

The role of modelling in prioritising and planning clinical trials

J Chilcott

A Brennan

A Booth

J Karnon

P Tappenden



Health Technology Assessment
NHS R&D HTA Programme





INAHTA

How to obtain copies of this and other HTA Programme reports.

An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (<http://www.hta.ac.uk>). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

- fax (with **credit card** or **official purchase order**)
- post (with **credit card** or **official purchase order** or **cheque**)
- phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque

If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card

The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order

You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of HTA on CD?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

The role of modelling in prioritising and planning clinical trials

J Chilcott*
A Brennan
A Booth
J Karnon
P Tappenden

School of Health and Related Research (ScHARR), University of Sheffield, UK

* Corresponding author

Declared competing interests of authors: none

Published August 2003

This report should be referenced as follows:

Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P. The role of modelling in prioritising and planning clinical trials. *Health Technol Assess* 2003;**7**(23).

Health Technology Assessment is indexed in *Index Medicus/MEDLINE* and *Excerpta Medica/EMBASE*.

NHS R&D HTA Programme

The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS.

Initially, six HTA panels (pharmaceuticals, acute sector, primary and community care, diagnostics and imaging, population screening, methodology) helped to set the research priorities for the HTA Programme. However, during the past few years there have been a number of changes in and around NHS R&D, such as the establishment of the National Institute for Clinical Excellence (NICE) and the creation of three new research programmes: Service Delivery and Organisation (SDO); New and Emerging Applications of Technology (NEAT); and the Methodology Programme.

The research reported in this monograph was identified as a priority by the HTA Programme's Methodology Panel and was funded as project number 96/50/02.

The views expressed in this publication are those of the authors and not necessarily those of the Methodology Programme, HTA Programme or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

Criteria for inclusion in the HTA monograph series

Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

Methodology Programme Director: Professor Richard Lilford

HTA Programme Director: Professor Kent Woods

Series Editors: Professor Andrew Stevens, Dr Ken Stein, Professor John Gabbay,
Dr Ruairidh Milne and Dr Rob Riemsma

Managing Editors: Sally Bailey and Sarah Llewellyn Lloyd

The editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

© Queen's Printer and Controller of HMSO 2003

This monograph may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising.

Applications for commercial reproduction should be addressed to HMSO, The Copyright Unit, St Clements House, 2-16 Colegate, Norwich, NR3 1BQ.

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.

Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.



Abstract

The role of modelling in prioritising and planning clinical trials

J Chilcott,* A Brennan, A Booth, J Karnon and P Tappenden

School of Health and Related Research (SchARR), University of Sheffield, UK

* Corresponding author

Objectives: To identify the role of modelling in planning and prioritising trials. The review focuses on modelling methods used in the construction of disease models and on methods for their analysis and interpretation.

Data sources: Searches were initially developed in MEDLINE and then translated into other databases.

Review methods: Systematic reviews of the methodological and case study literature were undertaken. Search strategies focused on the intersection between three domains: modelling, health technology assessment and prioritisation.

Results: The review found that modelling can extend the validity of trials by: generalising from trial populations to specific target groups; generalising to other settings and countries; extrapolating trial outcomes to the longer term; linking intermediate outcome measures to final outcomes; extending analysis to the relevant comparators; adjusting for prognostic factors in trials; and synthesising research results. The review suggested that modelling may offer greatest benefits where the impact of a technology occurs over a long duration, where disease/technology characteristics are not observable, where there are long lead times in research, or for rapidly changing technologies. It was also found that modelling can inform the key parameters for research: sample size, trial duration and population characteristics. One-way, multi-way and threshold sensitivity analysis have been

used in informing these aspects but are flawed. The payback approach has been piloted and while there have been weaknesses in its implementation, the approach does have potential. Expected value of information analysis is the only existing methodology that has been applied in practice and can address all these issues. The potential benefit of this methodology is that the value of research is directly related to its impact on technology commissioning decisions, and is demonstrated in real and absolute rather than relative terms; it assesses the technical efficiency of different types of research. Modelling is not a substitute for data collection. However, modelling can identify trial designs of low priority in informing health technology commissioning decisions.

Conclusions: Good practice in undertaking and reporting economic modelling studies requires further dissemination and support, specifically in sensitivity analyses, model validation and the reporting of assumptions. Case studies of the payback approach using stochastic sensitivity analyses should be developed. Use of overall expected value of perfect information should be encouraged in modelling studies seeking to inform prioritisation and planning of health technology assessments. Research is required to assess if the potential benefits of value of information analysis can be realised in practice; on the definition of an adequate objective function; on methods for analysing computationally expensive models; and on methods for updating prior probability distributions.



Contents

| | | | |
|---|-----|---|----|
| List of abbreviations | vii | Current approaches | 33 |
| Executive summary | ix | Critique of general modelling approach | 34 |
| 1 Introduction | 1 | Recommended prioritisation approaches | 35 |
| The research questions | 1 | Conclusions | 36 |
| Defining terms | 1 | 6 The direct assessment of the cost- effectiveness or 'payback' of research | 37 |
| Background: why the research questions are important | 2 | Introduction | 37 |
| Scope of the review | 3 | Review of case studies | 37 |
| Audience for the report | 4 | Methodological debates | 45 |
| Structure of the report | 4 | Evaluation against criteria | 48 |
| 2 Literature search and review methods | 7 | Conclusions and recommendations | 51 |
| Introduction | 7 | 7 The expected value of information approach | 53 |
| Study identification | 10 | Introduction | 53 |
| Study selection | 14 | The concepts involved | 53 |
| Data extraction and appraisal of studies | 14 | Steps in the EVI method | 55 |
| Presentation of results | 16 | Issues around the application of the EVI approach | 60 |
| Conclusions and recommendations from the systematic review process | 17 | Evaluation against criteria | 62 |
| Conclusions regarding systematic review methodology | 18 | Conclusions and recommendations | 65 |
| Results | 18 | 8 Expected value of information case studies | 69 |
| 3 The use of mathematical modelling in health technology assessment | 19 | Introduction | 69 |
| Introduction to the three important modelling methods | 19 | Decision-analytic determination of study size: the case of electronic foetal monitoring | 69 |
| The roles of modelling in health technology assessment | 22 | Focusing technology assessment using medical decision theory | 71 |
| Use of standard cost-effectiveness modelling and sensitivity analysis to inform the design of clinical research | 25 | Meta-analysis by the confidence profile method | 72 |
| Conclusions and recommendations | 26 | An economic approach to clinical trial design and research priority setting | 73 |
| 4 Good practice and critical appraisal of modelling studies in health technology assessment | 27 | Designing a cost-effective clinical trial | 74 |
| Introduction | 27 | The cost-benefit of a randomised trial to a healthcare organisation | 75 |
| Review of good modelling practice | 28 | The irrelevance of inference: a decision- making approach to the stochastic evaluation of healthcare technologies | 77 |
| Consensus on principles | 31 | Case study: EVI of research on diagnosis of urinary tract infections | 77 |
| Conclusions | 31 | Sensitivity analysis and the EVPI | 79 |
| 5 Background to the research prioritisation process | 33 | Bayesian approach to sensitivity analysis | 79 |
| Introduction | 33 | | |
| Criteria for prioritisation | 33 | | |

