economic involvement. The trial was under design with a planned duration of three years costing approximately £1.65 million, when the opportunity arose to power the trial for economic endpoints, in the context of a frequentist trial design. This chapter explores whether the same, or an alternative conclusion (in terms of the desire for an RCT, and the appropriate sample size) would have been drawn had an iterative approach been followed whereby a DAM had been developed in advance and explored to its full potential.

This chapter undertakes a re-analysis of the fFN study, adapting it so that the cost and effect estimates are determined probabilistically. The probabilistic results are then used to explore the EVPI and EVSI to compare the resultant research priorities and sample size under an iterative approach with the outcomes that were determined through conventional methods for the trial proposal, as detailed in Chapter 5. This chapter then discusses the practicalities of employing the EVSI approach and some of the complexities.

6.1 Re-analysis of the fFN case study

6.1.1 Probabilistic analysis of fFN model

With regards to the fFN model developed in Chapter 5, naturally there is uncertainty surrounding the parameter point estimates in the pre-trial model; and variability in the willingness to pay value (λ) for the NMB calculation and the non-inferiority margin which are subjective (172), leading to uncertainty surrounding the model outcomes and variation in the sample size estimates. Therefore, it is appropriate to undertake a probabilistic sensitivity analysis (PSA) on the baseline model to reflect uncertainty in the model parameters and explore the impact that this has on the predicted net benefit for both the treatment and control arms, as well as allowing further analysis on the value of further information.

The parameters from the pre-trial model for which we are most uncertain, are the diagnostic test accuracy (sensitivity and specificity) of the fFN test, the risk of hospital transfer, the risk reduction of pre-term morbidity provided by steroids and the probability of hospitalisation given the fFN test results. The model in Chapter 5 used point estimates of these parameters to predict deterministic estimates for the outcome parameters of interest: the proportion of hospitalisations and the proportion of neonatal morbidity experienced in each arm. The deterministic model parameters as detailed in Table 5-1 were used to produce the predictions detailed in Table 6-1.

Table 6-1: fFN model predictions

Parameter	Point estimate	Source
fFN test: Proportion hospitalised	0.52423	model prediction
No test: Proportion hospitalised	0.90000	model prediction
fFN test: Proportion morbidity	0.03036	model prediction
No test: Proportion morbidity	0.02860	model prediction

fFN = fetal fibronectin

If a trial were undertaken, the model predictions are the key parameters which would be informed. Therefore, for the purpose of the PSA and VOI analyses, these four key parameters of interest (proportion hospitalised and proportion of pre-term morbidity, for each arm) were assigned distributions. In this way the model was kept simple which was advantageous for undertaking the EVSI calculation.

The four model prediction parameters are represented by proportions, so when considering distributions for the probabilistic analysis, Beta distributions were deemed appropriate (as detailed in Chapter 2). There was no prior study information from which to determine the alpha and beta parameters. Ideally formal elicitation of clinical opinion could have been undertaken (for example, using Delphi techniques), to inform the uncertainty surrounding the parameter mean estimates, however, due to time constraints a best guess approach was used. Wide confidence intervals were applied to handle the resultant uncertainty in the model parameters in the probabilistic sensitivity analysis. Standard errors were assumed and the alpha and beta values were derived using the Method of Moments (4), as detailed in Chapter 2. A standard error of 20% of the mean value was assigned to the probability of pre-term morbidity in the model for both arms and to the probability of hospitalisation under the treatment (fFN test) arm. There is less uncertainty about the probability of hospitalisation in the control arm, as current practice is an admit all approach, therefore a smaller standard error of 6% of the mean value was assumed to represent the uncertainty for this parameter. Table 6-2 details the mean values and probabilistic details for the four key parameters which would be determined in a trial.

Table 6-2: fFN Parameters for probabilistic model

Parameter of interest	Mean value	Standard error	Distribution	α	β
Hospitalisation treatment arm	0.524	0.105	Beta	11	10
Hospitalisation control arm	0.9	0.05	Beta	31	3
Morbidity treatment arm	0.03	0.0061	Beta	24	773
Morbidity control arm	0.029	0.0057	Beta	24	824

The PSA used Monte Carlo simulation to draw randomly from the specified beta distributions for the proportion of hospitalisations and proportion of neonatal morbidity in each arm, and generate 10,000 probabilistic results. The average cost of hospitalisation was combined with the proportion of hospitalisation in each arm to determine the cost for each of the iterations, from which the incremental cost could be determined. The incremental effect was determined as the difference in probability (neonatal morbidity) between each arm, for each of the iterations.

Figure 6-1 depicts the resultant cost-effectiveness plane, plotting the incremental costs and incremental effects, for each of the 10,000 PSA iterations. It illustrates that we can be almost certain that the treatment will reduce cost in comparison to control (only six of the iterations do not fall below the x axis), but as the probabilistic results cross through the y axis, we can not be certain that the two interventions differ in effectiveness. The x axis in Figure 6-1 illustrates incremental effectiveness, where an improvement in effectiveness means a reduction in pre-term morbidity. Therefore, points in the western quadrant reflect a reduction in effectiveness, where the treatment arm has a greater proportion of pre-term morbidity than the control arm; whereas points in the eastern quadrants reflect an improvement in effectiveness, i.e. where the treatment arm has less morbidity than the control arm. The spread across the y axis is not unexpected in this non-inferiority type situation, whereby we would expect a marginal reduction in effectiveness (increase in morbidity) in the treatment arm, within the pre-specified margin of non-inferiority (as discussed in Chapter 5).
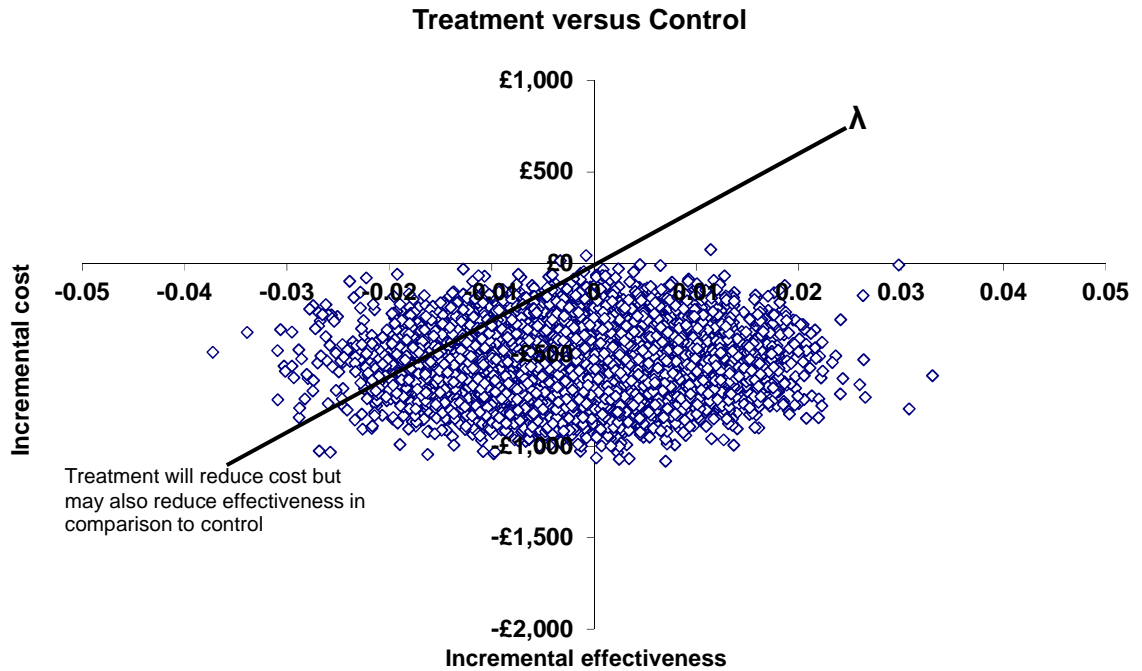
Figure 6-1: The CE plane for Treatment (fFN test) versus Control (no test)

By employing the NMB approach as discussed in Chapter 5; the impact on effectiveness can be monetarised using a willingness to pay value (λ) to avoid morbidity, in this case the pre-defined value of £25,700 (as detailed in Chapter 5). Given this monetary threshold (λ), in Figure 6-1 all points that fall to the right of the monetary threshold (λ) can be considered cost-effective, while all points that fall to the left are not cost-effectiveness. The NMB for each intervention was calculated for all 10,000 iterations. The expected net benefits under the treatment and control arms were then calculated, that is, the average net benefit in each arm across the 10,000 iterations. The treatment arm had the greatest ENB (£26) compared to the control arm (-£559), and would therefore be the treatment of choice, maximising NB at the willingness to pay (λ) value of £25,700. It should be noted that in Chapter 5 it was the incremental net monetary benefit that was reported, i.e. the difference in net benefit between the treatment and control. Given the ENB for the treatment and control arms, the incremental net monetary benefit from the probabilistic analysis is £584, and therefore the treatment arm is the optimal strategy. In the fFN case study the NBs for both interventions are low and sometimes negative, which is due to the nature of the fFN case study, i.e.

in comparison to the control the treatment arm leads to cost savings in terms of hospitalisations avoided, but at the expense of a marginal reduction in effectiveness (some additional pre-term morbidity).

Figure 6-2 illustrates the cost-effectiveness acceptability curve, showing the probability of the Treatment and Control interventions being cost-effective at different willingness to pay values for avoiding morbidity.
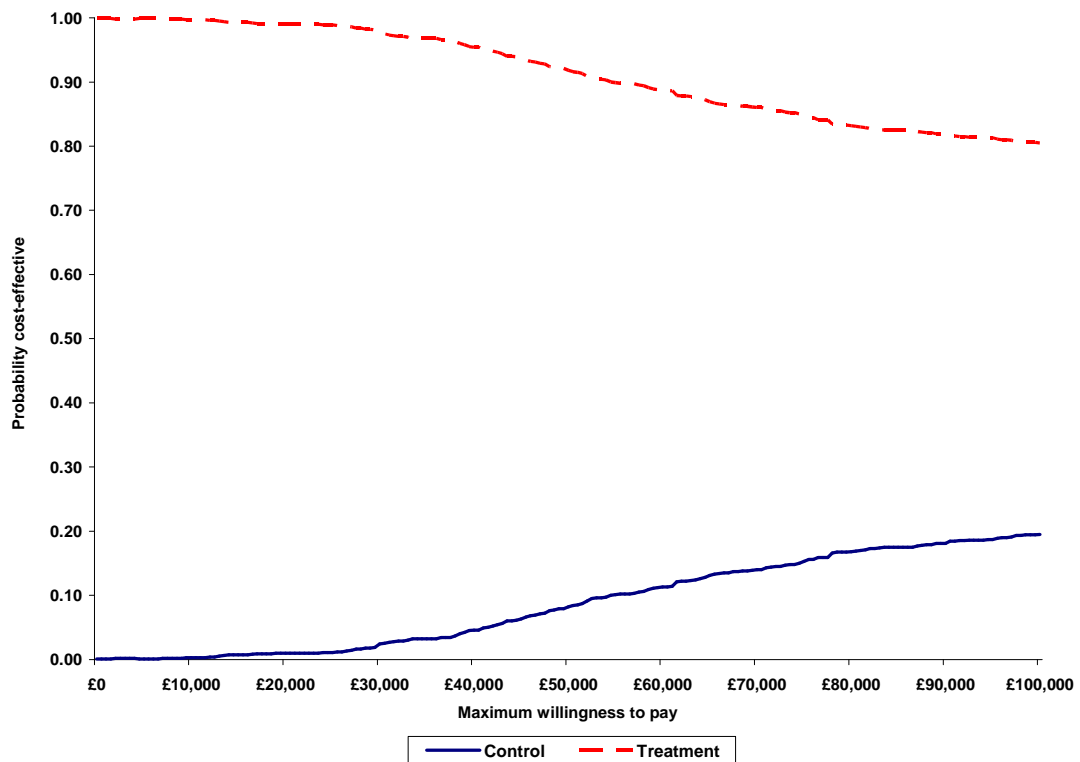
Figure 6-2: The CEAC for fFN

At a willingness to pay value (λ) to avoid a case of morbidity of £30,000, the probability that the Treatment intervention is cost-effective is 98% and there is only a 2% probability that the Control is cost-effective. Even if the ceiling ratio per pre-term morbidity is raised to £50,000, there remains a 92% probability that the Treatment will be cost-effective compared to the control. As the ceiling ratio increases, the probability of Treatment being the most cost-effective intervention falls, however, at a value of £100,000 per pre-term morbidity, the curves flattens out and there is still an 80% probability that

Treatment is the cost-effective choice, with only a 20% probability that the Control is cost-effective.

The CE plane in Figure 6-1 illustrated considerable uncertainty as to whether there was a difference in effectiveness, which is what we would expect in this non-inferiority type case. However, when the willingness to pay threshold of £25,700 per morbidity avoid (λ) was added to Figure 6-1, it could be seen that the majority of points fall to the right of this threshold, indicating that they would be considered cost-effective. Following this, the CEAC in Figure 6-2 illustrates that there is little decision uncertainty, i.e. we can be confident that the Treatment is cost-effective in comparison to the Control over a wide range of ceiling ratios. At the ceiling ratio of £25,700 per morbidity, the CEAC illustrates a 98.8% probability that the Treatment is cost-effective. Utilising the probabilistic model results further, an expected value of perfect information analysis could be undertaken to explore whether it is worthwhile undertaking further research or make the decision to adopt or reject the Treatment intervention based on the current information.

The EVPI was calculated (as described in Chapter 2) and at a ceiling ratio of £25,700 per morbidity the expected value of perfect information is £2.10 per decision/patient. At a ceiling ratio of £30,000 the EVPI per patient is £3.13. Even if society were willing to pay a maximum of £50,000 per neonatal morbidity avoided, the EVPI per patient would only be £15. These EVPI values are very low, and reflect the low level of uncertainty in the cost-effectiveness decision that was illustrated in the CEAC. Figure 6-2 illustrated that at a ceiling ratio of £25,700 there is a 98.8% probability that Treatment is cost-effective. Given this very low uncertainty surrounding the cost-effectiveness of the optimal intervention (Treatment), the expected value of further information to aid the decision is low.

The EVPI can be scaled up to the population of patients, multiplying the EVPI per person by the population over the technology life time, and can then be interpreted as the maximum amount the health care system should be willing to pay for additional evidence to inform this decision in the future, i.e. an upper bound on the value of conducting further research (4;86). The

population EVPI for fFN was calculated, based on an annual population of 120,000 suspected pre-term labour cases (181). A four year technology lifetime was assumed on the basis that this is the appropriate time period (until it is replaced by another more accurate test). Discounted at 3.5% (24) this gives an effective population who could potentially benefit from further information regarding the fFN test of 456,196 women suspected of pre-term labour. The EVPI for the population is depicted in Figure 6-3.
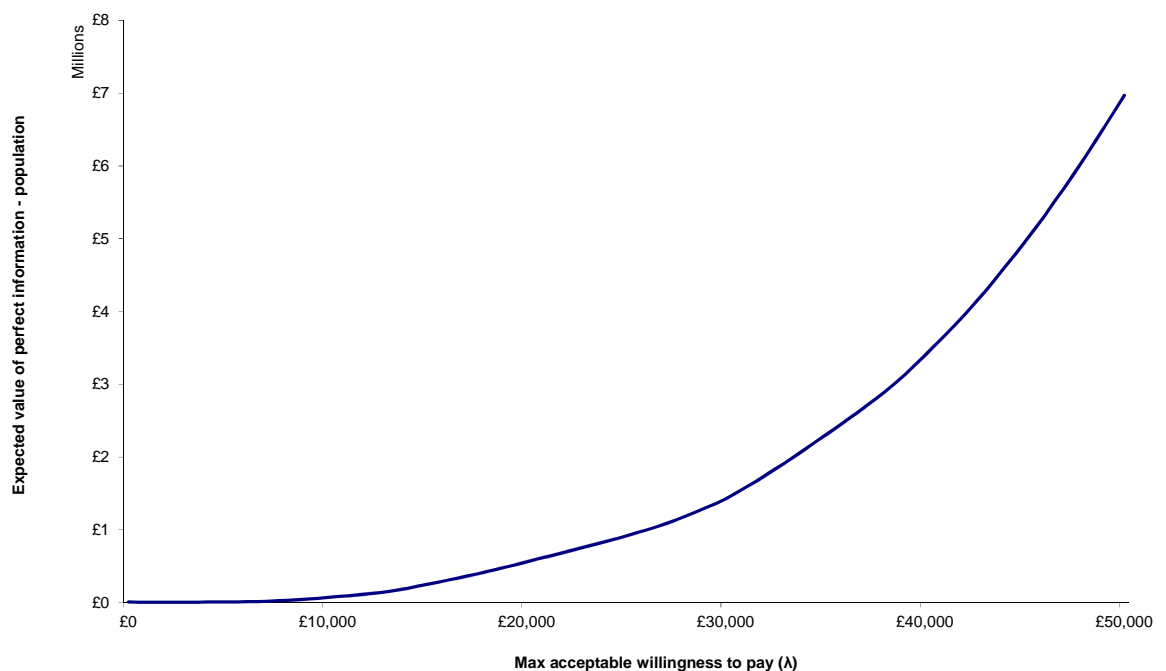
Figure 6-3: EVPI for fFN – population level

Figure 6-3 indicates that at a willingness to pay of £30,000 per morbidity avoided the EVPI for the population is approximately £1.4 million; therefore, research which costs less than this value is potentially worthwhile. However, when this ceiling ratio is lowered, i.e. at a monetary threshold of £10,000 per morbidity avoided, the EVPI is only £70,000. At this lower EVPI value the cost of further research (such as an RCT) is likely to exceed the EVPI and therefore would not be worthwhile. Implementing the pre-specified ceiling ratio of £25,700 as derived in Chapter 5, the EVPI is approximately £957,000. If further research costs more than £957,000 then it is not worthwhile and current evidence can be regarded as sufficient to support the decision to accept fFN

technology. However, if research costs less than £957,000 then it may potentially be worthwhile undertaking that research in order to reduce uncertainty in the decision to adopt or reject the fFN test.

If the maximum that society is willing to pay for avoiding a pre-term morbidity is £25,700, then the EVPI of £957,000 can be interpreted as the maximum amount the health care system should be willing to pay for additional evidence to inform the decision regarding the fFN test. The fFN trial proposal estimated the cost of the proposed research to be approximately £1.65million, which is substantially greater than the maximum willingness to pay for the proposed research. Even at a greater ceiling ratio of £30,000 per morbidity avoided the population EVPI was £1.4 million, which is still lower than the research cost of £1.65 million. Therefore it would be concluded that further research will cost more than the maximum amount the health care system is willing to pay for additional evidence, and therefore the fFN test should be adopted based on current evidence with no further research.

This conclusion contradicts the funding body call for an RCT in fFN. Under an iterative approach the research priorities indicate that further research is unlikely to be worthwhile and therefore the research process would end, without undertaking any further value of information analyses. An EVSI would only be undertaken if the EVPI had exceeded the fixed costs of research, i.e. if further research was potentially worthwhile. However, for the sake of this thesis; it is of interest to demonstrate the EVSI process that could have been followed had a trial been deemed potentially worthwhile and again compare the resultant sample size with those determined through conventional methods for the trial proposal, as detailed in Chapter 5.

6.1.2 EVSI for the fFN case study

In order to calculate the EVSI for the fFN case study, the steps and algorithms outlined in Ades, et al. (83) were followed. The fFN pre-trial model was a simple linear model with four key parameters of interest, as described earlier in this chapter. Therefore the EVSI algorithms for a linear model were used, i.e. utilising the posterior estimates directly to calculate net benefits, rather than re-running the probabilistic Monte Carlo simulation to sample from the posterior distribution as would be required in a non-linear model. The EVSI process is based around specific parameters of interest, which are the uncertain parameters (θ_i) that require further information from the proposed piece of research. The calculation undertaken is now outlined as follows:

1. A probabilistic sensitivity analysis (PSA) was undertaken on the model: Monte Carlo simulation generated 10,000 prior estimates (p_1 to $p_{10,000}$) for the four uncertain parameters of interest.

2. A sample size (N) was assumed for a new piece of research.

3. Data collection was simulated for the new sample size (N) from the binomial likelihood using the prior estimate for the parameter(s) of interest from the first iteration of the PSA in step 1 (p_1).

    - Draw 1 sample result (x) from the binomial likelihood: $\beta \sim (n,p)$ where n is the new sample size and p is the prior probability estimate from the Monte Carlo simulation.

    - This draw represents the number of events (x) which can then be used to calculate a posterior probability.

4. The prior estimate and the simulated number of events (x) are then used to calculate the Beta posterior distribution $\beta(\alpha + x, \beta + N - x)$ for the parameter(s) of interest. The posterior estimate (P_{post}) for the parameter(s) of interest can then be derived, as detailed in Equation 6.1.

$$p_{post} = \frac{(\alpha + x)}{\alpha + \beta + N}$$ (6.1)

5. The next step is to put the posterior probabilities (P_{post}) for the parameter(s) of interest back into the model (replacing the prior estimates) in order to recalculate net benefit for treatment and control (NB_T NB_C) using the posterior probabilities for the parameters of interest and the prior probabilities of the other parameters in the model. The NBs for each intervention were stored.

6. Steps 3 to 5 were repeated for each prior estimate from the PSA in step 1. I.e. in turn using p_2, p_3....$p_{10,000}$ in step 3 to simulate new data (x_2, x_3...$x_{10,000}$) and calculate the Beta posterior and net benefit for treatment and control each time (NB_{T2} NB_{C2}, NB_{T3} NB_{c3}....$NB_{T10,000}$ $NB_{C10,000}$).

7. The Expected Net Benefit for Treatment and Control was calculated across the 10,000 posterior net benefits, as illustrated in Table 6-3 to establish the which intervention is the optimal choice $\left[\max(ENB_T : ENB_C)\right]$ given current information.

Table 6-3: Net Benefits from posterior calculation

Net Benefit Treatment	Net Benefit Control	Maximum Net Benefit
NB_{T1}	NB_{C1}	$\max(NB_{T1}:NB_{C1})$
NB_{T2}	NB_{C2}	$\max(NB_{T2}:NB_{C2})$
NB_{T3}	NB_{C3}	$\max(NB_{T3}:NB_{C3})$
......	....	......
$NB_{T10,000}$	$NB_{C10,000}$	$\max(NB_{T10,000}:NB_{C10,000})$
ENB_T	ENB_C	E max(NB)

8. For each of the individual iterations the intervention which had the maximum net benefit was chosen and stored, essentially making the optimal decision based on sample information for each of the iterations. However, we still do not know the actual results of the sample in advance, and therefore the expected value of a decision taken with sample

information is found by averaging the maximum net benefits over the distribution of possible values (4). The average (or expectation) of the maximum NBs is calculated [E max(NB)], as illustrated in Table 6-3.

9. The intervention which has the greatest ENB across the 10,000 PSA outcomes is the intervention of choice under current information. The average of the 10,000 maximum net benefits is the expected value of a decision made with sample information. Subtract the ENB of the intervention of choice from the maximum ENB; this is the expected value of sample information, as illustrated in Equation 6.1:

$$EVSI = E\big[max\big(NB_{T[i]}, NB_{C[i]}\big)\big] - max\big[ENB_T : ENB_C\big]$$ (6.1),

or

$$EVSI = E_\theta\big[max_t\big(E_\theta NB(t,\theta)\big)\big] - max\big[E_\theta NB(t,\theta)\big]$$ (6.2).

In order to undertake the EVSI calculations in a timely manner, the fFN model was re-programmed into the programming language FORTRAN and the EVSI calculation was undertaken for sample sizes n = 1, 10, 20, 50, 100, 150, 250, 500, 750, 1000, 2000, 3000; then it was re-run for all possible sizes between 1 and 50,000 to explore where the maximum EVSI was reached.

6.1.3 EVSI Results

Table 6-4 presents the EVSI results for a range of sample sizes. All values are calculated for a willingness to pay value (λ) of £25,700 as specified in Chapter 5. Table 6-4 also presents the probability that further research of size n would change the current decision regarding the cost-effective intervention (Treatment). This is calculated using a similar process to that of the cost-effectiveness acceptability curve, detailed in chapter 2. For the set ceiling ratio and sample size n, the probability that each intervention will be cost-effective under perfect sample information is calculated and the optimal

strategy given perfect information is chosen. Given that Treatment was the optimal strategy under current information, the probability of a change in decision is equivalent to the probability that Control is the optimal strategy given perfect information, i.e. 1 minus the probability that Treatment is cost-effective under perfect sample information.

Table 6-4: EVSI calculations for a range of sample sizes

Sample size (n)	EVSI per person	EVSI population	Probability of change in decision
0	£0.00	£0	0.0000
20	£0.00	£0	0.0000
80	£0.00	£904	0.0001
100	£0.00	£2,131	0.0001
150	£0.01	£6,334	0.0003
200	£0.02	£10,935	0.0005
250	£0.04	£17,333	0.0007
500	£0.12	£53,973	0.0020
750	£0.22	£98,226	0.0031
1000	£0.30	£136,234	0.0041
1500	£0.44	£200,966	0.0058
2000	£0.55	£249,086	0.0069
3000	£0.70	£319,945	0.0083
5000	£0.88	£399,780	0.0099
10000	£1.04	£474,983	0.0114
20000	£1.13	£516,572	0.0121
30000	£1.17	£532,408	0.0124
40000	£1.18	£540,431	0.0125
50000	£1.20	£545,446	0.0126

As the proposed sample size increases from 0 to 50,000 the value of sample information increases from 0 to £1.20 per person. The EVSI results from this case study are extremely low, showing that a sample size of 500 is only valued at 12 pence per person, and even with a sample of 2000 participants in the treatment arm, the value is only 55 pence per person. As with EVPI, the per person/decision level of EVSI can be scaled up by the estimated population (here judged to be 456,169 women over a four year period) in order to determine the population level EVSI. At a sample size of 3000 the EVSI per person is only 70 pence; however this translates into a population level value of nearly £320,000.

The EVSI is clearly related to the probability that the proposed study will change the decision from that under current information; identified in Table 6-3 as iterations where: $\max(NB_T:NB_C) \neq \max(ENB_T:ENB_C)$. In Table 6-4 the very low probabilities for a change in decision explain the extremely low EVSI results. The probability that a study with 50,000 participants will change the decision on cost-effectiveness (from the treatment strategy being optimal) is only 0.013 and therefore, the value of such a study is also very low, £1.20 per person. Figure 6-4 plots how the EVSI changes as the sample size increases from 0 to 50,000 in the treatment arm at a monetary threshold of £25,700. The EVSI is very low at all sample sizes, with the curve continuing to rise (albeit very slowly) beyond a sample size of 40,000 and 50,000.
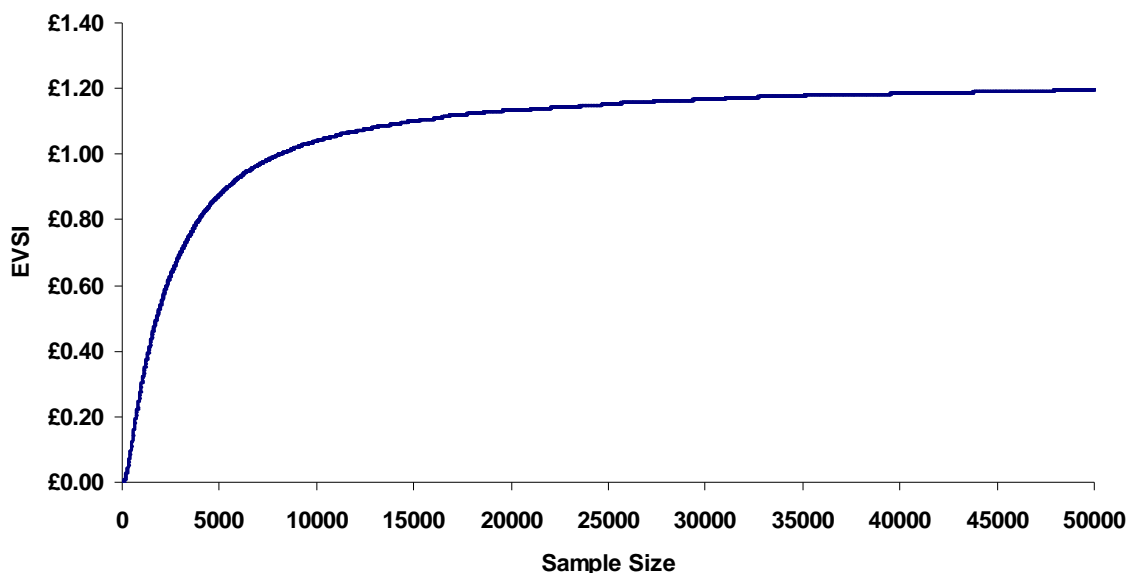
Figure 6-4: EVSI with varying sample size – per patient/decision

The shape of the curve is as would be expected, with the EVPI increasing with sample size but with diminishing returns (85;199), however, as sample size tends towards infinity the EVSI should tend towards the value of EVPI (200) and level off. The EVPI per person at the monetary threshold of £25,700 (λ) is £2.09, and as can be seen from Figure 6-4, at a sample size of 50,000 the EVSI per person is only £1.20. The value of sample information is still increasing beyond a sample size of 50,000, but at a diminishing rate, and therefore a sample much greater than 50,000 would be required to reach the EVPI value of £2.09. The EVSI calculation for this example was not re-run for sample sizes

beyond 50,000 in order to limit the computational time, however, it can be concluded that a sample size beyond 50,000 in each arm of an RCT is unrealistic in terms of patient recruitment.

Figure 6.5 presents the population level EVSI (assuming an annual incidence of threatened pre-term labour of 120,000 with a technology lifetime of 4 years discounted at 3.5%).
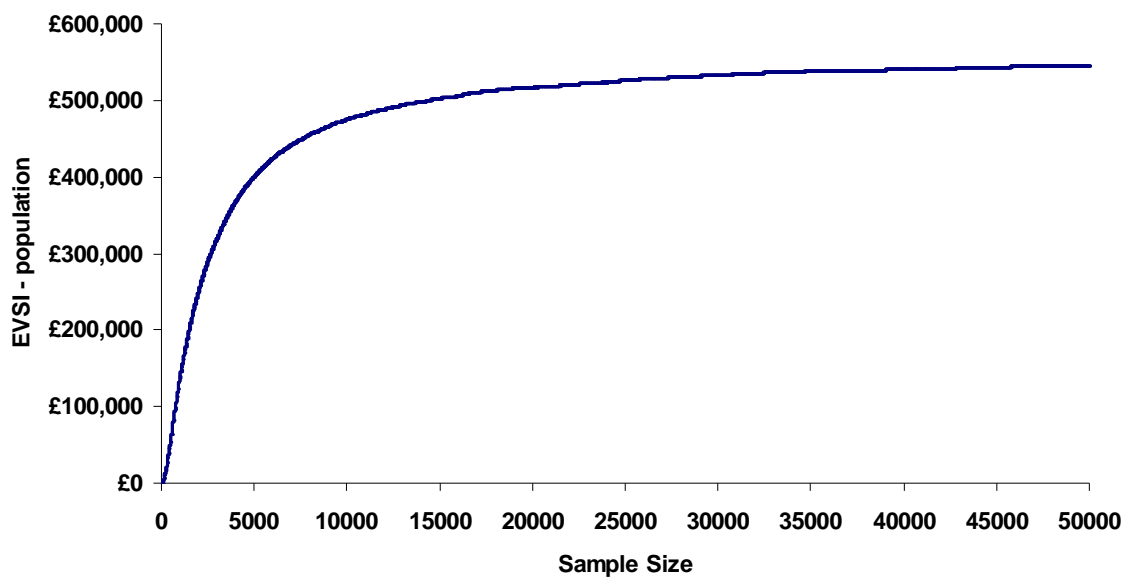
Figure 6-5: EVSI with varying sample size – population level

With an effective population of 456,196 women, the population level EVSI is £136,000 for a sample size of 1000 per arm, while a sample size of 2000 per arm gives a value of £249,000 at the population level.   The EVSI at the population level is the same shape as that for the EVSI per decision/patient, and again the value of sample information is still increasing beyond a sample size of 50,000, but at an ever diminishing rate.

The difference between the EVSI and the cost of acquiring the sample information is the expected net benefit of sample information (ENBS) (9;86); with the optimal sample being the value of N that generates the maximum expected net benefit of sampling.   The cost of the fFN trial as detailed in Chapter 5 was budgeted at £1.65million and therefore in order to generate an

expected net benefit with sample information, the EVSI would need to be greater than this value.