| | | | |
|--|----|--|-----|
| Bayesian value-of-information analysis: an application to a policy model of Alzheimer's disease | 80 | The EVI approach: informing both the design and prioritising of research studies | 99 |
| EVPI in liquid cytology for cervical screening | 81 | Recommendations for further R&D in modelling methodology in planning and prioritising trials | 102 |
| The clinical-effectiveness and cost-effectiveness of machine perfusion versus cold storage of kidneys for heart-beating and non-heart-beating donors | 82 | Acknowledgements | 103 |
| EVPI in screening for inborn errors of metabolism using tandem mass spectrometry | 83 | References | 105 |
| 9 Other related literature | 85 | Appendix 1 Review of case studies addressing the planning and design of future research | 111 |
| Overview | 85 | Appendix 2 Summary of identified modelling guidelines | 119 |
| Commercial investment appraisal approaches | 85 | Appendix 3 Data extraction sheet for identifying the key roles of modelling | 121 |
| Mathematical modelling of underlying biological processes | 89 | Appendix 4 Data extraction sheet for discussion on the roles of modelling | 123 |
| Pretrial simulation of clinical trials | 90 | Appendix 5 Data extraction sheet for case studies on the role of modelling | 125 |
| 10 Conclusions and recommendations | 93 | Health Technology and Assessment reports published to date | 127 |
| Objectives | 93 | Health Technology and Assessment Programme | 135 |
| What do we mean by 'modelling'? | 93 | | |
| Obtaining the relevant literature | 93 | | |
| The role of modelling in health technology assessment | 94 | | |
| Good practice guidelines and critical appraisal of modelling in health technology assessment | 95 | | |
| The current place of modelling in research prioritisation | 96 | | |
| Modelling the likely consequences of proposed research: the cost-benefit or 'payback' approach | 97 | | |



List of abbreviations

| | | | |
|--------|--|--------|---|
| CEAC | cost-effectiveness acceptability curve | HEED | Health Economic Evaluations Database |
| CNS | central nervous system | HRT | hormone replacement therapy |
| CPM | confidence profile method | HSR | health services research |
| CRD | Centre for Reviews and Dissemination | ICER | incremental cost-effectiveness ratio |
| CS | cold storage | INB | incremental net benefit |
| DCCT | Diabetes Control and Complications Trial | ITT | intention to treat |
| DEALE | declining exponential approximation of life expectancy | MAICER | maximum acceptable incremental cost-effectiveness ratio |
| DEC | Development and Evaluation Committee | MCAD | medium-chain acyl-coenzyme A dehydrogenase |
| DES | discrete event simulation | MeSH | Medical Subject Headings |
| DGF | delayed graft function | MP | machine (pulsatile) perfusion |
| EFM | electronic foetal monitoring | MS/MS | tandem mass spectrometry |
| ENBS | expected net benefits of sampling | MSUD | maple syrup urine disease |
| EVCI | expected value of clinical information | NB | net benefit |
| EVI | expected value of information | NEED | NHS Economic Evaluations Database |
| EVImpI | expected value of imperfect information | NICE | National Institute for Clinical Excellence |
| EVPI | expected value of perfect information | NPV | net present value |
| EVSI | expected value of sample information | PEVPI | population expected value of perfect information |
| | | QALY | quality-adjusted life year |

continued

List of abbreviations *continued*

| | | | |
|-----|------------------------------------|-------|--|
| R&D | research and development | RR | relative risk |
| RCT | randomised controlled trial | TAPSS | Technology Assessment Priority Scoring Scale |
| ROC | receiver operating characteristics | UTI | urinary tract infection |

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



Executive summary

Background

Most decision-analytic models in health technology assessment describe pathways through health states and events in a population. Mathematical models of the natural history of a disease are used to estimate health outcomes, resource usage and costs, and to compare the clinical and economic effectiveness of the technologies under assessment. The most common mathematical techniques used are decision analysis, state transition models and discrete event simulation. The appropriate technique depends on the characteristics of the treatment under evaluation.

Objectives

- To assess modelling methods used in the construction of disease models to support health technology assessment, and methods for their analysis and interpretation.
- To identify the role of mathematical modelling in planning and prioritising trials. 'Trials' is defined as all forms of primary research supporting health technology assessment of the clinical and economic consequence of alternative interventions.

Methods

Systematic reviews of the methodological and case study literature were undertaken. Search strategies focused on the intersection between modelling, health technology assessment, and priorities and prioritisation.

Results and conclusions

Five central questions were addressed.

(1) In what ways can modelling extend the validity of trials?

By:

- generalising from trial populations to specific target groups
- generalising to other settings and countries
- extrapolating trial outcomes to the longer term
- linking intermediate outcome measures to final outcomes

- extending analysis to relevant rather than trial comparators
- adjusting for prognostic factors in trials
- synthesising primary research results.

These conclusions are drawn from the review of methodological and case studies of economic models from the general health technology assessment literature that claims some value in research planning and design. In undertaking modelling or interpreting the results of modelling studies, the degree of reliance that can be placed on these studies is important, so close attention must be paid to guidelines for good practice.

(2) What characteristics of the trial/technology affect the success of modelling?

The review does not highlight specific success factors within the trials or technologies; given analytical expertise, there are no theoretical distinctions between alternative disease areas. Modelling may offer greater benefits as an evaluative tool for certain forms of health technology, such as diagnostics and screening, which may have an impact over a long period and where key disease/technology characteristics may not be directly observable. It may also provide more substantial benefits for technologies with long lead times in research, or for rapidly changing technologies.

A limited evidence base will reduce the 'success' of modelling, if the criterion is usefulness of a model in deciding on the adoption of the technology in practice. However, if the criterion for a model's success is its usefulness in helping to decide on further research, then a limited evidence base is inevitable, and provides the key source material to describe the current uncertainty.

(3) What aspects of trial design can modelling feasibly inform?

Cost-effectiveness modelling and sensitivity analysis can inform research design by: identifying key parameters requiring further investigation, specifying the minimum clinical difference needed for sample size calculations for a proposed trial, and defining the duration and population characteristics of a proposed trial.

Some methodological discussion and case studies use standard methods of sensitivity analysis in informing these aspects, but these methods have weaknesses. Analytical methods focusing on trial design and prioritisation are required. Two methods identified in the literature are payback methods and expected value of information (EVI) analysis.

- Payback methodology presupposes a specific trial design and therefore does not explicitly address this issue. Specific applications have focused on its role in informing the sample size of trials.
- EVI analysis of economics models has been applied in practice and can address all these issues.

(4) How feasible, costly and beneficial might modelling be as part of the prioritisation process?

Although the payback approach has not always been implemented successfully, it has potential feasibility. There are no published results on its implementation costs. The benefits are unproven but are often conceived as increased explicitness of the prioritisation process and improved decision-making. The main requirement for research into payback methods is the implementation of stochastic sensitivity analysis within exemplar case studies.

EVI analyses have been shown to be possible within the financial, resource and time constraints of the NHS HTA R&D Programme. The potential benefits of EVI are:

- The value of further research relates directly to its impact on technology commissioning decisions and the consequential health and economic benefits, and is demonstrated in real and absolute rather than relative terms.
- It avoids the misleading rankings of uncertainties that may result from conventional sensitivity analyses.
- It does not start from a prespecified research design, but identifies key uncertainties and allows the technical efficiency of many different types of research to be assessed. Further research is required to establish the benefits in practice.

(5) How far can modelling substitute for low-priority trials?

Modelling is not a substitute for data collection. By identifying the absolute and relative value of further research on specific parameters, EVI analysis directly identifies trial designs of low priority in informing technology commissioning decisions.

Recommendations for further research

- To report issues of good practice in undertaking and reporting economic modelling. Areas for development include model validation, stochastic sensitivity analyses, and specifically the cost-effectiveness acceptability curve presentation of uncertainty and the explicit reporting of assumptions. The guidelines identified here should be recommended to journals that publish economic evaluations to provide a structure for peer review.
- To develop case studies using stochastic sensitivity analyses within the payback approach to prioritisation of research.
- To encourage the calculation of the overall expected value of perfect information for a decision problem in modelling studies seeking to inform the prioritisation and planning of health technology assessment.
- To identify the potential benefits of EVI analysis and assess whether they can be realised in R&D prioritisation and planning in practice.
- To define an objective function that captures the issues of importance to decision-makers in health technology assessment planning and prioritisation, and includes quantifiable aspects to incorporate into a process that supports the arbitration of subjective judgement.
- To develop approximation methods to allow the general application of EVI methods.
- To develop a general method to estimate expected value and expected net benefit of sample information, through methodological research into updating of prior probability distributions. These methods should be demonstrated in case studies.

Chapter I

Introduction

The research questions

This report is the result of a project commissioned by the National Co-ordinating Centre for Health Technology Assessment (NCCHTA) with the original title: 'A systematic review of the use of modelling in planning and prioritising clinical trials'. The commissioning brief identified five questions to be addressed:

- How feasible, costly and beneficial might modelling be as part of the prioritisation and design process?
- What aspects of trial design can modelling feasibly inform?
- What characteristics of the trial/technology affect the success of modelling?
- How far can modelling substitute for low-priority trials?
- In what ways can modelling extend the validity of trials (e.g. through adding to their generalisability)?

Other research questions have evolved during the course of the review, including:

- Where might modelling be most beneficial in the cycle of tasks in the NHS research and development (R&D) process (as described in *Figure 1*)?
- Which methodological approach should be recommended for attempting to assess the potential value of a research project?

In addition, several subquestions evolved concerning the appropriate methodology for the different steps in the methods. These questions were of the form 'what is the best way to do step X of the method?' The report makes recommendations on these detailed aspects of method, which might be considered good practice.

Defining terms

In order to complete the research successfully, it was important to define clearly each of the terms used in the statement of the research question. Being clear on the definitions of these terms from the beginning will help the reader to understand the scope of the report, its findings and its recommendations.

Systematic review

A systematic review is the systematic search for, and review of, published and unpublished work concerning aspects of the research question.

Clinical trials and health technology assessments

A clinical trial usually involves the comparative assessment of two or more interventions on the health outcomes of a group of people. Study participants can be randomised to the different treatments to minimise selection bias. Both participants and providers of the treatment can be blinded so that knowledge of the intervention received cannot influence patient response. This

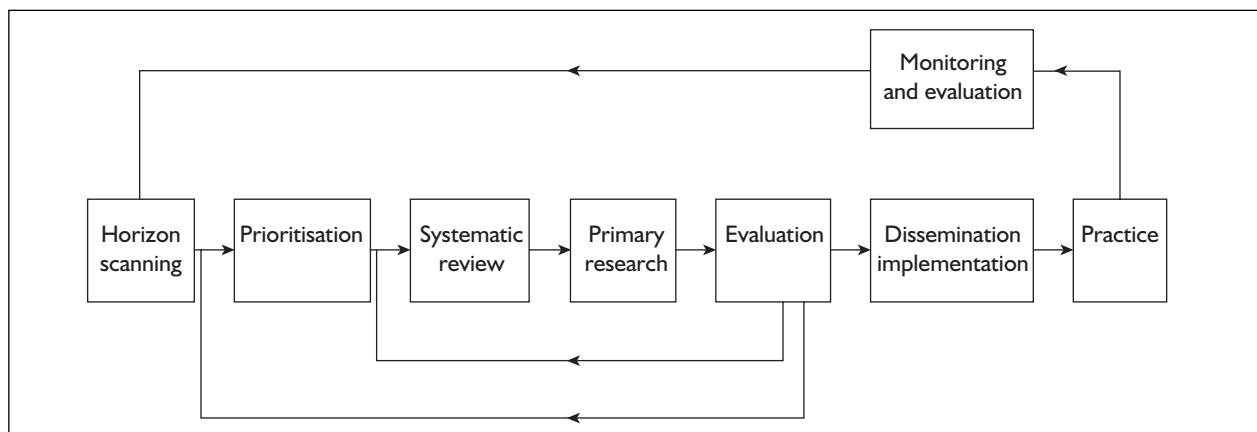


FIGURE 1 The cycle of tasks in the NHS R&D process

double-blind, randomised control trial (RCT) is the gold standard for the scientific assessment of the efficacy of a treatment. Alternative forms of retrospective and prospective observational studies can also be used to inform the effectiveness of competing interventions. The choice of study will depend on the exact nature of the interventions and research questions concerned.

This review has widened the scope slightly, to include the broader subject area of health technology assessment. Health technology assessment is the systematic examination of the effectiveness of a health technology, to assist decision-making in policy and practice. The outcomes of health technology assessments are intended to capture both the comparative clinical effectiveness and the cost-effectiveness of new and existing technologies against their appropriate comparator therapies. In essence, health technology assessments, which can incorporate clinical trial evidence, are aimed at answering not whether an intervention works at all, but rather whether it should be used in practice.

Modelling

Modelling is the use of a quantitative, mathematical approach to assess the potential effects of different decision options. A model is a representation of the world, based on explicit assumptions concerning the reality of a particular decision problem.¹ In the context of this review, the modelling studies focus on quantifying the comparative costs and benefits of competing health technologies.

The central role of any modelling technique is to develop a representation of the treatment area of interest at an appropriate level of detail to support the reasoning of the practitioner. The model acts as an aid, offering practitioners insight into the complex relationships between variables associated with patient pathways.

Models also enable the integration of evidence from a wide range of relevant sources, such as the incorporation of expert opinion alongside the results of primary and secondary data analyses. Another important dimension of modelling is that the assumptions and values used in the model are explicit.

The use of models provides an opportunity to explore the sensitivity of the results to variations in the assumptions that underpin the model.

Variables and parameters

The language of modelling has several other useful terms. Variables are items in the model

whose value can change, for example the cost of a medical procedure that may vary between countries. A parameter is a quantifiable characteristic associated with a variable and used in defining the mathematical model, for example the mean or standard deviation of a variable. A model also has structure, which defines the relationships between one variable and another. The model structure is based on explicit assumptions concerning the state of reality represented by the model.

Planning (a trial)

By planning a clinical trial or a wider health technology assessment study we mean such tasks as deciding on:

- definition of primary outcomes
- definition of comparator therapies
- definition and size of the study population
- length of the study.

Prioritising

By prioritising a clinical trial or a wider health technology assessment study, we mean the task of deciding whether the study is important or valuable enough to go ahead. Assessing the potential value of a trial might suggest that the resources required would be better used on another research study. Prioritising can be absolute (estimating the potential benefits or value of the study on its own terms against some kind of benchmark) or relative (ranking a series of proposed studies into priority order).