In order to plot the cost of the research and the resultant ENBS for the fFN trial, the costs from the commissioned research proposal were consulted. There was no maximum budget set by the funders, and therefore the cost of the research was established based on the research professionals' time and estimates of the resources involved in the trial. The total cost of the proposed three year RCT aiming to recruit 2312 participants (1156 per arm as calculated in Chapter 5) was £1.65 million. The variable costs incurred in a trial would incorporate the cost of the test (which at £50 a test would only be used in the participants randomised to the intervention arm, and would therefore cost £25 per participant), plus an additional £5 per person to cover administration. Therefore the variable costs can be assumed to be approximately £30 per additional participant recruited. The fixed costs of the trial can be assumed to be approximately £1,600,000 with a sample size of zero (to cover project set-up, research staff time, travel etc). Figure 6-6 plots the EVSI at the population level, along with the cost of the research and the resultant ENBS. The figure illustrates that the cost of the research exceeds the value, at all potential sample sizes.
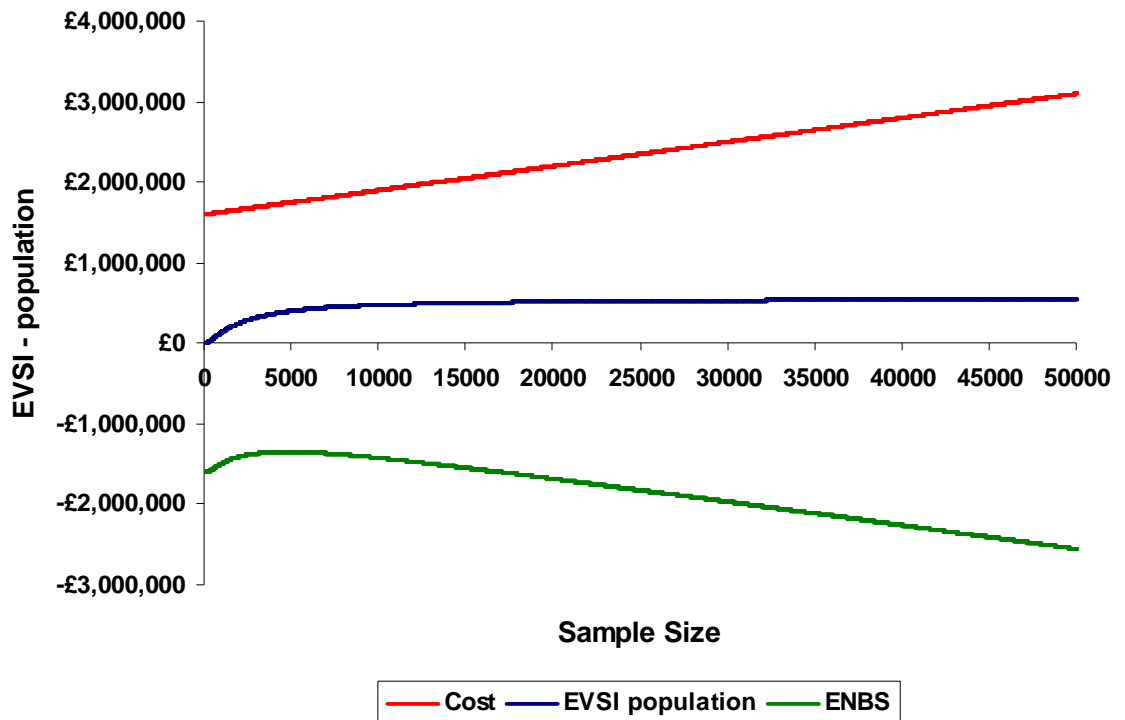
Figure 6-6: EVSI, Cost & ENBS for fFN trial

Figure 6-6 does not display the typical illustration of ENBS and cost in relation to the EVSI. The cost of the research is greater than the EVSI at a sample size of zero (as is typically the case), however the EVSI rises with sample size but in this case the EVSI never reaches the cost of the trial and therefore the expected net benefit of sampling (which is the EVSI minus the cost) remains negative regardless of the sample size. The sample size where ENBS is at its highest is a sample of 5000 participants per arm, at an ENBS of -£1.35 million. As this value is negative, clearly the optimal option would be not to undertake a trial of this magnitude (i.e. a sample size of zero).

Out with the demonstration of EVSI for this thesis chapter, an EVSI would not have been undertaken in the fFN case study as it was found not to be worthwhile at the EVPI stage (EVPI < fixed cost of research). To illustrate how the EVSI and ENBS curves would typically look, let us assume that the cost of research was lower than the EVPI value of £957,000 (i.e. much less than that for the fFN case study). Had a six months observational study of fFN been commissioned (rather than a three year RCT) the cost of research may have

had a fixed cost of £100,000, with an additional variable cost of £10 per participant recruited. If this were the case, the trial could potentially have been worthwhile, and the ENBS may have taken a more typical form, as illustrated in Figure 6-7.
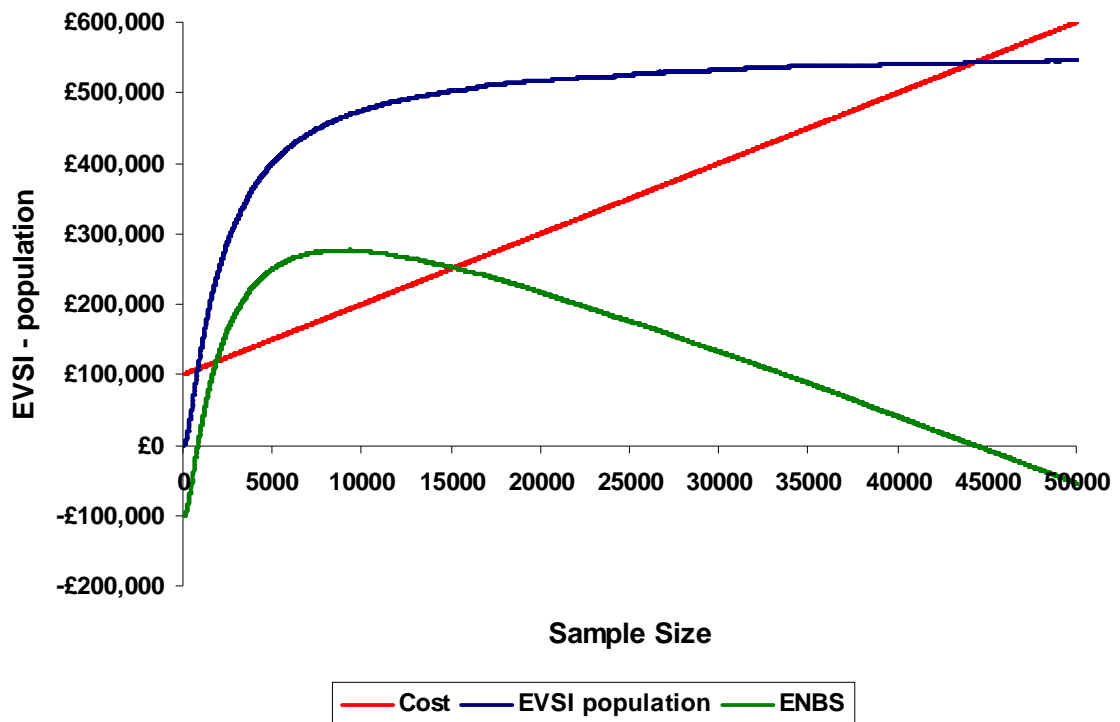
Figure 6-7: Alternative EVSI, Cost & ENBS for an observational study

Figure 6-7 illustrates the hypothetical case where the fixed cost of the proposed research was £100,000 with a small variable cost of £10 per additional person recruited to the sample. At the outset, where the sample size is zero, there is no value in the research and the ENBS is negative due to the fixed £100,000 cost of the trial. As the sample size increases so does the EVSI, and at a sample size greater than approximately 900 participants the EVSI exceeds the cost of the research, and at this point the ENBS is no longer negative. Figure 6-7 demonstrates that in this hypothetical case the ENBS reaches the maximum at a sample size of 9000 participants per arm, and therefore, this would be considered to be the optimal sample size for the study.

The purpose of this chapter was to explore whether the same or an alternative conclusion would have been drawn had the iterative Bayesian approach to health technology assessment been adopted in the design of the fFN trial, rather than the sample size being determined pragmatically through traditional methods to inform the design of the trial as it was being prepared. Under the iterative approach outlined in this chapter, the EVSI outcomes indicate that a sample size of greater than 50,000 participants in the treatment arm will be required to be worth any value, as lower sample sizes are unlikely to change the cost-effectiveness decision based on current information. A sample size of 5,000 participants in each arm will generate the greatest expected net benefit of sampling (-£1.35 million), however, as this is a negative value clearly the optimal option would be not to undertake a trial (sample size of zero). The sample conclusions that were determined previously, under the traditional approach described in Chapter 5, suggested a sample size of approximately 1000 participants in each arm. Table 6-5 details the alternative sample sizes for the fFN trial, derived from the various approaches, and shows that the EVSI approach highlights that any kind of trial is unlikely to be worthwhile in terms of altering the cost-effectiveness decision.

Table 6-5: Sample sizes under different approaches

Power	Sample per arm	Total sample size
NMB approach	925	1850
NI approach	1156	2312
EVSI approach	Zero/no trial	Zero/no trial

A study with a total sample of approximately 2000, as prescribed by the NI and NMB calculations in Chapter 5, is likely to be a waste of resources in terms of the opportunity cost for research spending. The EVSI results in Table 6-4 showed that at a sample size of 1000 per arm the EVSI is only 30 pence (£136,000 at the population level) and the probability of the study changing the decision from that under current information is 0.004, i.e. it is extremely unlikely that the study would change the decision, which under current information was to adopt the fFN test. At a cost of £1.65 million the proposed

trial costs more than the expected value of sample information, £136,000, and therefore, undertaking a trial at this scale is not providing additional value in term of reducing uncertainty about cost-effectiveness. The EVPI analysis in section 6.1.1 demonstrated that at a monetary threshold of £25,700 per morbidity avoided, the EVPI of £957,000 can be interpreted as the maximum amount the health care system should be willing to pay for additional research and therefore the fFN test should be adopted based on current evidence with no further research. This presents resource allocation issues when the opportunity cost of a study costing in the range of £1.6million is considered. The EVSI calculation highlights that a much larger study would be required to generate any value, however, considering the additional cost of a larger trial, it becomes clear that for this case study any kind of trial is unlikely to be worthwhile in terms of altering the cost-effectiveness decision.

Based on these results, under an iterative framework the likely conclusions would have been that rather than undertaking a clinical trial the fFN test should be adopted in practice, freeing up research resources for more valuable research. The fFN test could be filtered into practice on a small scale, and then the economic model updated at a later stage with routine data to further support the reimbursement of fFN by the NHS. Given this suggestion, it is of interest to consider whether any funding body would change policy or practice without evidence from an RCT. In the case of drugs, this is unlikely; however, in the context of health services and public health interventions, many NHS boards introduce new services without formal assessment of effectiveness in a trial/study setting. In the case of fFN, this diagnostic test had begun to be implemented in some hospitals in the UK as the trial application was being put together. As a result of this, the fFN trial was not funded. Implicitly, the funding body for the fFN trial came to the same conclusion as this VOI analysis (which was undertaken post-application for demonstration purposes in the context of this thesis). Ultimately the review committee queried the value of undertaking the trial because fFN had already begun to be implemented in some UK hospitals, given existing evidence from USA. Implicitly the review committee judged the proposed trial to be not worthwhile, given current practice which was beginning to implementing it anyway.

6.2 Practicalities with employing the EVSI approach

Having demonstrated the use and potential value in undertaking early modelling and incorporating an EVSI analysis to inform trial design, it is important to discuss some of the potential reasons that this approach may have been less well adopted in practice, despite strong support from some quarters of the health economics community (9;85;199).

6.2.1 Interpretation of the EVSI process

The EVSI process has been described by numerous authors (4;83;199) however, many health economists are still unfamiliar with the exact process, if not the methodology. Unfamiliarity with the process may present problems with interpretation, particularly for researchers undertaking an EVSI analyses for the first time. The outline and various algorithms presented by Ades et al. (83) are comprehensive, yet the EVSI calculation is complex, involving a choice between different algorithms and complexities depending on the situation and type of model employed. In undertaking the EVSI calculation for this chapter, some interpretation issues were experienced with regards to the simulation of binomial data, and also with regards to the appropriate approach to adopt with a linear model design, which is different if the model design is non-linear.

Simulating sample data

In the various publications outlining the EVSI process, the terminology used to describe simulating the data under perfect sample information is open to misinterpretation. Following the guidelines "simulate a dataset of a specified sample size and design" (199) or "draw a sample from the distribution of the sufficient statistics arising from a new study of size N" (83), it is unclear whether the sample drawn should be one single draw or numerous draws to generate a full sample, i.e. 1, 1000 or 10,000 draws for each prior estimate. After consultation with health economists who have previous experience of employing the EVSI technique (111), it was clarified that only one single draw is required, regardless of the type of model employed.

Therefore, in step 3 of the EVSI calculation, detailed in section 6.1.2, only one sample (x) is simulated from the beta binomial distribution, and this single draw can be assumed to reflect the distribution. The terminology used in the EVSI guidelines with regards to this does not make it clear that only one sample (x) is required to be drawn. Misinterpretation of this step could lead to simulating an additional 9,999 binomial estimates for each prior estimate unnecessarily. It makes intuitive sense to draw from the beta binomial distribution numerous times for each prior estimate to reflect the true events across the whole distribution. However, this is unnecessary in the EVSI calculation as the process (step3 to step 6) is repeated 10,000 times, drawing one sample (x) for each of the prior iterations from the Monte Carlo simulation and this is considered to be sufficient to represent the beta binomial distribution. For each iteration of the outer loop of the EVSI calculation the single binomial draw each time is different (because it is an independent binomial draw) and therefore an additional inner loop of numerous binomial draws within each iteration is unnecessary.

In order to check this premise, the EVSI calculation was re-run with a modification to step 3 in the process to draw 10,000 estimates from the beta binomial likelihood for each of the 10,000 prior estimates (rather than just once for each prior) and the outcomes were compared. Table 6-6 details the results of this comparison, and shows that both approaches have very similar outcomes.

Table 6-6: EVSI results for 1 and 10,000 binomial draws

Sample size (n)	EVSI: 1 binomial draw	Probability of change decision	EVSI: 10,000 binomial draws	Probability of change decision
20	£0.00	0.0000	£0.00	0.0000
50	£0.00	0.0000	£0.00	0.0000
80	£0.00	0.0001	£0.00	0.0000
100	£0.01	0.0002	£0.00	0.0001
150	£0.01	0.0002	£0.01	0.0002
250	£0.03	0.0006	£0.03	0.0006
500	£0.11	0.0021	£0.11	0.0018
1000	£0.31	0.0033	£0.28	0.0040
5000	£0.83	0.0095	£0.78	0.0090
10000	£0.99	0.0108	£0.92	0.0101
20000	£0.93	0.0100	£1.00	0.0107
30000	£1.03	0.0111	£1.02	0.0109
40000	£1.00	0.0104	£1.04	0.0110
50000	£1.06	0.0108	£1.04	0.0110

As the sample becomes greater than 1000 the additional draw generates a slightly lower EVSI, but only marginally. Therefore, it is appropriate and more efficient to just draw once from the binomial distribution for each of the prior estimates, substantially reducing computational running time. It is however, interesting to note that the 9,999 additional draws from the binomial generates a smoother EVSI curve than the single draw.   This is demonstrated in Figure 6-8.
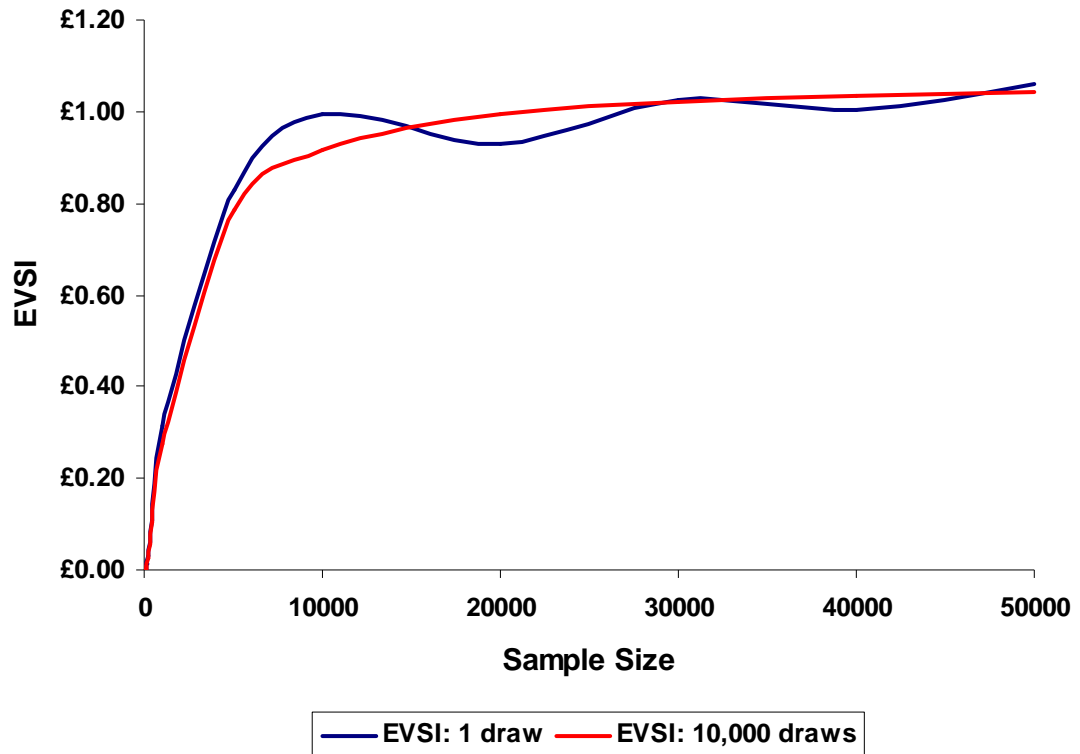
Figure 6-8: EVSI curves under one and 10,000 binomial draws

Figure 6-8 plots the EVSI results for each sample size using the two different approaches. The 9,999 additional binomial draws generated a smoother curve, which is unsurprising given the increase in iterations. If the initial number of iterations from the prior was increased from 10,000 to 100,000 the 'noise' shown in the one binomial draw curve would be ironed out, because as the number of iterations tends towards infinity, the true distribution is represented.