Background: why the research questions are important

Health-related research is a significant endeavour worldwide. Total funding for health R&D in England is over £2000 million, of which industry funds almost 60%, the Department of Health 15% and the research councils 10%, with the rest coming from charities and universities. Spending has risen over recent years to around 1.5% of the total health expenditure. R&D commissioning continues to be the subject of close scrutiny. In the UK, Research for Health² in 1993 identified the need for a more robust relationship between health problems and scientific multidisciplinary investigation, and a move from investigator-led research to problem-led research.

The growth in new health technologies, together with the backlog of existing unassessed therapies, far outstrips the capacity of the agencies

undertaking those assessments. Thus, there is a need for prioritisation of the health technology assessment process to ensure that resources for assessment are applied to those disease areas and technologies where an assessment exercise is likely to have most impact on decision-making.³ Furthermore, the design of funded assessments should be such that their impact on decision-making is maximised.

There is growing interest in the use of modelling methods to help to prioritise and plan research because modelling:

- **may help to make prioritisation more explicit and rational** by identifying:
 - which technologies have the greatest potential to be cost-effective
 - where crucial uncertainties lie
 - what form of research could best be used to inform these uncertainties
- **can demonstrate the key areas of data paucity** and show that these may not actually be addressed by a particular trial. As a consequence, a different and potentially simpler study could be undertaken
- **can also have a role in the design of trials**, for example, by indicating key parameters to which costs and benefits are sensitive
- **can facilitate the synthesis of evidence from diverse data sources** to address decision problems that have not been the subject of direct primary research.

However, there are also doubts concerning the value of mathematical modelling. Modelling could be costly in terms of time, scarce skills might be better used elsewhere and the models produced may not provide improvements in decision-making.

Although there have been case studies and discussion papers, a thorough methodological review in this area has not been conducted. The findings of this report and the recommendations based upon them should help:

- **to increase the appropriate use of modelling** in the R&D commissioning and prioritisation process
- in some cases, **to reduce expenditure** both within the NHS and by other bodies on trials that are either unnecessary or designed with limited added value to the evidence base. (Given that even small trials can cost between £100,000 and £200,000, the benefits of small, focused modelling exercises on the particular topic could be significant if abandonment or design changes occur)

- **to enable more cost-effective trials to be designed** by ensuring that the right questions and data are considered
- **to reduce the inappropriate use of modelling** by aiming to develop practical guidance and quality-control criteria
- **to highlight the potential of already existing models**, often built to produce an economic evaluation, but which could be reused with a focus on future research questions and trial design.

Scope of the review

Included in the review are methodological papers, discussion papers and case studies that concern the use of a modelling approach either to plan clinical trials and health assessment technology studies or to evaluate their priority (or both).

In the review, three separate but linked areas for modelling are clearly identified:

- the use of modelling to attempt to assess a particular technology, the results of which can inform future research
- the use of modelling in planning a future trial or study
- the use of quantified approaches to assess the priority of several studies competing for funding.

As time progresses and the evidence base for a particular technology increases, a developing model of a technology's cost-effectiveness can be used in all three areas.

The review is not intended to cover the literature describing standard sample size calculation, general simulation, the use of modelling in engineering design and the general literature on extrapolation, prediction and survival analysis. These areas are outside the scope of the review, unless a study has specifically involved some kind of modelling of disease or associated clinical pathways. There is also literature on priority scoring mechanisms for health research funding; this is not modelling (see definitions earlier); however, the relevant discussion papers have been reviewed because they constitute conventional current practice in R&D priority setting and thus comprise both the comparator for modelling and the context in which modelling for decision support may be used. There is also a very small body of published literature concerning

approaches taken by commercial companies. These papers are reviewed, although they are probably not representative of the large amount of (unpublished and commercial-in-confidence) effort to prioritise R&D, particularly in the pharmaceutical industry.

Audience for the report

This report is intended for three separate audiences:

- **Commissioners of clinical trials and health assessment technology in various contexts:** to support them in deciding how and when to incorporate modelling into their prioritisation and planning processes.
- **Expert modellers:** to provide a review of the state of the art, a discussion on methodological issues for further study, and a discussion of the ways in which established modelling methods can be adapted to the context of research planning and prioritisation.
- **Researchers involved in planning clinical trials and other studies:** to provide a better understanding of how a modelling approach can help to ensure that studies are well planned (i.e. they include the appropriate patient groups, number of people, length of the study and, most importantly, items measured) and hence provide enough value to make the proposed research a priority.

Structure of the report

Chapter 2 provides a detailed account of the systematic search and review methods and includes recommendations on further research requirements in the methodology of undertaking methodology reviews.

Chapter 3 reports on the roles, and indeed the value, of mathematical modelling for health technology assessment, focusing on the general health economic modelling methodology literature and case studies of models that claim some value in trial planning and prioritisation.

Chapter 4 reports a review of the literature on good practice and critical appraisal of modelling studies in health technology assessment.

Chapter 5 introduces the issue of research prioritisation within the healthcare field through analyses of previous work that has assessed

alternative approaches to the prioritisation of research. The chapter describes a range of criteria that should be accounted for in a research prioritisation process, and establishes the arguments that have previously been made both for and against alternative prioritisation processes.

Chapter 6 reviews the literature on direct attempts to assess the cost-effectiveness of research itself. The purpose of this method is to inform the priority of the research by establishing whether the proposed trial or study will be cost-effective. The principle is to consider the likely outcomes of the research and its potential benefits in improving health and healthcare. The approach attempts to address the question: 'Given a particular proposal for research, what are its estimated costs and benefits?' The likely outcomes of the research are called delta results, defined as 'a result of an assessment that can potentially cause a change in the use of the technology'.⁴ The approach has an intuitive appeal, because it is exactly the question facing research funding bodies. The chapter also includes the literature on payback, that is, retrospective assessment of the value gained by doing earlier research.

Chapter 7 reviews the literature on the value of information approach, also known as the Bayesian or decision-analytic approach. This analytical technique provides a framework that can be helpful both for the planning of trials and for prioritisation across different disease areas. It uses a well-developed methodology for decision-making under uncertainty, developed in the 1950s and 1960s.⁵ The questions addressed by this approach are: "Suppose we had more information about a new technology, how would that reduce the likelihood of making the wrong policy decision? How would it reduce the 'expected loss' associated with such a decision?" The approach is based on the decision tree for choosing between intervention treatment options (e.g. between the current standard treatment and a new experimental health technology). The consequence of each option has an estimated value that is subject to some uncertainty. The decision problem is to choose the option with the greatest expected net benefit. The decision could, however, be informed by gathering more information through further research. The value of gathering the further information lies in enabling the decision-maker to avoid making the wrong decision (e.g. incorrectly deciding to retain the standard intervention when in fact the experimental treatment is more cost-effective).

That is, the value of information is related to the reduction in the decision-maker's uncertainty. The approach requires an assessment of the 'correct' decision given the current information, together with a quantification of the probability that this 'prior to research' decision might be wrong and, finally, the calculation of the 'expected loss' incurred by making the wrong decision. The approach can be used at two levels: 'perfect information', where the uncertainty about a parameter is removed completely, and 'sample information', where research produces a better estimate with a narrower confidence interval.

Chapter 8 presents a review of methodology and case study papers concerned with the value of information approach and makes recommendations both on detailed aspects of the methodology and on its use in planning and prioritising research.

Chapter 9 reports on other miscellaneous studies, which do not fit easily into the other chapters.

These include the small amount of published work on commercial approaches to R&D prioritisation in the pharmaceutical and healthcare technology industries.

Chapter 10 presents conclusions and recommendations on the role of modelling in the process of planning and prioritising research.