Misinterpretation of this step could lead to simulating 10,000 draws unnecessarily, generating an additional (unnecessary) 9,999 binomial estimates for each prior estimate, and requiring a second loop in the EVSI calculation which substantially slows the EVSI computational time, i.e. running through the calculation 100,000,000 times rather than just 10,000 times.

Alternative approaches for linear and non-linear models

Another area open to confusion and misinterpretation in the EVSI process is the alternative approaches and algorithms required depending on whether the model is linear or non-linear. Following the steps and algorithms outlined by Ades et al. (83) the researcher must decide whether the model is linear or non-linear for the parameters of interest (θ_i) and for the complementary parameters (θ_c) and then choose from four possible algorithms to calculate the posterior expected net benefits for each intervention. Linearity is determined by whether the relationship between the parameters is directly proportional or not, i.e. in a linear model a change in one of the parameters would lead to a directly proportional change in the other parameters and the model outcomes, whereas in non-linear models outcomes are a multiplicative function of the input parameters (4;72). In step 5 of the EVSI calculation, detailed in section 6.1.2, the posterior probabilities for the parameters of interest (θ_{ip}) must be put back into the model (replacing the prior estimates for the parameters of interest) in order to calculate the net benefit (NB) for each intervention (t), $NB(t,\theta_{ip})$. A linear model requires only the posterior probability for the parameter(s) of interest to calculate the net benefits. A non-linear model requires the use of the posterior distribution to re-run the Monte Carlo simulation, making the process considerably more time consuming for non-linear models.

Researchers undertaking an EVSI analysis for the first time may become confused as to which of the algorithms presented in the guidelines (83;85) is most appropriate for their model, and also in interpreting the algorithms. Table 6-7 sets out the four alternative procedures which can be used to calculate the posterior net benefits, depending on the type of model involved.

Table 6-7: Alternative methods for calculating posterior net benefits in EVSI

	Model specifications	Process to calculate the posterior net benefits
1	**Linear** parameters of interest (θ_i) & **linear** complementary parameters (θ_c)	Use posterior mean for θ_i and the prior mean for θ_c. Plug them into model and calculate NB for each intervention. No need to re-run the PSA.
2	**Linear** parameters of interest (θ_i) & **non-linear** complementary parameters (θ_c)	Use posterior mean for θ_i and the prior distribution for θ_c. Re-run the PSA drawing from the prior distribution for complimentary parameters but using the set posterior mean for parameters of interest. Calculate the ENB for each intervention.
3	**Non-linear** parameters of interest (θ_i)& **linear** complementary parameters (θ_c)	Use posterior distribution for θ_i and the prior mean for θ_c. Re-run the PSA drawing from the posterior distribution for the parameters of interest but using the set prior mean for the complimentary parameters. Calculate the ENB for each intervention.
4	**Non-linear** parameters of interest (θ_i) & **non-linear** complementary parameters (θ_c)	Use posterior distribution for θ_i and the prior distribution for θ_c. Re-run the PSA drawing from the prior distribution for complimentary parameters and the posterior distribution for the parameters of interest. Calculate the ENB for each intervention.

Developed based on information in Ades et al. 2004 (83) & personal communication with experts (111)

If either the complementary or parameters of interest in the model are non-linear, then the EVSI process involves an inner loop to re-run the PSA (nested Monte Carlo integration(83)) with 10,000 iterations, for each of the prior estimates from the outer loop of the calculation.   The process is therefore more complex from step 5 onwards than the linear process followed in section 6.1.2. The non-linear process is as follows:

1. Undertake a probabilistic sensitivity analysis (PSA) on the model: Monte Carlo simulation to generate 10,000 prior estimates (p_1 to $p_{10,000}$) for the uncertain parameters of interest.

2. Assume a sample size (N) for a new piece of research.

3. Simulate data collection for the new sample (N), from the binomial likelihood using the prior estimate for the parameter(s) of interest from the first iteration of the PSA in step 1 (p_1).

- Draw 1 sample result (x) from the binomial likelihood

4. Calculate the Beta posterior distribution $\beta(\alpha + x, \beta + N\text{-}x)$ for the parameter(s) of interest and derive the posterior estimate for the parameter of interest.

5. Put the posterior distribution for the parameters of interest back into the model (along with the complimentary parameters and their prior distributions which have not changed). Re-run the PSA using Monte Carlo simulation with 10,000 iterations, drawing from the posterior distributions for the parameters of interest and the prior distributions for the complimentary parameters. This is the inner loop.

6. Calculate the net benefit for treatment (NB_T) and control (NB_C) for each iteration of this inner loop and calculate the expected net benefit $E(NB)$ across all 10,000 for each intervention.

7. Store the ENB for each intervention and then choose the maximum of the expected net benefits $max(ENB_T:ENB_C)$ and store this too.

8. Repeat steps 3 to 7 for each prior estimate from the PSA in step 1. I.e. use p_2, p_3....$p_{10,000}$ in step 3 to simulate a single draw from the binomial for each (x_2......$x_{10,000}$) and calculate the posterior distribution for each, then re-run the 10,000 iteration PSA Monte Carlo simulation each time. Calculating the expected net benefit for treatment and control, storing the maximum each time as detailed in Table 6-8.

9. Average across the maximum expected net benefits, as illustrated in Table 6-8, this is the expected value of a decision based on sample information. The intervention which has the greatest ENB across the 10,000 PSA outcomes is the intervention of choice under current information. Subtract the ENB of the intervention of choice from the maximum ENB, this is the expected value of sample information.

Table 6-8: Calculating ENB for EVSI in a non-linear model

Net Benefit Treatment	Net Benefit Control	Maximum ENB
$ENB_{T\,p1}$	ENB_{Cp1}	$\max(ENB_T : ENB_C)\ p_1$
ENB_{Tp2}	ENB_{Cp2}	$\max(ENB_T : ENB_C)\ p_2$
ENB_{Tp3}	ENB_{Cp3}	$\max(ENB_T : ENB_C)\ p_3$
……	…….	……
$ENB_{Tp10,000}$	$ENB_{Cp10,000}$	$\max(ENB_T : ENB_C)\ p_{10,000}$
$E(ENB_T)$	**$E(ENB_C)$**	**$E(\max(ENB_T : ENB_C)$**

Interpreting the Value of Current Information

A further potential area for confusion in undertaking an EVSI calculation is misinterpretation of what exactly the 'value under current information' is.   A researcher new to the EVSI process may interpret the 'value under current information' literally, and therefore misinterpret it as the value of ENB from the intervention of choice from the prior Monte Carlo simulation. However, as with EVPPI and following the EVSI algorithm detailed in Equation 6.1, the 'value of current information' is actually generated within the EVSI simulation. Given existing evidence the optimal decision is the intervention that generates the maximum expected net-benefit, so from the EVSI simulation we choose the intervention with the maximum net benefits over all the iterations because each iteration represents a possible future realisation of the existing uncertainty in the parameter(s) of interest (4). Therefore the intervention with the greatest ENB across all 10,000 iterations in the EVSI simulation is the best we can do without sample information, aka 'the value of current information'; while the Expected Maximum NB is the best we can do with sample information as it is the average of the 'best choice' chosen for each iteration.

Current EVSI methodology papers provide the relevant EVSI algorithm (Equation 6.1) and simply refer to it as the 'value of current information', however they

could be more explicit, explaining the reasoning behind its generation from within the EVSI simulation to avoid misinterpretation and confusion amongst first time or unfamiliar EVSI users.

6.2.2 Computation

As demonstrated throughout this chapter, undertaking an EVSI calculation can vary in complexity depending on the form of model used and the relationship between the model parameters. The EVSI process requires intensive computation (83;86) and dependent on the complexity of the model can be even more burdensome. The ability to undertake an EVSI analysis depends on the processing power of the computer and the programming language used to undertake the calculation. Brennan and colleagues (85;199) discuss computational and mathematical issues surrounding EVSI calculations and explore alternative methods to try and reduce computational time, for example, using an approach called Laplace approximations (199;199).

Health economists may be familiar with a variety of computer packages, such as TreeAge (201), R (202) and STATA (203), but for many researchers the basic Microsoft Office package Excel TM will suffice for developing a probabilistic model and using macros to run Monte Carlo simulations. However, Excel TM is not sufficiently powerful to undertake EVSI calculations, even with a simple linear model such as that developed for the fFN case study. The fFN EVSI calculation used a simple linear model with four parameters of interest, therefore requiring 10,000 iterations of the model simultaneously for each of the four parameters. The calculation was initially attempted using Excel TM, but it was extremely cumbersome and the running time was approximately 66 hours to complete the full 10,000 iterations (for the linear model with four parameters of interest) for just one sample size. The calculation then had to be repeated for a range of sample sizes. Additionally, if a non-linear model had been used, 10 thousand iterations times 10 thousand would have been required, making a non-linear EVSI calculation impractical in Excel TM. Therefore, the model was reprogrammed using the programming language FORTRAN for the EVSI calculations. However, even in more sophisticated

languages, such as FORTRAN, it is still not an instant process and can be time consuming unless the code is manipulated to run efficiently.

For the EVSI calculation 10,000 iterations from the prior distribution was deemed to be a sufficient number of iterations to be representative (85;111), however, after the initial calculation was undertaken, variation in the EVSI results was explored using alternative numbers of iterations from the prior. Figure 6-9 details variation in the EVSI curve under alternative numbers of draws from the prior distribution.
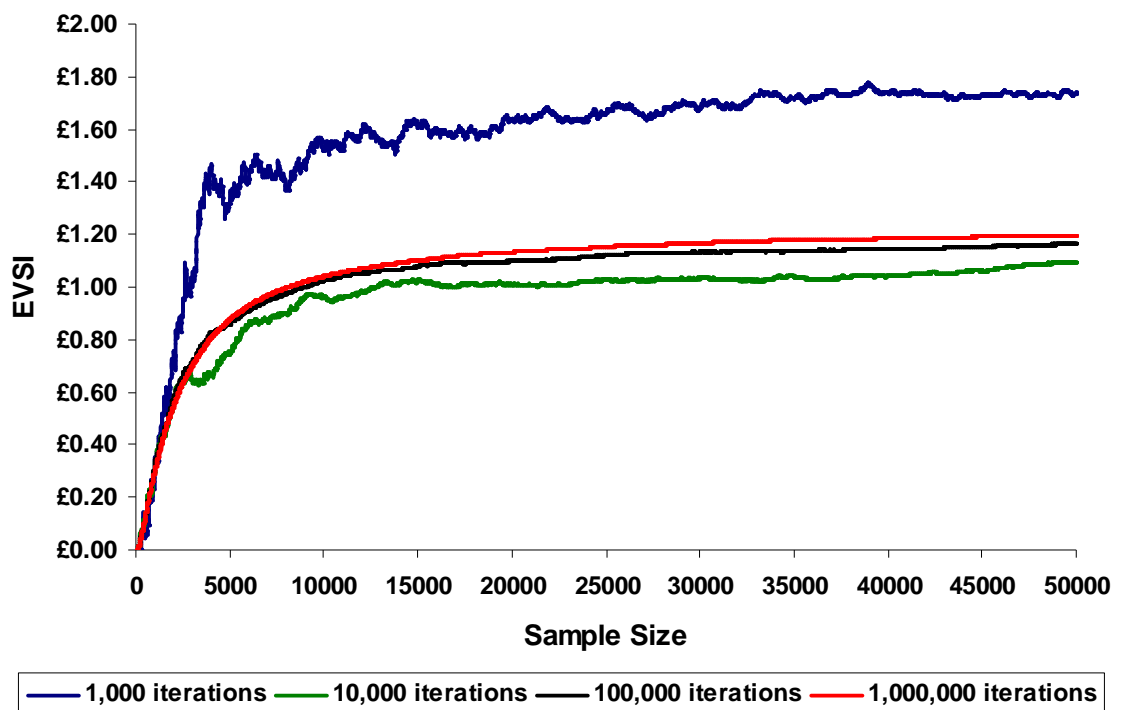
Figure 6-9: EVSI curves under alternative number of prior draws

As can be seen drawing 1000 iterations is not nearly enough to generate a smooth curve, and as recommended (85) 10,000 simulations are more likely to give representative outcomes. However, as can be seen, the 10,000 iteration draw is still subject to some 'noise' for this case study. As the number of iterations increases, the EVSI curves converge towards the 'true' solution. This figure shows that 100,000 to 1 million iterations are required in the fFN EVSI calculation to generate a realistic 'smooth' EVSI curve. Therefore, the EVSI

calculation for this case study was re-run using 1 million iterations of the calculation as opposed to 10,000. The EVSI results reported in section 6.1.3 are these more thorough results[8].

Running the EVSI calculation with an ever increasing number of iterations from the prior substantially increases computational running time, and therefore necessitates efficient code manipulation to reduce the time. Table 6-9 details the duration of the EVSI calculations in FORTRAN for the variety of prior iterations presented in Figure 6-9; for each iteration running the EVSI simulation for 5000 different sample sizes (simulating from n=10 to n=50,000 in increments of 10).

Table 6-9: Duration to run linear simulation with alternative number of priors

N iterations	Time (minutes)	(seconds)
1,000		1.551
10,000		14.363
100,000	3	0.825
1,000,000	39	48.004

Increasing the iterations from 100,000 to 1 million puts a substantial strain on the time duration, and this was only for the simple linear fFN model. Had the fFN model been non-linear, the calculation would have required an additional inner loop of 1 million Monte Carlo iterations for the posterior PSA; i.e. 1 million times 1 million iterations.

Therefore, in order to undertake an EVSI calculation a modern computer with strong processing power is required, as well as access to and knowledge of programming languages that are sufficiently powerful to undertake the complex simulations required. Extensive knowledge of how to manipulate specific programming languages is also beneficial for eliciting more efficient running times for complex, multi-loop simulations. Therefore, health

---

[8] Note that for consistency the binomial draws comparison in section 6.2.1 used 10,000 priors for both approaches.

economists or researchers considering utilising the EVSI methodology for the first time may not have the capacity, or programming know-how to undertake an EVSI calculation. This is likely to be a considerable hindrance to the wide spread adoption of EVSI methodology.

Eckermann, Karnon & Willan (81) contend that VOI needs to be useful in addressing real decisions yet simple enough to be applied by analysts and understood by decision makers in practice. They suggest that calculating the EVSI over all possible ranges of sample size using the methods proposed by Ades et al. (83) is too complex and they advocate a simpler method based on assumptions using the central limit theorem (CLT). They propose that EVSI can be considered equivalent to the current value of EVPI minus the expectation of EVPI with updated information from the trial, and that under the CLT the information from a new trial will reduce the EVPI; and as sample sizes increases, so does the amount that EVPI will be reduced by. While this approach may be more straightforward than undertaking two-level Monte Carlo simulations under the approach proposed by Claxton and Ades (83), it does not consider what is driving the uncertainty regarding cost-effectiveness. It is based upon the concept that a trial has already been designed based on a frequentist approach. As demonstrated in this Chapter, a trial commissioned by a funding body without early DAM may not be necessary in the first place, and will therefore have a specified sample size that does not add any value in terms of reducing uncertainty. By using a decision analytic modelling approach, EVPPI can help inform on the type of research required. Brennan & Kharroubi (84;199) continue to explore ways to simplify the EVSI calculation, improve efficiency and reduce computational running time.