Appendix 1 presents a review of case studies addressing the planning and design of future research.

Appendix 2 presents a summary of three identified modelling guidelines which describe good practice for use, and critical appraisal of decision-analytic modelling as a tool for the assessment of health technologies.

Appendices 3–5 reproduce the data extraction sheets used.

Chapter 2

Literature search and review methods

Introduction

Recent work on the methodology of methodological systematic reviews has revealed that they have a number of distinguishing characteristics compared with the more orthodox Cochrane-type review of effectiveness.⁶ These differences question assumptions about the review process and also present a challenge to the validity of methodological reviews through the potential introduction of bias. As a preface to the description of the methods of this particular review, therefore, some of these issues are confronted and their implications for other methodological reviews in health technology assessment spelt out. There are four main elements to a systematic review and attention to each of these is required for a review to be classed as truly systematic. These are: study identification, study selection, data extraction and appraisal of studies, and presentation of results.

Study identification

Issues

The assumption of a literature search in an effectiveness systematic review is that it will be comprehensive in order to identify as many items in the study population (i.e. studies) as possible.⁷ Failure to do this can result in database bias^{8,9} (selectivity occasioned by the journal coverage policies of a database), language bias^{10,11} (a predominance of a language that is more comprehensively covered by databases over other important, yet less accessible, languages) or publication bias^{12,13} (a preoccupation with articles that have positive results and that are therefore more likely to be published in a journal).¹⁴

Approaches

Approaches to counter, or at least acknowledge, these various types of bias include using the widest possible selection of database and unpublished sources, reporting references identified in other languages even if not formally covered by the review and using a funnel plot technique¹⁵ (plotting trials' effect estimates against their sample size to identify whether a certain size of study or effect size data appears to be missing).

Implications for this review

Given the size of the literature on the use of modelling, it is clearly not possible to identify all items from the methodological literature. The methodological review, therefore, attempts to be systematic in the following ways.

- Identifying all the major schools of thought in a particular area while being alert to the identification of variants, minority views and dissenters. In this review this was done through following up cited references and by using citation searching techniques. It has to be recognised, however, that the most effective way of countering a discordant argument in research terms is to ignore it. This can result in reference or citation bias.¹⁶
- Searching within a broad range of disciplines so as to bring different views (e.g. health economist, statistician, health technology assessment commissioner) to bear on the topic in hand. This review included literature searches of operational research, economics and general science databases in addition to general health and specific health technology assessment databases.
- Using a broad range of electronic and manual search techniques to ensure that materials were not missed either through the inadequacies of indexing or through selective coverage of databases.^{16,17} In addition to database searching, this review used contact with expert agencies, handsearching and the Internet.

Optimally, a methodological search will reach a point of data saturation, where no further perspectives or schools of thought are added by further acquisition of articles. However, this is more likely to occur around the use of a specific technique rather than in a broader domain such as modelling.¹⁸

Study selection

Issues

An effectiveness review typically specifies a threshold of study designs, or at the very least, a hierarchy of study designs within which retrieved items will be located. For a methodology review

study selection is more problematic; the potential range of types of article that can inform such a review is diverse. The range includes discursive articles, methodology texts and varied study designs such as individual case reports, case series and even RCTs. A methodology relating to clinical trials, that has potential benefit to health technology assessment need not necessarily have been trialled in an experimental setting. However, extra weight can at least be afforded to methodologies when a case study has established that a particular technique is feasible over methodology studies that just hint at its potential. Judgements on inclusion can be made either from a reading of the article itself or in a two-stage process involving the initial review of abstracts followed by a more detailed consideration upon receipt of candidate articles.

Approaches

Two independent reviewers usually conduct study selection for a systematic review and a test of inter-rater reliability (kappa) is performed to indicate the level of agreement. The reviewers will be provided with inclusion criteria and study-type selection or methodological checklists to ensure rigorous and consistent application of selection methods. In instances where it is not feasible for all abstracts to be read by two observers it has been known for kappa tests to be performed on a subset of articles and selection rules progressively clarified until an acceptably high degree of agreement is achieved.

Implications for this review

The large volume of abstracts retrieved and a proportionately low yield of articles from database searching meant that, once repeated kappa tests had contributed to the refinement of the study selection decision rules, selection was done by either one of two independent reviewers. Poor indexing and a low frequency of abstracts meant that a very forgiving standard was used for identification of 'candidate' articles (if there was any doubt about inclusion, a photocopy of the item was obtained). Tighter imposition of inclusion criteria was therefore applied at the stage of review of the actual articles.

The absence of criteria for selection according to study design, a characteristic of effectiveness reviews, necessitated a simple binary classification that separated methodology articles from case study articles. Within these two broad categories there were a number of subcategories that could either be established from the abstract, where present, or, more commonly, at the time the article

TABLE 1 Classifications used in the literature search

| M: Methodology | C: Case studies |
|--|--|
| M1: Discussion paper clearly about the use of modelling in prioritising research | C1: Modelling concerning prioritisation of research (pretrial) |
| M2: Paper clearly about the use of modelling in the design of research | C2: Modelling concerning the design of a trial (pretrial) |
| M3: General discussion paper on modelling but not clearly about prioritisation or design | C3: Modelling in relation to economic evaluation in HTA (post-trial) |
| M4: General discussion of research prioritisation but with no explicit modelling | O: Other |
| M5: General discussion of trial design (not included in review) | O (UN): Unconnected to review topic |
| M6: General economic evaluation discussion but with no explicit modelling | O (NK): Not known. Holding category, reclassified following receipt of the article |

was obtained and read. These subcategories are not mutually exclusive; rather, a paper is included in the highest possible ranking. Thus, a paper classified as M1, while certainly addressing prioritisation issues, may also address trial design issues (M2), but a paper classified as M2 will not discuss prioritisation issues (*Table 1*).

Data extraction and appraisal of studies

Issues

In an effectiveness review, the appraisal of studies is usually conducted according to an already published checklist with criteria covering both study quality and level of informativeness. There is an implicit assumption that good studies will often be characterised by a good level of reporting. In a methodology review there is likely to be a broad range of types of evidence, hence a single checklist orientated to a particular study design is unlikely to suffice. In addition, in case studies as opposed to methodological articles, the description of the methodology may be less well defined, as more weight is afforded to its application in a specific context. The level of information provided is likely to fall short of the ideal required by the methodology review.

Approaches

A plethora of scales and checklists have been developed, each with an apparent justification for its independent existence. However, problems arise from issues of validation and reproducibility. It is generally accepted that scoring systems, which appraise the quality of a study, are prone to oversimplification and flawed logic.¹⁹ However, a tailored checklist or data extraction form can at least achieve consistency of analysis and subsequent reporting. Finally, in contrast to quantitative effectiveness reviews, methodology reviews require an approach that handles the appraisal of qualitative statements or judgements.

Implications for this review

Purpose-specific data extraction forms were developed for the methodology and case study articles. In both cases the emphasis was on their informativeness rather than any judgement of study quality. These data extraction criteria are reproduced in Appendices 3–5.

The study team carried out an extensive literature search to try to identify a methodology by which the qualitative statements made by the various authors could be brought into a common framework for analysis and synthesis. The only qualitative review methodology that offers potential in this regard was meta-ethnography.²⁰ This technique, initially used in an educational context to compare published reports of schools, takes the researcher through progressive and iterative approaches of identification and then synthesis of major themes. This process is, however, highly labour intensive and requires considerable powers of analysis, interpretation and subject knowledge to identify commonalities expressed using different terminologies. The benefits of this particular methodology can be seen in the Results section of this chapter (for example, in the identification of a commonality in the meanings of the terms ‘delta results’ and ‘exemplar outcomes’). It is recommended that future methodology systematic reviews investigate the use of this structured approach to qualitative overview.

Presentation of results**Issues**

Quantitative systematic reviews of effectiveness are able to indicate strength of agreement, direction of results and, in many cases, a single summary estimate by drawing on a range of accepted statistical methods. In qualitative methodological reviews it is very difficult to indicate a weight that can be attached to a particular qualitative

argument or line of reasoning. For example, an attempt to attribute the importance of a particular concept according to how often it is repeated or how often it is cited is analogous to deciding the outcome of a consensus conference according to whoever speaks the loudest or most often. Having said this, there is precedence for bibliometric analysis of impact in many of the citation studies conducted by funding agencies or scientific organisations. However, it has to be acknowledged that such approaches are more suited to an evaluation of the importance of a single study than in deciding on the relative merits of conflicting or competing approaches. Quite clearly in the context of methodology, particularly in an area such as health technology assessment where established ideas or approaches are subject to continual challenges from newer and less well-known approaches, acknowledgement must be made of minority voices.

Approaches

It is interesting to note that even a methodological review aimed at looking at the methodology behind systematic reviews and meta-analysis does not attempt to attribute relative weight or importance to particular views or opinions.²¹ This has the unfortunate effect of occasionally implying equal value. The absence of a relative weight can imply that an isolated pronouncement has equal value to an established theory entrenched in orthodox dogma. Clearly, there is a need for methodologies that are able objectively to record and present different lines of thought or reasoning according to their level of acceptance or popularity. One approach exemplified in the above-mentioned report is to characterise particular schools of thought (an artificial but nevertheless practically useful concept) and then to analyse these according to supporters, variants, critics and opponents. However, it should be noted that the most popular line of thought is not necessarily the most accurate and this should be recognised in the systematic review process.

Implications for this review

It can be seen from the Results section that one school identified as being particularly prevalent and influential in this context is the cost–benefit school based on work by Eddy.⁴ Description of this work then leads to a discussion of a variant on this approach²² before rehearsing arguments and limitations of these and then exploring conflicting schools (e.g. the expected value of perfect information school). This approach can be seen to steer the reader through the essential stages of theory formulation, modification, feedback and

counter-proposition in attempting to arrive at an interpretation of the methodological review's results.

Conclusion regarding chosen methodologies

The above discussion is essential to an understanding of this review as it locates the following work in the context of an increasing awareness of the limitations of the Cochrane-type review paradigm for systematic reviews of methodology.⁶ The chosen methodologies of this review are informed by past experience in conducting a methodology review,²³ and by extensive consideration of studies of review methodology. The choices made have been pragmatic, rather than ideal, and this emphasises the need for further work in producing guidelines on methodological review, comparable to those already available from the Cochrane Collaboration²⁴ and the NHS Centre for Reviews and Dissemination.²⁵

Study identification

Search strategy for identification of studies

The search strategy employed was constructed empirically using a number of intersects of key concepts. The four concepts used were:

- concept A: modelling
- concept B: health technology assessment
concept B1: clinical trials
- concept C: priorities and prioritisation
- concept D: specific named health technology assessment organisations.

For each concept a detailed search strategy was initially developed in MEDLINE and then translated into other databases. Combining each permutation (e.g. concept A and concept B) produced an intersect. Each intersect was then evaluated on the MEDLINE database in terms of its yield and a decision was made regarding whether that particular intersect should be pursued across additional databases. Samples were taken from each intersect and the hit rate and the incremental hit rate were examined to ascertain the likely yield from each approach. It should be emphasised at this stage that the initial list of intersects was quite exhaustive in its approach, so that strategies with a poor yield could be subsequently omitted with a fairly high degree of confidence that relevant articles would not be missed by other intersects. Samples examined in order to evaluate each

intersect were taken from the more recent articles from the database as it was noticeable that there is a marked degradation in yield as one goes back in time, as a result of both imperfect indexing and the immaturity of the concepts involved. The hypothesis behind concept D was that there could possibly exist documents mentioning the activities of a specific health technology assessment organisation that did not contain any of the health technology assessment-related terms contained in concept B. However, on analysis it was revealed that this did not retrieve any relevant papers over and above those already identified by concept B, and this intersect was subsequently discontinued.

Itemising the various intersects gives the following permutations:

1. health technology assessment intersect modelling (B AND A)
 - 1a. clinical trials intersect modelling (B1 AND A)
2. health technology assessment intersect priorities (B AND C)
3. health technology assessment agencies intersect modelling (D AND A)
4. health technology assessment agencies intersect priorities (D AND C).

The intersect of health technology assessment, modelling, and priorities (A AND B AND C) was expected to yield the most specific materials, but this benefit was not realised because of the lack of specificity of indexing terminology with regard to modelling and to prioritisation processes. It was therefore decided to maximise the sensitivity of the search strategy by breaking this intersect into two separate intersects (namely A AND B and B AND C). It was considered preferable to have expert assessment of a large number of references retrieved by sensitive searches than to risk premature exclusion owing to indexing irregularities.

Other facets initially considered but subsequently rejected include:

- (a) modelling and the specific names of health technology assessment agencies
- (b) prioritisation and the specific names of health technology assessment agencies
- (c) screening and modelling
- (d) screening and prioritisation
- (e) modelling and clinical trials
- (f) prioritisation and economic evaluation.

It was also decided not to use the methodological term 'health services research' (HSR) which

embraced all HSR methodologies; the preference was to search using 'health technology assessment' as this term is used more contextually.

Decisions on the viability of the above-listed facets (a–f) were based on review of a substantial number of abstracts (typically between 500 and 1000 records). If retrieval of relevant articles fell below 1%, or if it failed to yield articles that were not already covered by the indexing of the principal intersects (1–4), searching was discontinued. It is recognised that such an approach, analogous to the implementation of clinical trial 'stopping rules', would not be appropriate for a clinical effectiveness systematic review. However, such an approach is methodologically defensible within the context of a methodological review with the following aims:

- representation of all major bodies of opinion or schools of thought relating to modelling
- an indicative, not comprehensive, sample of modelling case studies.

It is acknowledged that further research is needed with regards to exactly what constitutes a comprehensive search for methodological literature. However, we are confident that the multifaceted approach to searching described below, using subject searching, manual and automated citation searching, contact with experts and handsearching, is vastly superior to that advocated for the more demanding systematic review of clinical effectiveness:

"To be pragmatic (there is no empirical evidence supporting this), we suggest that a review with a comprehensive search uses at least 3 sources and provides a description of efforts to identify unpublished trials. A particularly effective combination could be 1 bibliographic database (e.g. MEDLINE or The Cochrane Library), a hand search of reference lists of eligible trials, and direct contact (by mail, fax, e-mail, and/or telephone) with the corresponding authors of eligible trials asking for additional published or unpublished trials. Such a review should include a discussion of the search's limitations."⁷

Databases used

The following databases were searched electronically:

- **medical and health:**
MEDLINE
HEALTHSTAR
EMBASE
HELMIS
DHSS-DATA
King's Fund Catalogue

- **health economics:**
Office of Health Economic Evaluation Database (HEED)
NHS Economic Evaluations Database (NEED)
- **general science:**
Science Citation Index
Social Science Citation Index
- **operational research:**
INFORMS
IAOR.