6.3 Conclusion

This chapter has demonstrated the use of EVSI methodology and highlighted that it is a more efficient means of designing a trial through considering the value of additional information. By calculating the expected net benefit of sampling the opportunity cost of research funding can be considered. The EVPI analysis undertaken on the fFN case study at the beginning of this chapter showed that any additional research exceeding a cost of £957,000 would not be worthwhile and therefore, the EVSI approach (undertaken for illustrative purposes) called for an impractically high sample size in order to generate any value, and was therefore not worthwhile. These findings are contrary to those in Chapter 5 where an external funding body commissioned a call for an RCT in fFN, signifying that they considered such research to be worthwhile, and in response to this call a trial was designed and powered on economic endpoints. The resultant sample determined by the conventional sample calculations in Chapter 5 was far too low to have an impact on reducing uncertainty about cost-effectiveness. It also demonstrated that the sample size prescribed by conventional methods in Chapter 5 (n=1156 per arm) would generate an expected population value of £136,000 and was therefore worth less than the cost of the study, budgeted at £1.65 million. Based on these results, under an iterative framework the likely conclusions would have been: rather than undertaking a clinical trial the fFN test should be adopted in practice, freeing up research resources for more valuable research. The fFN test could be filtered into practice on a small scale, and then the economic model updated at a later stage with routine data to further support the reimbursement of fFN by the NHS. Implicitly the peer review committee for the potential funders came to the same conclusion as the VOI undertaken for this thesis, that the fFN trial added little value, given current practice in some areas which was beginning to incorporate the fFN trial based on evidence from the USA, and therefore the fFN trial was not funded.

Despite a strong case being presented in support of value of information techniques for designing trials, there remain some practical difficulties in employing the EVSI approach which may explain the reluctance amongst some

health economists, trial clinicians and statisticians to adopt this approach. There is a general perception in the medical decision making community that EVSI in particular is time-consuming, computationally expensive and has a reputation for being 'too difficult'. When these drawbacks are combined with the current research set-up, in which VOI in general is not supported (or required) by funding bodies in research applications, then it is unsurprising that many decision modellers and some health economists do not seem to think EVSI is a worthwhile exercise. Another drawback appears to be lack of experience with EVSI (for both health economists and clinicians etc.) which may compound the negative perceptions with EVSI. Despite these drawbacks, I maintain that VOI and EVSI is a worthwhile exercise, as demonstrated in this chapter. Just because it is computationally heavy and time consuming, does not mean that it should not be attempted. However, unless the current research funding set-up changes (i.e. unless a requirement for formal VOI to justify research applications is required), unfortunately many researchers are unlikely to adopt this 'good practice' in practice.

7 Summary & Conclusions

7.1 Introduction

The aim of this thesis was to examine the role of early decision analytic modelling for informing research priorities and the design of future studies in health care within the context of an iterative framework for economic evaluation. The thesis explored the feasibility, merits and drawbacks of undertaking early decision analytic modelling in practice, and considered potential reasons as to why it is not more widely implemented.

The iterative approach to economic appraisal is a framework that has been proposed as good practice (1) in which evaluations should begin with explorative modelling using indicative studies, and progress to more rigorous assessments, updating the model over time as more data becomes available. Therefore, decision analytic modelling is a key process within the iterative framework. An 'early' DAM, undertaken prior to primary research, allows explorative evaluation of cost-effectiveness based on existing evidence and can be used to assess any uncertainty surrounding the cost-effectiveness decision. Developing a DAM and undertaking probabilistic analysis at an early stage also enables the use of value of information analyses which can be used to help inform research priorities, as recommended as part of the iterative framework (1). If developed in advance of primary research, a decision model can enable full exploitation of VOI techniques and therefore help determine whether further research is potentially worthwhile, help explore the type of research required to address uncertainty in current evidence, and even help design a trial, for example with regards to an appropriate sample size.   Despite these advantages, in practice support and funding for early stage decision analytic modelling (and full exploitation of VOI techniques) is rare, and in the health care sector economic evaluations still tend to be funded as a one-off exercise alongside a trial to justify reimbursement decisions.

This thesis explored the feasibility and benefits of decision analytic modelling in practice, using case study examples. Practical applications of building early

decision analytic models were used to demonstrate the importance of early DAM; in both an 'ideal' setting where early stage modelling was adequately funded and also in a less desirable, time-constrained setting where early DAM had not been funded but proved a valuable tool for informing trial design nonetheless.   This Chapter now provides an overview of the main points from the thesis followed by a section considering some practical limitations and finally, areas for future research.

7.2 Overview

7.2.1 Chapter 1

Chapter 1 introduced the basic concepts and rationale for this thesis, as well as providing an overview of the structure of the thesis. Background was provided regarding economic evaluation in the health care sector and decision analytic modelling was introduced as a means to undertake these evaluations, regardless of whether the evaluation is being undertaken alongside a prospective trial, or from an entirely retrospective perspective. The iterative approach to economic evaluation has been proposed as a framework for good practice in economic evaluation (1), but in practice there has been little support and many stakeholders in the health care sector remain reluctant to adopt the iterative approach. The aim of this thesis was to examine the role for early decision analytic modelling through case studies of practical applications.

7.2.2 Chapter 2

This chapter introduced economic evaluation in the context of public and commercial health care research. An overview of economic evaluation in the health care sector was provided followed by a closer look at the role for decision analytic modelling. The various methodologies involved in DAM for economic evaluation were described in detail, including building a decision

analytic model, undertaking probabilistic sensitivity analysis, analysing decision uncertainty, and using value of information (VOI) techniques in order to provide relevant conclusion to decision makers and recommendations for further research.

The iterative approach to economic appraisal has been suggested as an appropriate framework for undertaking research in the health care sector, within which DAM plays a key role. Chapter 2 explored the benefits of an iterative approach for undertaking research in the health care sector, such as improvements for decision making in terms of reduced uncertainty, reduction of costs and efficiency gains. Given these potential benefits, this chapter supported the case for early stage decision modelling as part of an iterative economic process.

7.2.3 Chapter 3

Chapter 3 demonstrated a case of good practice whereby the development of an early stage decision analytic model was funded by a research body. This case study highlighted that some (national) funding bodies do recognise the importance of undertaking explorative economic analysis prior to commissioning primary research and in some cases are willing to fund and support early stage economic research along with the more conventionally funded systematic reviews.

The chapter detailed the development of three early DAMs to assess the potential cost-effectiveness of an emerging technology (PET/CT) and in doing so demonstrated the benefit and ease of synthesising current evidence and clinical expertise to develop such models. Additionally the chapter illustrated the feasibility and viability of modelling techniques in the face of limited and poor quality evidence. When little evidence is found from a systematic review it is still possible, and indeed appropriate, to develop a probabilistic model using clinical expertise and research group consensus decision making to verify the limited existing data, apply an appropriate mean estimate and wide uncertainty intervals to represent the uncertainty in probabilistic analysis. The economic models developed in Chapter 3 relied heavily on clinical expertise,

given the dearth of previous economic evaluations and poor quality, biased diagnostic test accuracy evidence in this area. Some reviewers may consider this to be a major drawback to this type of decision analytic modelling approach; however, this is precisely the point of employing early decision analytic modelling: to attempt to capture what knowledge is available and what isn't, in a systematic manner. Just because evidence is of poor quality does not mean that the economic evaluation will be poor quality.

Chapter 3 demonstrated the feasibility and benefit of synthesising current evidence and clinical expertise in an early decision analytic model, which is then used to assess potential cost-effectiveness given current information.

7.2.4 Chapter 4

Following on from the development of thee decision models for the PET/CT case study in Chapter 3, Chapter 4 explored uncertainty in the model results and demonstrated that in using cost-effectiveness acceptability curves the results can be examined in terms of decision uncertainty, followed by value of information techniques to give meaningful recommendations to funders and decision making bodies. This chapter demonstrated the advantages of early stage probabilistic models which can be used to inform on future research needs, as opposed to a more traditional approach whereby a research problem or lack of evidence is identified and used to support a case for primary research, without involving any decision analytic modelling.

Using VOI can identify cases where further research is not worthwhile, and if undertaken would be a waste of resources which could be used elsewhere, i.e. funding other, more promising research or used to fund clinical practice. As illustrated in the case of primary colorectal cancer, when current evidence is limited or of poor quality, this does not necessitate a trial. Additionally, in cases where further research is warranted, a large scale, randomised controlled trial is not necessarily required. Rather, the type of research will depend on the different parameters that require further information. In the PET/CT case study, the recurrent and metastatic analyses determined that further research would be of potential value and that it was the DTA

parameters that were driving uncertainty in the cost-effectiveness decision. Therefore the type of further research that would be required is likely to be a randomised trial to evaluate unbiased DTA characteristics for the PET/CT, ceCT and MRI technologies, however, such a trial would not necessarily be a long term trial (as the EVPPI showed that there is little or no value in further research for the 5 year survival parameters). A short term trial would be adequate to determine unbiased DTA characteristics, and could possibly also be used to derive more accurate data on patients' quality of life and the cost of PET/CT.

In addition to identifying research priorities, Chapter 4 demonstrated that applying the formal process of evidence synthesis and decision modelling at an early stage (in advance of a trial), can help ensure the appropriate research question is set. The PET/CT case study highlighted that the original decision problem identified by the funding body was not quite appropriate with regards to primary CRC, and had the funding body undertaking an iterative process, whereby explorative research and possibly a DAM was used in defining the research question the research question for Primary CRC may have been re-specified.

Chapter 4 demonstrated that even though early decision analytic modelling is not often funded and value of information analyses are even less frequently utilised in practice, on occasion they are supported by funding bodies. In such situations the practicality of undertaking these analyses is relatively straightforward and can have substantial benefits in terms of understanding outcomes and determining future research priorities. Just because current evidence is limited or of poor quality, does not necessarily mean that further research is required, and therefore, applying early DAM and VOI techniques allows decision makers to make informed decisions as to whether a new intervention should be adopted (or rejected) based on current evidence, or whether further information (and what kind of information) is required to help determine cost-effectiveness, as opposed to making decisions based on subjective reasoning.

7.2.5 Chapter 5

Chapter 5 verified that even when early decision analytic modelling has not been funded, a simple DAM can be constructed relatively quickly and fairly simply which can be used to help inform the design of a study. In the context of this thesis it would have been preferable to inform a clinical trial design following an iterative approach, whereby a comprehensive decision analytic model was developed and EVSI undertaken. However, this was not feasible within the context of the fFN study research bid. Chapter 5 offered a practical example of how an economic perspective can still be incorporated into trial design and sample size calculations in the context of a frequentist design. Given that a basic model constructed within a short time frame, such as that developed for the fFN trial, can help in the design and development of a standard frequentist trial, there is no reason for economists not to be involved in the design stage of a trial, when the purpose of that trial is to determine cost-effectiveness. This chapter highlighted that economic considerations for evaluations alongside clinical trials can and should be used to guide conventional trial design when an iterative approach to economic evaluation is not possible.

7.2.6 Chapter 6

Chapter 6 developed the fFN case study from Chapter 5, expanding the analysis to consider whether the same or an alternative conclusion would have been drawn had it been possible to adopt an iterative economic approach for the fFN trial design.

Formal exploration of uncertainty in the decision model highlighted that the fFN test was likely to be cost-effective, with a very high probability of being the optimal choice across a wide range of monetary thresholds. Given the low decision uncertainty, the EVPI analysis showed that the value of further information per decision was very low, but given the large effective population, further research up to a cost of £957,000 could potentially be of value. Considering that the cost associated with undertaking a randomised controlled trial is likely to exceed this level of population EVPI, the findings

under an iterative approach are in contrast to the research commission brief which called for research bids to undertake an RCT in fFN. The EVSI calculation was undertaken for illustrative purposes and called for an impractically high sample size in order to generate any value, which is unsurprising given the low EVPI per decision. The EVSI results illustrated that a sample of the size determined by the conventional calculations in Chapter 5, (n=1156 per arm) would generate an expected population value of £136,000 and was therefore worth less than the cost of the trial, budgeted at £1.65 million. Therefore, at a cost of £1.65 million the proposed fFN trial was likely to be a waste of resources in terms of the opportunity cost for research spending.

Based on these results, under an iterative framework the conclusions would have been that (rather than undertaking a clinical trial) the fFN test should be adopted in practice, freeing up resources for more valuable research. The fFN test could be filtered into practice on a small scale, and then the economic model updated at a later stage with routine data to further support the reimbursement of fFN by the NHS.

While the results of Chapter 6 promote the use of an iterative approach, and in particular the use of EVSI, there remain some practical difficulties in employing the EVSI approach (in terms of correct interpretation of the process, computation and time constraints) which may explain the reluctance amongst some health economists, trial clinicians and statisticians to adopt this approach. This chapter demonstrated the EVSI methodology and highlighted that it is a more efficient means of designing a trial through considering the value of additional information and through ENBS, the opportunity cost of research funding.

7.3 Practical limitations for early DAM & VOI

Time constraints in the process of designing a piece of research are an important practical hindrance to undertaking VOI calculations. The design and development of trial protocols are typically undertaken in short periods of time in order to meet tight protocol submission deadlines, and therefore, even if researchers have familiarity with the VOI methodology and processes, as well as knowledge and access to programming languages; a further constraint is the time required to develop a decision model, undertake a probabilistic analysis, followed by an EVPI analysis and finally undertake an EVSI calculation for a wide range of sample sizes.

In practice the decision problem under question (stage 1 of the iterative approach to economic evaluation) will often have been identified by a funding body who issue a call for trial proposals to address the specific question of interest, as in the fFN case study described in Chapter 5. This method means the rigorous evidence synthesis and early modelling stage of the iterative process may have been skipped, and if health economists want to contribute to the design of the trial they must do so within a short time period specified by the funding body, potentially limiting the capacity for full exploration of VOI techniques. In such circumstances adequate resources and time are not allocated to the development of a comprehensive early decision analytic model to inform the design of the trial, and any attempt to use the VOI process will likely involve a rushed, simplified analysis, open to considerable uncertainty. However, that is not to say that such an analysis should not be undertaken.

Despite a strong case being presented in this thesis in support of value of information techniques for designing trials, there remain some practical difficulties in employing the EVSI approach in particular which may explain the reluctance amongst some health economists, trial clinicians and statisticians to adopt this approach. There is a general perception in the medical decision making community that EVSI in particular is time-consuming, computationally expensive and has a reputation for being 'too difficult'. When these drawbacks are combined with the current research set-up, in which VOI in

general is not supported (or required) by funding bodies in research applications, then it is unsurprising that many decision modellers and some health economists do not consider undertaking VOI or EVSI analyses in practice. Despite these drawbacks, this thesis has demonstrated that VOI methodologies and even EVSI is a worthwhile useful exercise for explicitly informing research funding decisions. Just because EVSI techniques are computationally heavy and time consuming does not mean that they should not be attempted. However, unless the current research funding set-up changes and formally require VOI analysis, unfortunately many researchers are unlikely to adopt VOI techniques.

The fFN case study in Chapter 5 demonstrated that it is possible to build a simple model when subject to strict time constraints and that even a basic deterministic model can be informative. A simple DAM was constructed within a short time frame and was fit for purpose with regards to helping identify which of the potential clinical endpoints was the most appropriate (neonatal morbidity) and was then used to undertake the sample size calculation in the context of a frequentist trial design. Chapter 6 demonstrated that if funding bodies adopted a more flexible, iterative approach to commissioning research, it would have been possible to apply VOI techniques to the simple model and potentially feedback to the funding body regarding a more appropriate research design, within the short timeframe. A basic probabilistic analysis was undertaken on the fFN model using just the four key parameters of interest, and this in turn enabled an EVPI analysis which informed that further research is only likely to be worthwhile at a cost of less than £957,000. Even though the model was simple and undertaken in a short time period, such a conclusion could be useful in terms of highlighting to the funding body that a commissioned call for a large scale RCT is likely to be an inefficient use of funding. In such a situation funding bodies could make better use of their finances by funding the researchers to develop the basic model into a more comprehensive model, and reanalyse the PSA to get a more accurate representation of uncertainty, and explore EVPPI to determine whether an alternative smaller scale study is appropriate, rather than the initially envisaged large scale RCT.

7.4 Areas for further research

7.4.1 Fully exploiting VOI within an iterative context

Chapter 4 demonstrated that in the PET/CT case study no further research was required in primary CRC, but that in recurrent and metastatic CRC further research was potentially worthwhile. The EVPPI in both models determined that the diagnostic test accuracy parameters were the main drivers for uncertainty and therefore a short-term randomised trial would most likely be an appropriate future research design. Having submitted the research recommendations to the funding body, no further primary research has yet been commissioned. If funding were made available for further research in this area, the early model developed for this piece of work could be fully exploited by extending the VOI analysis to undertake EVSI and explore ENBS in order to determine an appropriate sample size and design criteria for such a trial. This would also be optimal in the context of the iterative framework, and in the longer term the decision model could be updated with the new DTA information from the trial.

7.4.2 Exploring uncertainty in sample size calculations

Chapter 5 demonstrated the feasibility of developing a simple DAM and undertaking sample size calculations based on cost-effectiveness endpoints within the context of a frequestist trial design. The design and development of trial protocols are typically undertaken in short periods of time in order to meet submission deadlines, resulting in potential uncertainty in the parameter estimates and sample size calculations. The role of uncertainty in sample size calculations has been given further attention in published literature and it has been proposed that probabilistic sensitivity analysis is a good way to deal with this uncertainty (198). By making a 'pre-trial' decision model probabilistic, uncertainty regarding the probability of cost-effectiveness could be incorporated into the sample size calculations. However, in practical terms the time constraints imposed by protocol submission deadlines may interfere

with such an approach and dealing with this uncertainty may only be feasible at a later time.

7.4.3 A simple guide to undertaking EVSI

Chapter 6 demonstrated the EVSI calculation; however this process also highlighted a number of issues, particularly with regards to complexity and interpretation, which is likely to hinder more widespread adoption of the EVSI technique. In moving EVSI forward, a potential means to overcome some of these issues would be the development of a simple guide to EVSI. The EVSI process has been described by numerous authors; however, existing guidelines are complicated to follow and are open to misinterpretation at numerous points, as highlighted in Chapter 6 (section 6.2). A simple guide to EVSI which outlined the process involved at each step and explicitly addressed potential pitfalls could be of considerable help to first time users of EVSI or even for researchers less familiar with the process. Such a guide could address each of the misinterpretation issues described in Chapter 6.

Additionally an area that could provide further clarity on EVSI would be with regards to determining whether a model is linear or non-linear in the parameters of interest and in the complementary parameters. The four alternative model specifications detailed in Table 6-7 highlight the four different approaches for the EVSI calculation; however, a practical worked example of the process under each different specification would provide additional clarity on the appropriate specification to adopt for those undertaking EVSI calculations. Such examples could potentially be incorporated into the EVSI user guide discussed above.

7.5 Conclusions

This thesis discussed and demonstrated the application of early decision analytic modelling in health care. The aim was to examine the role for early decision analytic modelling through case studies of practical applications.

The thesis has shown that developing 'early' decision modelling in advance of primary research is feasible and of considerable merit, regardless of whether the DAM was a fully funded comprehensive model, or an unfunded simple model constructed under relatively short/strict time constraints. Undertaking early DAM can help explore appropriate endpoints for a planned trial and, in the context of an iterative framework for economic evaluation; it can help in determining whether further research is potentially worthwhile. Applying early DAM and VOI techniques enables meaningful recommendations to decision makers, who can then make informed decisions as to whether a new intervention should be adopted (or rejected) based on current evidence, or whether further information is required to help make the decision, as opposed to making decisions based on subjective reasoning. There is considerable merit in terms of efficiency with employing early DAM prior to primary research, such as reduced uncertainty, reduction of costs and efficiency gains, however, some drawbacks exists. It may not always be viable to fully exploit VOI analyses and, with regards to undertaking EVSI calculations to inform the design of a trial, some issues remain which hinder the widespread support both inside and out-with the health economics community.

Appendix 1: Literature Search Histories

Medline (OvidSP) 1950-Nov week 2 2009

Base search for PET/CT and Colorectal cancer

1. exp Colorectal Neoplasms/

2. ((rectal or rectum or colonic or colon or colorectal or bowel* or sigmoid or anus or anal) adj3 (cancer* or carcinoma* or neoplas* or tumor* or tumour* or sarcoma* or adenocarcinoma* or adeno?carcinoma* or adenom* or lesion* or CRC)).mp.

3. or/1-2

4. exp Tomography, emission-computed/

5. positron emission tomography.ti,ab,rw,sh.

6. pet$.ti,ab,rw,sh.

7. animal/ not (human/ and animal/)

8. 6 not 7

9. exp Deoxyglucose/

10. deoxyglucose.ti,ab,rw,sh.

11. deoxy-glucose.ti,ab,rw,sh.

12. fluorodeoxyglucose.ti,ab,rw,sh.

13. ^{18}fluorodeoxyglucose.ti,ab,rw,sh.

14. fludeoxyglucose.ti,ab,rw,sh.

15. ^{18}FDG$.ti,ab,rw,sh.

16. ^{1818}FDG.ti,ab,rw,sh.

17. f-^{18}-dg.ti,ab,rw,sh.

^{18}. fluoro-2-deoxy-d-glucose.ti,ab,rw,sh.

19. 2fluoro-2deoxyglucose.ti,ab,rw,sh.

20. fluoro-d-glucose.ti,ab,rw,sh.

21. or/4-5,8-20

22. animals/ not (humans/ and animals/)

23. (3 and 21) not 22

Economics search

1. exp "Costs and Cost Analysis"/

2. Economics/

3. Cost allocation/

4. Cost control/

5. Cost savings/

6. Cost of illness/

7. Cost sharing/

8. Health care costs/

9. Direct service costs/

10. Drug costs/

11. Employer health costs/

12. Hospital costs/

13. Health expenditures/

14. Capital expenditures/

15. Value of life/

16. exp economics, hospital/

17. exp economics, medical/

18. Economics, nursing/

19. Economics, pharmaceutical/

20. exp "fees and charges"/

21. exp budgets/

22. (low adj cost).mp.

23. (high adj cost).mp.

24. (health?care adj cost$).mp.

25. (fiscal or funding or financial or finance).tw.

26. (cost adj estimate$).mp.

27. (cost adj variable).mp.

28. (unit adj cost$).mp.

29. (economic$ or pharmacoeconomic$ or price$ or pricing).tw.

30. exp models, economic/

31. ec.fs.

32. or/1-31

Toxicity, Adverse events, QoL search

1. ae.xs.
2. "Quality of Life"/
3. mo.fs.
4. quality-adjusted life years/
5. "cost of illness"/
6. (QALY or QALM or Quality-Adjusted Life Month or DALY or Disability Adjusted Life-Years).mp
7. or/1-6

Decision-making search

1. Decision Trees/
2. algorithms/
3. exp decision making, computer-assisted/ or exp decision support techniques/ or decision support systems, clinical/
4. Decision Making/
5. exp Patient Care Planning/
6. or/1-5

Embase (OvidSP) 1980 to 2009 Week 47

Base search for PET/CT and Colorectal cancer

1. exp anus tumor/ or exp colon tumor/ or exp rectum tumor/
2. ((rectal or rectum or colonic or colon or colorectal or bowel* or sigmoid or anus or anal) adj3 (cancer* or carcinoma* or neoplas* or tumor* or tumour* or sarcoma* or adenocarcinoma* or adeno?carcinoma* or adenom* or lesion* or CRC)).mp.
3. 1 or 2
4. exp computer assisted emission tomography/ or exp positron emission tomography/ or exp whole body tomography/
5. positron emission tomography.mp.
6. (pet* not (animal not (human and animal))).mp.
7. Deoxyglucose/
8. Fluorodeoxyglucose/
9. Fluorodeoxyglucose F ^{18}/
10. deoxyglucose.mp.
11. deoxy-glucose.mp.

12. fluorodeoxyglucose.mp.

13. ^{18}fluorodeoxyglucose.mp.

14. fludeoxyglucose.mp.

15. ^{18}FDG*.mp.

16. ^{1818}FDG.mp.

17. f-^{18}-dg.mp.

18. fluoro-2-deoxy-d-glucose.mp.

19. 2fluoro-2deoxyglucose.mp.

20. fluoro-d-glucose.mp.

21. exp tomography/

22. or/ 4-21

23. (rat or rats or mouse or mice or monkey* or rabbit* or hamster* or bovine or sheep).mp.

24. animal/ or experimental animal/

25. 23 or 24

26. (3 and 22) not 25

Economics search

1. Socioeconomics/

2. Cost benefit analysis/

3. Cost effectiveness analysis/

4. Cost of illness/

5. Cost control/

6. Economic aspect/

7. Financial management/

8. Health care cost/

9. Health care financing/

10. Health economics/

11. Hospital cost/

12. (fiscal or financial or finance or funding).tw.

13. Cost minimization analysis/

14. (cost adj estimate$).mp.

15. (cost adj variable$).mp.

16. (unit adj cost$).mp.

17. pe.fs.

^{18}. or/1-17

Toxicity, Adverse events, QoL search

1. exp "Quality of Life"/
2. "cost of illness"/
3. (QALY or QALM or Quality-Adjusted Life Month or DALY or Disability Adjusted Life-Years).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]
4. ae.fs.
5. to.fs.
6. or/1-5

Decision-making search

1. "decision tree"/
2. decision support system/
3. decision making/
4. algorithm/
5. clinical pathway/
6. or/1-5

Web of Science all content up to 25 Nov 2009

Base search for PET/CT and Colorectal cancer

Topic=(((rectal or rectum or colonic or colon or colorectal or bowel* or sigmoid or anus or anal) same (cancer* or carcinoma* or neoplas* or tumor* or tumour* or sarcoma* or adenocarcinoma* or adeno-carcinoma* or adenom* or lesion* or CRC)) AND (positron emission tomography or Fluorodeoxyglucose or ^{18}fluorodeoxyglucose or deoxy-glucose or Deoxyglucose or fludeoxyglucose or ^{18}FDG* or ^{1818}FDG or f-^{18}-dg or fluoro-2-deoxy-d-glucose or 2fluoro-2deoxyglucose or fluoro-d-glucose or (PET* same (CT or computer tomography))) NOT (rat or rats or mouse or mice or monkey* or rabbit* or hamster* or bovine or sheep))

Economics search

Topic=(Economic* OR cost*)

Toxicity, Adverse events, QoL search

Topic= (toxic* or adverse or "quality of life" or QALY or "quality adjusted life years" or QALM or "quality adjusted life month" or DALY or "disability adjusted life years")

Decision-making search

Topic= (decision* OR algorithm* OR pathway* OR (patient SAME management))

CINAHL Plus via Ebsco download (30 Nov 2009)

Base search for PET/CT and Colorectal cancer

((MH "Colorectal Neoplasms+") or (TX (cancer* OR carcinoma* OR neoplas* OR tumor* OR tumour* OR sarcoma* OR adenocarcinoma* OR adeno?carcinoma* OR adenom* OR lesion* OR CRC) AND (rectal OR rectum OR colonic OR colon OR colorectal OR bowel* OR sigmoid OR anus OR anal))) AND ((MH "Tomography, Emission-Computed+") or TX ("positron emission tomography" OR "[18]FDG PET-CT" OR "[18]FDG PET*") or TX fluorodeoxyglucose or ((MH "Fludeoxyglucose F [18]")))

Economics search

MW EC OR (TX cost or costs or economic* OR pharmacoeconomic* OR price* OR pricing*) OR (MH "Health Resource Utilization") OR (MH "Health Resource Allocation") OR (MH "Business+") OR (MH "Financing, Organized+") OR (MH "Financial Support+") OR (MH "Financial Management+") OR (MH "Economics+")

Toxicity, Adverse events, QoL etc search

(QALY or QALM or DALY) OR quality adjusted life years OR quality adjusted life months OR disability adjusted life years OR (MH "Ferrans and Powers Quality of Life Index") OR MW "AE" OR MW "TO" OR MW "MO" OR (MH "Quality of Life")

Decision-making search

(MH "Decision Making+") OR (MH "Algorithms") OR (MH "Triage") OR pathway* OR policy OR policies

Cochrane Library (NHSEED, HTA, CENTRAL, DARE) Issue 4, 2009

#1 MeSH descriptor Colorectal Neoplasms explode all trees

#2 (rectal OR rectum OR colonic OR colon OR colorectal OR bowel* OR sigmoid OR anus OR anal) NEAR/3 (cancer* OR carcinoma* OR neoplas* OR tumor* OR tumour* OR sarcoma* OR adenocarcinoma* OR adeno?carcinoma* OR adenom* OR lesion* OR CRC)

#3 MeSH descriptor Tomography, Emission-Computed explode all trees

#4 (positron emission tomography OR Fluorodeoxyglucose OR ^{18}fluorodeoxyglucose OR deoxy-glucose OR Deoxyglucose OR fludeoxyglucose OR ^{18}FDG* OR ^{1818}FDG OR f-^{18}-dg OR fluoro-2-deoxy-d-glucose OR 2fluoro-2deoxyglucose OR fluoro-d-glucose OR (PET* NEAR/5 (CT OR computer tomography)))

#5 MeSH descriptor Deoxyglucose explode all trees

#6 ((#1 OR #2) AND (#3 OR #4 OR #5))

HMIC Health Management Information Consortium (OvidSP) November 2009

1. positron emission tomography.mp.

2. ^{18}FDG PET-CT.mp.

3. computed tomography scanners/ or tomography/

4. colorectal cancer/

5. ((rectal or rectum or colonic or colon or colorectal or bowel* or sigmoid or anus or anal) adj3 (cancer* or carcinoma* or neoplas* or tumor* or tumour* or sarcoma* or adenocarcinoma* or adeno?carcinoma* or adenom* or lesion* or CRC)).mp.

6. 1 or 2 or 3

7. 4 or 5

8. 6 and 7

CEA Registry

"positron emission tomography"

(Anything more detailed yielded nil results)

Reference List

(1) Sculpher M, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Economics 15, 677-687. 2006.

(2) Fenwick E, Sculpher M, Claxton K, Briggs AH. Improving the efficiency and relevance of health technology assessment: The role of decision analytic modelling. Discussion paper. 1999. Birmingham, UK, Health Economics Study Group.

(3) 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 & Policy 2[1], 26-30. 1997.

(4) Briggs AH, Claxton K, Sculpher M. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press, 2006.

(5) Briggs AH. Handling uncertainty in cost-effectiveness models. Pharmacoeconomics 17[5], 479-500. 2000.

(6) Briggs AH. Handling uncertainty in economic evaluations and presenting the results. In: Drummond M, McGuire A, editors. Economic Evaluation in Health Care: merging theory with practice. Oxford: Oxford University Press, 2001: 172-214.

(7) Fenwick E, Claxton K, Sculpher M. Representing Uncertainty: the role of cost-effectiveness acceptability curves. Health Economics 10, 779-787. 2001.

(8) Claxton K. Bayesian approaches to the value of information: implications for the regulation of new pharmaceuticals. Health Economics 8[3], 269-274. 1999.

(9) Claxton K. The irrelevance of inference: a decision-making apoproach to the stochastic evaluation of health care technologies. Journal of Health Economics 18, 341-364. 1999.

(10) Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. Health Technology Assessment 8[31], 1-118. 2004.

(11) Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K, Sutton A. An iterative Bayesian approach to health technology assessment: application to a policy of preoperative optimization for patients undergoing major elective surgery. Medical Decision Making 26, 480-496. 2006.

(12) Claxton K, Eggington S, Ginnelly L, Griffin S, McCabe C, Philips Z et al. A Pilot Study of Value of Information Analysis to support research reccomendations for NICE. CHE Research Paper 4. 2005. York, Centre for Health Economics, Univerity of York.

(13) Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. Economic Evaluation 5, 513-524. 1996.

(14) Department of Health. Spending Review 2010. cited online 30/10/11. 2010.

(15) Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the economic evaluation of health care programmes. Third ed. Oxford: Oxford University Press, 2005.

(16) Hjelmgren J, Berggren F, Andersson F. Health Economic Guidelines: Similarities, Differences and Some Implications. Value In Health 4[3], 225-250. 2001.

(17) Hutton J, McGarth C, Frybourg J, Tremblay M, Bramley-Harker E, Henshall C. Framework for describing and classifying decision-making systems using technology assessment to determine the reimbursement of health technologies (fourth hurdle systems). International Journal of Technology Assessment in Health Care 22[1], 10-18. 2006.

(18) Sculpher M, Claxton K, Akehurst R. It's just evaluation for decision-making: recent developments in, and challenges for, cost-effectiveness research. In: Smith PC, Ginnelly L, Sculpher M, editors. Health Policy and Economics: Opportunities and Challenges. Maidenhead: Open University Press, 2005: 8-41.

(19) Elliott R, Payne K. Essentials of Economic Evaulation in Healthcare. London: Pharmaceutical Press, 2005.

(20) Vallejo-Torres L, Steuten L, Buxton M. Integrating health economics modelling in the product development cycle of medical devices: A Bayesian approach. International Journal of Technology Assessment in Health Care 24[4], 459-464. 2008.