The systematic search generated a database of over 8000 references.

Citation pearl growing

Pearl growing is an "application of the method used for searching citation indexes, in which the index terms accompanying a located citation are used to find a new set of documents".²⁶ It is particularly appropriate for identifying a corpus of knowledge where there are known deficiencies in indexing or terminology. The efficacy of citation retrieval within health and related subjects has been established in a field study that found that citation searching added an average of 24% recall to traditional subject retrieval.²⁷

Pearl growing is analogous to the key informant technique in qualitative research; key documents are identified and then references citing these documents are retrieved and reviewed for relevance. Its limitations are similar to the key informant technique in that it relies on the prior selection of a sufficiently diffuse sample of records in initiating the process. It has been demonstrated with regard to subject searching that the more cited references used for a citation search, the better the performance, in terms of retrieving more relevant documents, up to a point of diminishing returns.²⁸

Citations to a sample of known relevant references ($n = 34$) were selected for searching using the Science Citation and Social Science Citation Indexes. These items were selected on the basis of centrality to the review topic (in terms of subject content and relevant keywords) and no attempt was made to evaluate either the extent of the contribution of a particular article or whether it belonged to a specific school of thought. Twenty-seven of the 34 articles were identified in the two citation indexes and these were cited by an average of 97 articles on Science Citation Index ($n = 2639$, range 0–952) and by an average of 43 articles on Social Science Citation Index ($n = 1160$; range 0–263). Formal evaluation of the pearl articles was undertaken only after this stage of the searching had been completed.

Cited references were retrieved and reviewed for relevance. A follow-up search was then conducted for articles citing these additional relevant items ($n = 200$) in their turn. This process was then continued for a further round before being concluded at this third level of retrieval for pragmatic reasons ($n = 425$). Ideally, this process would be continued until a point of data saturation, that is, when no new references are identified. Such an exhaustive approach is only feasible for tightly defined areas such as works by a single author or development of a school of thinking. Within the context of a diffuse and uncontrolled area of methodological writing, where the law of diminishing returns can be observed to apply in a quite striking manner, the principal value of this approach is to identify the central corpus of the literature and to cross-validate items identified through subject searching. In this way further modification of search strategies, using terms suggested by the pearl literature, could be undertaken where necessary.

Web searching

Web searching in the context of systematic reviews is still in its infancy and the only attempt to evaluate its usefulness in this context is an unpublished abstract presented by Campbell at the Systematic Reviews Symposium in Oxford in January 1998. In this particular case report, yield from the Internet for a subject search had been very poor, with the only reference additional to those identified from traditional routes being a trial that was uncompleted. Such a situation was likely to appertain even more with a methodology search, particularly where non-context-specific terms such as planning and prioritisation appear. The experience of the project team in this connection confirmed these anticipated difficulties. Nevertheless, the team did bring to bear state-of-the-art techniques and technology in including web searching in their approach. Copernic99[®], a software tool that can be downloaded onto a local machine, can be used to specify simultaneous searches of a number of the major search engines, and to store and analyse the results. Two very sensitive searches were performed to maximise retrieval:

- Plan* AND clinical trial* AND model*
- Priorit* AND clinical trial* AND model*

The results are presented in *Table 2*.

Each entry was downloaded by an information specialist using the Copernic99 software and reviewed for relevance to the systematic review

TABLE 2 Results of web searches

| | Plan* AND model* AND clinical trial* | Priorit* AND model* AND clinical trial* |
|---------------------|--------------------------------------|---|
| Alta Vista | (0 ^a) 1457 | 80 |
| Euroseek | 14 | 0 |
| Excite | 300 ^b | 0 |
| Fastsearch | 295 | 0 |
| HotBot | 300 ^b | 300 |
| Infoseek | 11 | 0 |
| Lycos | 297 | 0 |
| Magellan | 48 | 0 |
| MSN Web Search | 287 | 279 |
| Netscape Net Center | 0 | 0 |
| Webcrawler | 48 | 0 |
| Yahoo | 0 | 0 |

^a The result for Alta Vista was not recognised by the search engine syntax and therefore the search was rerun using the search engine directly.
^b Upper limit on number of results.

question. A second reviewer, a modelling expert, then reviewed candidate items.

Contacting the experts

Although contact with experts is an acknowledged part of systematic review methodology, initial results as documented in the literature have been disappointing. The Cochrane Collaboration found that contact with 42,000 obstetricians and paediatricians identified only 18 studies that had been completed outside a 2-year potential publication timeframe.²⁹ However, in more recent years, possibly with the greater awareness of the importance of systematic reviews and a corresponding increased willingness to contribute to their production, results have been more impressive. Roberts and Schierhout recently emphasised the importance of writing letters to the authors of trials in order to locate additional references for systematic reviews,³⁰ while follow-up correspondence from McGrath and co-workers³¹ found that writing 133 letters to authors of papers elicited further data in 17% of cases. However, it has to be acknowledged that the latter instances involve writing to those who are known to be involved in a clinical trial for additional data, and therefore it is expected that these would elicit a better response than the speculative efforts mentioned above which sought to establish the existence of trials.

The authors contacted 131 health technology assessment agencies identified in the 1998 Directory of Health Technology Assessment Organizations Worldwide, published by the

Medical Technology and Practice Patterns Institute. In May 1998 a standard letter was sent to 149 agencies involved in undertaking health technology assessment reviews within the UK and 109 international agencies responsible for commissioning technology assessments describing the systematic review and characterising the information required. The overall response rate from agencies was 48%. Twenty organisations, 16% of those who responded, had undertaken modelling before initiating primary research and a further 12 organisations were aware of modelling that had been undertaken in these circumstances. Twelve of the agencies contacted gave references to relevant published material, and a further 18 cited examples or case studies of early pretrial modelling.

Handsearching

Handsearching had been seen as an important method for the identification of relevant studies. To investigate its utility and to assist in the identification of target journals, a pilot study was conducted whereby expert reviewers were required to identify candidate articles from the contents pages and subject indexing of a core set of journals. The objectives of the pilot journal searches were:

- to clarify the scope of the planned systematic searches
- to inform the definition of inclusion/exclusion criteria
- to inform the definition of indexed and free-text search terms.

It was argued that if a threshold limit of potentially useful articles could be identified from any of these journals then it would be worthwhile to conduct a handsearch. An initial list of eight key journals to be handsearched, prepared for the research proposal, was thus whittled down following review of a sizeable sample (between 40 and 100) of consecutive article titles and abstracts. Each abstract listing was reviewed by individuals of the project team and articles were marked as 'In', 'Out' or 'Questionable'. These were discussed at a project team meeting and a number of issues resolved. The eight journals involved, together with their corresponding hit rates, are given in *Table 3*.