(21) Fenwick E, Claxton K, Sculpher M, Briggs AH. Improving the efficiency and relevance of health technology assessment: The role of iterative decision analytic modelling. 179. 2000. York, Centre for Health Economics, University of York.

(22) Claxton K, Sculpher M, Drummond A. A rational framework for decision making by the National Institute of Clinical Excellence. The Lancet 360[9334], 711-715. 2002.

(23) Claxton K, Cohen JT, Neumann P. When is evidence sufficient? Health Affairs 24[1], 93-101. 2005.

(24) National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. 2008. National Institute for Health and Clinical Excellence.

(25) Scottish Medicines Consortium. Submissions Process: Economic Questions and Answers. cited online 30/10/11. 2011.

(26) Stinnett A, Mullahy J. Net Health Benefits : A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis. Medical Decision Making 18[2], S68-S80. 1998.

(27) Drummond M, McGuire A. Economic evaluation in health care:merging theory with practice. Oxford: Oxford University Press, 2001.

(28) Kuntz K, Weinstein M. Modelling in Economic Evaluation. In: Drummond M, McGuire A, editors. Economic Evaluation in Health Care. Oxford: Oxford University Press, 2001: 141-171.

(29) Sheldon T. Problems of using modelling in the economic evaluation of health care. Health Economics 5, 1-11. 1996.

(30) Weinstein M, 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 6[1], 9-17. 2003.

(31) Brennan A, Akehurst R. Modelling in Health Economic Evaluation: What is its place? What is its value? Pharmacoeconomics 17[5], 445-459. 2000.

(32) Brennan A, Chick S, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Economics 15, 1295-1310. 2006.

(33) Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models. Pharmacoeconomics 17[5], 461-477. 2000.

(34) Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: A review and consolidation of quality assessment. Pharmacoeconomics 24[4], 355-371. 2006.

(35) Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. British Medical Journal 342[1766], 1195-1198. 2011.

(36) Barton PM, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. Journal of Health Services Research & Policy 9[2], 110-118. 2004.

(37) Caro J, Briggs AH, Siebert U, Kuntz K. Modeling Good Research Practices - Overview: A Report of the ISPOR SMDM Modeling Good Research Practices Task Force Working Group Part 1. Draft Report. 2012.

(38) Glick H, Polsky D, Schuman K. Trial-based economic evaluations: an overview of design and analysis. In: Drummond M, McGuire A, editors. Economic evaluation in health care: merging theory with practice. Oxford: Oxford University Press, 2001: 113-140.

(39) Ramsey S, Willke R, Briggs AH, Brown R, Buxton M, Chawla A et al. Good Research Practices for Cost-Effectiveness Analysis Alongside Clinical Trials: The ISPOR RCT-CEA Task Force Report. Value In Health 2005; 8(5):521-533.

(40) Glick H, Doshi J, Sonnad S, Polsky D. Economic evaluation in clinical trials. Oxford: Oxford University Press, 2007.

(41) Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. British Medical Journal 342[1548], 1069-1073. 2011.

(42) Edwards R, Hounsome B, Linck P, Russell I. Economic evaluation alongside pragmatic randomised trials: developing a standard operating procedure for

(43) Gray A, Marshall M, Lockwood A, Morris J. Problems in conducting economic evaluations alongside clinical trials: Lessons from a study of case management for people with mental disorders. British Journal of Psychiatry 170, 47-52. 1997.

(44) Briggs AH. Economic evaluation and clinical trials: size matters. British Medical Journal 321, 1362-1363. 2000.

(45) Drummond M, BMJ Economic Evaluation Working Party. Guidelines for authors and peer reviewers of economic submissions to the BMJ. British Medical Journal 313, 275-283. 1996.

(46) Briggs AH, O'Brien B, Blackhouse G. Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. Annual Review of Public Health 23, 377-401. 2002.

(47) Fryback D, Chinnis J, Ulvila J. Bayesian cost-effectiveness analysis. International Journal of Technology Assessment in Health Care 17[1], 83-97. 2001.

(48) Eldridge S, Spencer A, Cryer C, Parsons S, Underwood M, Feder G. Why modelling a complex intervention is an important precursor to trial design: lessons from studying an intervention to reduce falls-related injuries in older people. Journal of Health Services Research & Policy 2005; 10(3):133-143.

(49) Briggs AH. New methods of analysing cost-effectiveness. British Medical Journal 2007; 335:622-623.

(50) National Institute for Health Research. About the HTA programme. cited online 30/10/11. 2011. National Institute for Health Research.

(51) Merck Sharp & Dohme Corp. Discovery & Development, Clinical Development. cited online 9/06/10. 2010.

(52) National Institute for Health Research. Identifying and prioritising HTA research. cited online 30/10/11. 2011.

(53) Research Councils UK. About the individual research councils. cited online 05/12/11. 2011. 5-12-0011.

(54) Medical Research Council. Our Research. cited online 02/12/11. 2011. 2-12-0011.

(55) Chief Scientist Office. About the Chief Scientist Office. cited online 10/01/12. 2011. 10-1-2012.

(56) Pfizer Ltd. Overview of the clinical trials process. cited online 9/06/10. 2010.

(57) Hoffmann-La Roche Ltd. Research & Development Value Chain. cited online 9/06/10. 2010.

(58) Janodia M, Sreedhar D, Ligade V, Pise A, Udupa N. Drug Development Process: A Review. cited online 8/06/11. 2007. Pharmainfo.net.

(59) Taylor R, Drummond M, Salkeld G, Sullivan D. Inclusion of cost effectiveness in licensing requirements of new drugs: the fourth hurdle. British Medical Journal 329, 972-975. 2004.

(60) Raffia H, Schlaifer R. Probability and statistics for business decisions. New York: McGraw-Hill, 1959.

(61) Snider D, Holtgrave D, Duñet D. Decision Analysis. In: Haddix A, Teutsch S, Shaffer P, Duñet D, editors. Prevention Effectiveness: A guide to decision analysis and economic evaluation. Oxford: Oxford University Press, 1996.

(62) Sonnenberg F, Beck RJ. Markov Models in medical decision making: a practical guide. Medical Decision Making 1993; 13:332-338.

(63) Briggs AH, Sculpher M. Commentary: Markov models of medical prognosis. British Medical Journal 314[354.2]. 1997.

(64) International Society for Pharmacoeconomics and Outcomes Research. Modeling Discrete Event Simulation: A Report of the ISPOR SMDM Modeling Good Research Practices Task Force Working Group Part 6. Draft Report. 2010.

(65) Glenny AM, Altman D, Song FJ, Sakarovitch C, Deeks J, D'Amico R. Indirect comparisons of competing interventions. Health Technology Assessment 9[26]. 2005.

(66) Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics 2006; 24(1):1-19.

(67) Han P, Klein W, Arora N. Varieties of uncertainty in health care: a conceptual taxonomy. Medical Decision Making 31, 828-838. 2011.

(68) Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing Structural Uncertainty in Decision Analytic Models:A Review and Application of Methods. Value In Health 12[5], 739-749. 2009.

(69) Jackson J, Bojke L, Thompson J, Claxton K, Sharples L. A Framework for Addressing Structural Uncertainty in Decision Models. Medical Decision Making 31, 662-674. 2011.

(70) Barton G, Briggs AH, Fenwick E. 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 11[5], 886-897. 2008.

(71) Fenwick E, Byford S. A guide to cost-effectiveness acceptability curves. British Journal of Psychiatry 187[106], 108. 2005.

(72) Armitage P, Berry G, Matthews J. Statistical methods in medical research. 4th ed. Oxford: Blackwell Publishing, 2002.

(73) Easton VJ, McColl J. Statistics Glossary V1.1. http://www.stats.gla.ac.uk/steps/glossary/ . 1997. 5-9-2010.

(74) Filliben J. Exploratory Data Analysis. In: Croarkin C, Tobias P, editors. Engineering Statistics Handbook. NIST/SEMATECH http://www.itl.nist.gov/div898/handbook/, 2012.

(75) Briggs AH, Ades AE, Price MJ. Probabilistic Sensitivity Analysis for Decision Trees with Multiple Branches: Use of the Dirichlet Distribution in a Bayesian Framework. Medical Decision Making 2003; 23(4):341-350.

(76) Ades AE, Claxton K, Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. Health Economics 15, 373-381. 2006.

(77) Appleby J, Devlin N, Parkin D, Buxton M, Chalkidou K. Searching for cost effectiveness thresholds in the NHS. Health Policy 2009; 91(3):239-245.

(78) O'Brien B, Gertsen K, Willan A, Faulkner A. Is there a kink in consumers' threshold value for cost-effectiveness in health care? Health Economics 11[2], 175-180. 2002.

(79) Donaldson C, Baker R, Jones-Lee M, Lancsar E, Loomes G, Mason H et al. Weighting and valuing quality adjusted life years: preliminary results from the Social Value of a QALY Project. 2008. The SVQ Research Team, Newcastle University .

(80) Claxton K, Eggington S, Ginnelly L, Griffin S, McCabe C, Philips Z et al. A Pilot Study of Value of Information Analysis to support research recommendations for the National Institute for Health and Clinical Excellence. CHE Research Paper 4. 2005. York, Centre for Health Economics, University of York.

(81) Eckermann S, Karnon J, Willan A. The Value of Value of Information: Best Informing Research Design and Prioritization Using Current Methods. Pharmacoeconomics 28[9], 699-709. 2010.

(82) Nosyk B, Sharif B, Sun H, Cooper C, Anis A. The Cost-Effectiveness and Value of Information of Three Influenza Vaccination Dosing Strategies for Individuals with Human Immunodeficiency Virus. PLoS ONE 2011; 6(12):e27059.

(83) Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modelling. Medical Decision Making 24[2], 207-227. 2004.

(84) Kharroubi SA, Brennan A, Strong M. Estimating Expected Value of Sample Information for Incomplete Data Models Using Bayesian Approximation. Medical Decision Making 31[6], 839-852. 2011.

(85) Brennan A, Chilcott J, Kharroubi SA, O'Hagan A, Cowan J. A two level Monte Carlo approach to calculating expected value of sample information: How to value a research design. ScHARR Research Paper. 2002. Available online at: https://sheffield.ac.uk/content/1/c6/02/96/29/EVSI.doc, School of Health & Related Research (ScHARR), University of Sheffield.

(86) Claxton K, Fenwick E, Sculpher M. Decision making with uncertainty: the value of information. In: Jones AM, editor. The Elgar Companion to Health Economics. Glos, UK: Edward Elgar Publishing Limited, 2012.

(87) Stevenson M, Llyod-Jones M. The cost effectiveness of a randomized controlled trial to establish the relative efficacy of Vitamin K compared with Alendronate. Medical Decision Making 31, 43-52. 2011.

(88) Spiegelhalter D, myles J, Jones D, Abrams K. An introduction to Bayesian methods in health technology assessment. British Medical Journal 319, 508-512. 1999.

(89)   Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: new guidance. 2008. Medical Research Council.

(90)   DiMasi J, Hansen R, Grabowski H. The price of innovation: new estimates of drug development costs. Journal of Health Economics 22, 151-185. 2003.

(91)   DiMasi J. The Value of Improving the Productivity of the Drug Development Process: Faster Times and Better Decisions. Pharmacoeconomics 20[3], 1-10. 2002.

(92)   Hill S, Freemantle N. A Role for Two-Stage Pharmacoeconomic Appraisal? Is There a Role for Interim Approval of a Drug for Reimbursement Based on Modelling Studies with Subsequent Full Approval Using Phase III Data? Pharmacoeconomics 21[11], 761-767. 2003.

(93)   National Institute for Health and Clinical Excellence. Scientific Advice Consultancy Service. cited online 02/07/10. 2010.

(94)   Cancer Research UK. Cancer Stats: Large Bowel Cancer UK. 2006. Cancer Research UK.

(95)   Edge S, Byrd D, Compton C, Fritz G, Greene F, Trotti A. Colon and Rectum. American Joint Committee on Cancer Staging Manual. New York: Springer, 2009.

(96)   Cancer Research UK. Treating Bowel Cancer: A quick guide. 2011.

(97)   The Royal College Radiologists. PET-CT in the UK: a strategy for development and integration of a leading egde technology within routine clinical practice. BRCR(05)0. 2005. London, The Royal College Radiologists.

(98)   Cohade C, Osman M, Leal J, Wahl RL. Direct comparison of (18)F-FDG PET and PET/CT in patients with colorectal carcinoma. Journal of Nuclear Medicine 2003; 44(11):1797-1803.

(99)   Hicks RJ, Ware RE, Lau EW. PET/CT: will it change the way that we use CT in cancer imaging? Cancer Imaging 2006; 6:S52-S62.

(100) British National Formulary. BNF 58. 2009.

(101) Department of Health. NHS Reference Costs 2008 - 2009. 2009. online, Department of Health.

(102) Personal Social Services Research Unit. Unit costs of health & social care 2009. 2009. Kent, UK, Personal Social Services Research Unit.

(103) Phelps C, Mushlin A. Focusing Technology Assessment Using Medical Decision Theory. Medical Decision Making 8[4], 279-289. 1988.

(104) Fenwick E. The application of decision analysis to the economic evaluation of diagnostic interventions. Department of Economics & Related Studies, University of York, 1998.

(105) Altman D, Bland J. Statistics Notes: diagnostic tests 1: sensitivity and specificity. British Medical Journal 1994; 308:1552.

(106) Tang M. On simultaneous assessment of sensitivity and specicity when combining two diagnostic tests. Statistics in Medicine 23[23], 3593-3605. 2004.

(107) Borowiak D, Reed J. Utility of combining two diagnostic tests. Computer Methods and Programs in Biomedicine 35, 171-175. 1991.

(108) Pandharipande P, Choy G, del Carmen M, Gazelle SG, Russell A, Lee S. MRI and PET/CT for triaging stage IB clinically operable cervical cancer to appropriate therapy: decision analysis to assess patient outcomes. American Roentgen Ray Society 192[March], 802-814. 2009.

(109) Tateishi U, Maeda T, Morimoto T, Miyake M, Arai Y, Kim EE. Non-enhanced CT versus contrast-enhanced CT in integrated PET/CT studies for nodal staging of rectal cancer. European Journal of Nuclear Medicine & Molecular Imaging 2007; 34(10):1627-1634.

(110) Tsunoda Y, Ito M, Fujii H, Kuwano H, Saito N. Preoperative diagnosis of lymph node metastases of colorectal cancer by FDG-PET/CT. Japanese Journal of Clinical Oncology 2008; 38(5):347-353.

(111) McKenna C, Claxton K, Soares M. Expected Value of Sample Information. Personal Communication: 21.09.2011. 2011. York, Centre for Health Economics, University of York. 21-9-0011.

(112) Gearhart SL, Frassica D, Rosen R, Choti M, Schulick R, Wahl R. Improved staging with pretreatment positron emission tomography/computed tomography in low rectal cancer. Annals of Surgical Oncology 2006; 13(3):397-404.

(113) Davey K, Heriot AG, Mackay J, Drummond E, Hogg A, Ngan S et al. The impact of 18-fluorodeoxyglucose positron emission tomography-computed tomography on the staging and management of primary rectal cancer. Diseases of the Colon & Rectum 2008; 51(7):997-1003.

(114) Bassi MC, Turri L, Sacchetti G, Loi G, Cannillo B, La Mattina P et al. FDG-PET/CT imaging for staging and target volume delineation in preoperative conformal radiotherapy of rectal cancer. International Journal of Radiation Oncology Biology Physics 2008; 70(5):1423-1426.

(115) Engledow AH, Bond-Smith GEL, Francis D, Pakzad F, Bomanji J, Groves A et al. The incremental value of dual modality PET/CT imaging over

PET imaging alone in advanced colorectal cancer. Indian Journal of Surgery 2009; 2009:71-68.

(116) Garin E, Devillers A, Prigent F, Bouriel C, Girault S, Boudjema K et al. Contribution of coregistrated PET/CT for patients with suspected recurrence of colo-rectal cancer. Medecine Nucleaire 2003; 27(12):665-675.

(117) Soyka JD, Veit-Haibach P, Strobel K, Breitenstein S, Tschopp A, Mende KA et al. Staging pathways in recurrent colorectal carcinoma: is contrast-enhanced 18F-FDG PET/CT the diagnostic tool of choice? Journal of Nuclear Medicine 2008; 49(3):354-361.

(118) Maroun J, Ng E, Berthelot J, Le Petit C, Dahrouge S, Flanagan W et al. Lifetime costs of colon and rectal cancer management in Canada. Chronic Disease Cancer 24[4], 91-101. 2003.

(119) The Royal College Radiologists. Radiotherapy Dose Fractionation. BFCO(06)1. 2006.

(120) Cancer Research UK.Treating Bowel Cancer. 2009.Cancer Research UK.

(121) Tenesa A, Theodoratou E, Din F, Farrington S, Cetnarskyj R, Barnetson R et al. Ten common genetic variants associated with colorectal cancer risk are not associated with survival after diagnosis. Clinical Cancer Research 16[14], 3754-3759. 2010.

(122) Ramsey S, Andersen MR, Etzioni R, Moinpour C, Peacock S, Potosky A et al. Quality of Life in survivors of colorectal cancer. American Cancer Society 2000; 88:1294-1303.

(123) National Institute for Health and Clinical Excellence. Capecitabine and oxaliplatin in the adjuvant treatment of stage III (Dukes' C) colon cancer. Technology Appraisal 100. 2006. National Institute for Health and Clinical Excellence. 2006.

(124) Kind P, Hardman G, Macran S. UK population Norms for EQ-5D. Discussion Paper 172. 1999. York, Centre for Health Economics, University of York.

(125) Office National Statistics. Deaths: age and sex, numbers and rates: Population Trends. 2009. online, Office National Statistics.

(126) Capirci C, Rubello D, Chierichetti F, Crepaldi G, Fanti S, Mandoliti G et al. Long-term prognostic value of 18F-FDG PET in patients with locally advanced rectal cancer previously treated with neoadjuvant radiochemotherapy. American Journal of Roentgenology 2006; 7(2):W202-W208.

(127) Fernandez FG, Drebin JA, Linehan DC, Dehdashti F, Siegel BA, Strasberg SM. Five-year survival after resection of hepatic metastases

from colorectal cancer in patients screened by positron emission tomography with F-18 fluorodeoxyglucose (FDG-PET). Annals of Surgery 2004; 240(3):438-447.

(128) Strasberg SM, Dehdashti F, Siegel BA, Drebin JA, Linehan D. Survival of patients evaluated by FDG-PET before hepatic resection for metastatic colorectal carcinoma: a prospective database study. Annals of Surgery 2001; 233(3):293-299.

(129) Medical Services Advisory Committee. Positron emission tomography for recurrent colorectal cancer. 35a. 2008. Australia, Medical Services Advisory Committee.

(130) Sloka J, Hollett P, Mathews M. Cost-effectiveness of Poitron Emission Tomography in recurrent colorectal cancer in Canada. McGill Journal of Medicine 7[2], 165-174. 2004.

(131) Chen LB, Tong IL, Song HZ, Zhu H, Wang YC. F-18-DG PET/CT in detection of recurrence and metastasis of colorectal cancer. World Journal of Gastroenterology 2007; 13(37):5025-5029.

(132) Tengs T, Wallace A. One Thousand Health-Related Quality of Life Estimates. Medical Care 2000; 38(6):583-637.

(133) American Cancer Society. Colorectal Cancer: Facts and Figures Special Edition. 2005. Atlanta, American Cancer Society.

(134) Park K, Schwimmer J, Shepherd J, Phelps M, Czernin J, Schiepers C et al. Decision Analysis for the cost-effective management of recurrent colorectal cancer. Annals of Surgery 233[3], 310-319. 2001.

(135) Lejeune C, Bismuth M, Conroy T, Zanni C, Bey P, Bedenne L et al. Use of a decision analysis model to assess the cost-effectiveness of F-FDG PET in the management of metachronous liver metastases of colorectal cancer. The Journal of Nuclear Medicine 46[12], 2020-2028. 2005.

(136) Zubledia J, Bednarczyk E, Baker J, Nabi H. The Economic Impact of FDG Positron Emission Tomography in the Surgical Management of Colorectal Cancer with Hepatic Metastases. Cancer Biotherapy & Radiopharmaceuticals 20[4], 450-456. 2005.

(137) Langenhoff B, Krabbe p, Peerenboom L, Wobbes T, Ruers T. Quality of life after surgical treatment of colorectal liver metastases. British Journal of Surgery 93, 1007-1014. 2006.

(138) American Joint Committee on Cancer. Colon and Rectum Cancer Staging. 7th. 2010. New York, American Joint Committee on Cancer.

(139) Barnetson R, Tenesa A, Farrington S, Nicholl I, Cetnarskyj R, Porteous M et al. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. New England Journal of Medicine 354[26], 2751-2763. 2006.

(140) Bipat S, Glas A, Slors F, Zwinderman A, Bossuyt P, Stoker J. Rectal Cancer: local staging and assessment of lymph node involvement with endoluminal US, CT and MR Imaging: a meta-analysis. Radiology 232[3], 773-783. 2004.

(141) Schaefer O, Langer M. Detection of recurrent rectal cancer with CT, MRI and PET/CT. European Radiology 2007; 17(8):2044-2054.

(142) Current Problems in Surgery. The curent management of rectal cancer. Current Problems in Surgery 42[2], 78-131. 2005.

(143) Eglinton T, Luck A, Bartholomeusz D, Varghese R, Lawrence M. Positron-emission tomography/computed tomography (PET/CT) in the intitial staging of primary rectal cancer. Colorectal Disorders 12[667], 673. 2010.

(144) Heriot AG, Hicks RJ, Drummond EG, Keck J, Mackay J, Chen F et al. Does positron emission tomography change management in primary rectal cancer? A prospective assessment. Diseases of the Colon & Rectum 2004; 47(4):451-458.

(145) Information Services Division. Scottish Health Service Costs Book 2009. 2009. online, Information Services Division.

(146) National Cancer Research Institute. A Framework for PET Research in the UK. 2007. National Cancer Research Institute. 2007.

(147) Price P, Laking G. How should we introduce clinical PET in the UK? The oncologists need to have a view. Clinical Oncology 16, 172-175. 2004.

(148) Department of Health. A Framework for the Development of Positron Emission Tomography (PET) Services in England. Gateway Number 5265. 2005.

(149) Guest J, Ruiz F, Greener M, Trotman I. Palliative care treatment patterns and associated costs of healthvare resource use for specific advanced cancer patients in the UK. European Journal of Cancer Care 15[1], 65-73. 2006.

(150) Saunders TH, Mendes Ribeiro HK, Gleeson FV. New techniques for imaging colorectal cancer: the use of MRI, PET and radioimmunoscintigraphy for primary staging and follow up. British Medical Bulletin 2002; 64:81-99.

(151) Brush J, Boyd KA, Chappell F, Crawford F, Dozier M, Fenwick E et al. The value of 18FDG positron emission tomography/computerised tomography (PET/CT) in pre-operative staging of colorectal cancer: a systematic review and economic evaluation. Health Technology Assessment 15[35]. 2011.

(152) Sarikaya I, Bloomston M, Povoski SP, Zhang J, Hall NC, Knopp MV et al. FDG-PET scan in patients with clinically and/or radiologically suspicious colorectal cancer recurrence but normal CEA. World Journal of Surgical Oncology 2007;(5):article no 64.

(153) Bellomi M, Rizzo S, Travaini LL, Bazzi L, Trifiro G, Zampino MG et al. Role of multidetector CT and FDG-PET/CT in the diagnosis of local and distant recurrence of resected rectal cancer. Radiologia Medica 2007; 112(5):681-690.

(154) Votrubova J, Belohlavek O, Jaruskova M, Oliverius M, Lohynska R, Trskova K et al. The role of FDG-PET/CT in the detection of recurrent colorectal cancer. European Journal of Nuclear Medicine & Molecular Imaging 2006; 33(7):779-784.

(155) Even-Sapir E, Parag Y, Lerman H, Gutman M, Levine C, Rabau M et al. Detection of recurrence in patients with rectal cancer: PET/CT after abdominoperineal or anterior resection. Radiology 2004; 232(3):815-822.

(156) Schmidt GP, Baur-Melnyk A, Haug A, Utzschneider S, Becker CR, Tiling R et al. Whole-body MRI at 1.5 T and 3 T compared with FDG-PET-CT for the detection of tumour recurrence in patients with colorectal cancer. European Radiology 2009; 19(6):1366-1378.

(157) Ramos E, Martinez L, Gamez C, Torras J, Valls C, Rafecas A et al. Use of PET-CT in pre-surgical staging of colorectal cancer hepatic metastases. Cirugia Espanola 2008; 84(2):71-77.

(158) Selzner M, Hany TF, Wildbrett P, McCormack L, Kadry Z, Clavien PA. Does the novel PET/CT imaging modality impact on the treatment of patients with metastatic colorectal cancer of the liver? Annals of Surgery 2004; 240(6):1027-1034.

(159) Sloka J, Hollett P. Cost effectiveness of Poitron Emission Tomography in Canada. Medical Science Monitor 11[10], PH1-PH6. 2005.

(160) D'Souza MM, Sharma R, Mondal A, Jaimini A, Tripathi M, Saw SK et al. Prospective evaluation of CECT and 18F-FDG-PET/CT in detection of hepatic metastases. Nuclear Medicine Communications 2009; 30(2):117-125.

(161) Chua SC, Groves AM, Kayani I, Menezes L, Gacinovic S, Du Y et al. The impact of 18F-FDG PET/CT in patients with liver metastases. European Journal of Nuclear Medicine & Molecular Imaging 2007; 34(12):1906-1914.

(162) Kong G, Jackson C, Koh DM, Lewington V, Sharma B, Brown G et al. The use of 18F-FDG PET/CT in colorectal liver metastases-comparison with CT and liver MRI. European Journal of Nuclear Medicine & Molecular Imaging 2008; 35(7):1323-1329.

(163)   Rappeport ED, Loft A, Berthelsen AK, der Recke P, Larsen PN, Mogensen AM et al. Contrast-enhanced FDG-PET/CT vs. SPIO-enhanced MRI vs. FDG-PET vs. CT in patients with liver metastases from colorectal cancer: A prospective study with intraoperative confirmation. Acta Radiologica 2007; 48(4):369-378.

(164)   Coenegrachts K, De GF, ter BL, Walgraeve N, Bipat S, Stoker J et al. Comparison of MRI (including SS SE-EPI and SPIO-enhanced MRI) and FDG-PET/CT for the detection of colorectal liver metastases. European Radiology 19(2)()(pp 370-379), 2009 Date of Publication: 2009 2009;(2):370-379.

(165)   Wildi SM, Gubler C, Hany T, Petrowsky H, Clavien PA, Jochum W et al. Intraoperative sonography in patients with colorectal cancer and resectable liver metastases on preoperative FDG-PET-CT. Journal of Clinical Ultrasound 2008; 36(1):20-26.

(166)   Scottish Intercollegiate Guidelines Network. SIGN 50: A guideline developer's handbook. 2011. Edinburgh, Scottish Intercollegiate Guidelines Network.

(167)   Harbord RM, Deeks J, Egger M, Whiting P, Sterne J. A unification of models for meta analysis of diagnostic accuracy studies. Biostatistics 8[2], 239-251. 2007.

(168)   Briggs AH. Statistical approaches to handling uncertainty in health economic evaluation. European Journal of Gastroenterology & Hepatology 2004; 16:551-561.

(169)   Drummond M. Economic analysis alongside clinical trials. London: Department of Health, 1994.

(170)   Briggs AH, Gray A. Power and sample size calculations for stochastic cost-effectiveness analysis. Medical Decision Making 18, S81-S92. 1998.

(171)   Briggs AH, Tambour M. The design and analysis of stochastic cost-effectiveness studies for the evaluation of health care interventions. Drug Information Journal 35[4], 1455-1468. 1998.

(172)   Committee for medicinal products for human use. Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99. 2005. London, UK, European Medicines Agency.

(173)   Information Services Division. Births and Babies. 2005. NHS National Services Scotland.

(174)   Kenyon S, Taylor D, Tarnow-Mordi W. Broad-spectrum antibiotics for spontaneous preterm labour: the ORACLE II randomised trial. The Lancet 357[9261], 989-994. 2001.

(175) Roberts D, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Database of Systematice Reviews 3[CD004454]. 2006.

(176) Phillipson B, Clark P. Baby steps to better care: Bliss Baby Report 2008. 2008. London, Bliss.

(177) Information Services Division. Scottish Perinatal Infant Morbidity and Mortality Report 2008. 2009. NHS National Services Scotland.

(178) Hospital Episode Statistics Online. Maternity Data Reports . 2010. online, The Information Centre for Health and Social Care, NHS.

(179) The Consortium on Safe Labor. Respiratory Morbidity in late Preterm Births. JAMA 304[4], 411-418. 2010.

(180) Altman M, Vanpee M, Cnattingius S, Norman M. Neonatal Morbidity in Moderately Preterm Infants: A Swedish National Population-Based Study. The Journal of Pediatrics 158[2], 239-244. 2011.

(181) Healthcare Commission. Towards better births: A review of maternity services in England. Commission for Healthcare Audit and Inspection, Concordat gateway number: 134. 2008.

(182) Hogg M, Penney G, Carmichael J. Audit of Care Provided and Outcomes Achieved by Community Maternity Units in Scotland 2005. Scottish Programme for Clinical Effectiveness in Reproductive Health, Publication Number 29. 2007. NHS Quality Improvement Scotland.

(183) Crowther C, Doyle L, Haslam R, Hiller J, Harding J, Robinson J. Outcomes at 2 years of age after repeat doses of antenatal corticosteroids. New England Journal of Medicine 357[12], 1179-1189. 2007.

(184) de Heus R, Mol B, Erwich J, van Geijn H, Gyselaers W, Hanssens M et al. Adverse drug reactions to tocolytic treatment for preterm labour: prospective cohort study. British Medical Journal 338[b744]. 2009.

(185) Tekesin I, Marek S, Hellmeyer L, Reitz D, Schmidt S. Assessment of Rapid Fetal Fibronectin in Predicting Preterm Delivery. Obstetrics & Gynecology 105[2], 280-284. 2005.

(186) Macintyre-Beon S, Skeoch C, Jackson L, Boot P, Cameron A. Perinatal Collaborative Transport Study (CoTS) Final report. DS30/LC. 2008. Glasgow, NHS Quality Improvement Scotland.

(187) Information Services Division. Cost Book: Maternity, Special Baby Care Unit. 2008. online, Information Services Division.

(188) Committee for proprietary medicinal products. Points to consider on switching between superiority and non inferiority. CPMP/EWP/482/99.

2000. London, European Agency for the Evaluation of Medicinal Products.

(189) Johansson P. On the definition and age-dependency of the value of a statistical life. The Journal of Risk and Uncertainty 25[3], 251-263. 2002.

(190) Department for Transport. Highways Economics Note No. 1: 2005 Valuation of the Benefits of Prevention of Road Accidents and Casualties. 1. 2007. Department for Transport.

(191) Abelson P. The value of life and health for public policy. The Economic

(192) Godfrey C, Pickett K, Parrott S, Mdege N, Eapen P. Estimating the Costs to the NHS of Smoking in Pregnancy for Pregnant Women and Infants. Final report. 2010. York, Public Health Research Consortium, University of York.

(193) Petrou S, Mehta Z, Hockley C, Cook-Mozaffari P, Henderson J, Goldacre M. The impact of pre-term birth on hospital inpatient admissions and costs during the first 5 years of life. Pediatrics 112[6], 1290-1297. 2003.

(194) Dolan P, Metcalfe R, Munro V, Christensen M. Valuing lives and life years: anomalies, implications, and an alternative. Health Economics, Policy & Law 2008; 3:277-300.

(195) Mason H, Jones-Lee M, Donaldson C. Modilling the monetary value of a QALY: A new approach based on UK data. Health Economics 18[8], 933-950. 2008.

(196) Office National Statistics. Life expectancy at birth: UK 2005-2007. 2008. online, Office National Statistics.

(197) Kay R. Issues in non-inferiority trials. Online lecture slides, cited 05/08/2010. 2009. Sheffield, Sheffield University.

(198) O'Hagan A, Stevens J. Bayesian assessment of sample size for clinical trials of cost-effectiveness. Medical Decision Making 21, 219-230. 2001.

(199) Brennan A, Kharroubi SA. Efficient computation of partial expected value of sample information using Bayesian approximation. Journal of Health Economics 26, 122-148. 2007.

(200) Brennan A, Chilcott J, Kharroubi SA, O'Hagan A. Cost-effectiveness models to inform trial design: Calculating the expected value of sample information . Online lecture slides, cited 3/01/2011. 2011. Sheffield, School of Health and Related Research.

(201) TreeAge Software Inc. TreeAge . 2012. USA, TreeAge Software Inc.

(202) The R Project for Statistical Computing. The R Project for Statistical Computing. 2012. Vienna, Austria, Institute for Statistics and Mathematics.

(203) StataCorp. STATA. 2012. USA, StataCorp TX.