That handsearching of journals would be less productive than in the context of a clinical Cochrane-type systematic review can be seen by the fact that 90 out of the 425 items retrieved by pursuing references from pearl articles were from books, book chapters or reports. Other

TABLE 3 Yield from handsearches of chosen journals

| Journal title | Yield |
|---|--------------------------|
| <i>Controlled Clinical Trials</i> | 15 articles/100 reviewed |
| <i>Health Economics</i> | 1 article/100 reviewed |
| <i>International Journal of Technology Assessment in Healthcare</i> | 3 articles/100 reviewed |
| <i>Journal of Health Economics</i> | 1 article/100 reviewed |
| <i>Journal of Health Services Research and Policy</i> | 4 articles/43 reviewed |
| <i>Medical Decision Making</i> | 13 articles/50 reviewed |
| <i>Pharmacoeconomics</i> | 6 articles/73 reviewed |
| <i>Statistics in Medicine</i> | 1 article/100 reviewed |

characteristics of the literature were that, given the prevalence of case study articles, a large number of the pearls were from general medical journals, for example, *British Medical Journal* (19), *Journal of the American Medical Association* (19), *New England Journal of Medicine* (18), *Annals of Internal Medicine* (10), *Lancet* (5) and *Archives of Internal Medicine* (5). Nevertheless, the candidate journals identified above were strongly represented in both the methodological and case study literature, with 77 articles identified from the pearl growing approach: *Pharmacoeconomics* (29), *Medical Decision Making* (17), *Health Economics* (17) and *International Journal of Technology Assessment in Health Care* (14).

Summary and conclusions from study identification

Within the specific context of a methodological systematic review, the conclusions are as follows.

- A comprehensive **subject search** is likely to yield about one-third of all relevant articles but between 97 and 99% of references initially reviewed will prove irrelevant.
- **Citation searching** is likely to yield the highest proportion of relevant articles, with references usually being placed in context within the citing article. However, citation searching is open to a number of biases, including editorial policies over which journals are included, the propensity to cite articles that are more easily identified or retrieved, and time delays before literature makes its way into the corpus of knowledge.
- **Handsearching**, one of the proposed methods for identifying the core literature, actually proved less significant, both because of the success of citation searching and the diffusion of case studies across a wide range of general medical journals.

- **Contact with experts** yielded little material of direct relevance but had the benefit of giving the review a higher profile among the international health technology assessment community.
- **Internet searching** yielded a very high proportion of material of low relevance or low quality. However, comprehensive methods of searching using the latest technologies ensured that coverage of the World Wide Web was achieved in a systematic and efficient manner.

Study selection

Inclusion/exclusion criteria

Although provisional criteria for inclusion and exclusion were identified *a priori* and recorded in a review protocol, it was necessary to firm up these criteria with reference to general methodological issues as well as to specific items identified from pilot literature searches. Inclusion and exclusion criteria were defined in the following ways:

- review of contents listings from key journals identified in the protocol, which identified relevant literature not previously defined in the protocol
- review of the preliminary results from the principal intersects
- discussions on scope with members of the Project Advisory Group, both collectively and through in-depth individual interviews.

As a result of this process of protocol refinement, the following decisions were taken and formally documented.

- Statistical methods papers including stratification, sample size, meta-analysis and unequal randomisation would be **excluded** from the review.
- Methods for eliciting expert opinion, such as the Delphi technique, would **not be included** *per se*; however, where such papers were retrieved because of the incorporation of a modelling concept they would necessarily be included and analysed.
- Distinctions between sensitivity analysis and modelling would not be made at the searching stage but would be resolved at the data extraction phase.

Application of criteria

Initial assessment of articles for inclusion was done on the basis of article abstracts and keywords by two research assistants (FS and MS). While

TABLE 4 Kappa statistic for inter-rater reliability

| | |
|---------------------|------|
| Number of reviewers | 2 |
| Sample size | 713 |
| Kappa statistic | 0.91 |

selection criteria were being developed, validated and refined, all database printouts were reviewed in parallel by both assistants, with disagreements being resolved with reference to the project team. However, following training, review and feedback, inter-rater reliability between the two investigators was assessed on the basis of a sample of 713 articles. It was found that the kappa for their relevance judgements was sufficient to justify single observer selection, particularly since the yield of around 6% was relatively low (*Table 4*).

A total of 601 articles was obtained and read by one or more reviewers (JC, AB, FS and MS). The nature of this review and the inadequacy of indexing terminology on the major databases meant that more reliable judgements could only be conducted where the article itself was obtained. A valuable source of additional literature was 'references in context', that is, where a citation in a relevant article provided a pointer to an additional relevant reference. The context and accompanying text around these citations provided a stronger indication of relevance than many of the bibliographic citations and abstracts reviewed by the team. Nevertheless, the possibility of citation bias should be acknowledged here. The identification of an 'Eddy school', for example, may well reflect an important body of literature that is thematically and conceptually linked. However, it conceivably may reflect a citation artefact, whereby the influence of Eddy's work is artificially augmented by lack of acknowledgement of citations from critics of his work and the perpetuation of the citation pearl by his proponents.

Summary of literature retrieved

Pretrial prioritisation modelling literature is considerably less plentiful than that dealing with use of modelling alongside trials or, more commonly, post-trial modelling. This is true for both methodological literature and case study literature (*Table 5*).

Data extraction and appraisal of studies

It was acknowledged at a very early stage in the project that the characteristics of case study

TABLE 5 Results of literature search

| M: Methodology | | C: Case studies | |
|--|-----|--|-----|
| M1: Discussion paper clearly about the use of modelling in prioritising research | 31 | C1: Modelling concerning prioritisation of research (pretrial) | 14 |
| M2: Paper clearly about the use of modelling in the design of research | 8 | C2: Modelling concerning the design of a trial (pretrial) | 18 |
| M3: General discussion paper on modelling but not clearly about prioritisation or design | 222 | C3: Modelling in relation to economic evaluation in HTA (post-trial) | 420 |
| M4: General discussion of research prioritisation but with no explicit modelling | 110 | O: Other | 101 |
| M5: General discussion of trial design (not included in review) | | O (UN): Unconnected to review topic | |
| M6: General economic evaluation discussion but with no explicit modelling | | O (NK): Not known. Holding category, reclassified following receipt of the article | |

articles were quite different and distinct from those of methodology articles.

For identification of the roles of modelling a two-stage process was used. An initial form was used to identify key issues from those references identified in the first phase of citation pearl growing. A second data extraction form was then designed to collect detailed methodological discussion concerning the key issues previously identified.

The data extraction form for the case studies included bibliographic information, its place in the R&D cycle, its setting, the nature of the intervention and of the management decision, a brief description of study objectives and the objectives of the modelling process. A detailed taxonomy was used for the modelling methods, which included how the model was structured, the outcomes that are modelled, how the model was validated and how sensitivity analyses were conducted. The data extraction form concluded by recording any implications for further research identified by the study.

The data extraction for the methodology papers focusing on prioritisation between disease areas was less structured, as befitting the topic, with provision for more extensive qualitative extracts from each article. Key statements were extracted from the articles and then these were matched against dominant themes, for example, 'valuation and prioritisation of potential trials', in a process of textual synthesis. Cross-references were also made to other articles cited in support of each statement and these were then followed up, where they had not previously been identified.

Despite extensive searching of the literature there is no extant checklist that can be readily applied to the methodological or case study articles retrieved for this study. In fact, the process of evaluating articles for this review has led towards the development of such criteria.

Criteria for assessing a methodology

The criteria used for assessing the value of a methodology have been developed during the course of this review. They are partly based on the review of the literature on good practice and critical appraisal in health economic modelling. In addition, the review team has examined previous health technology assessment methodology reviews to examine the criteria applied in judging other health technology assessment-related methodologies. Finally, in reviewing the particular literature on the use of modelling in planning and prioritising clinical trials, several authors point out pros and cons of their different approaches. These statements of judgement on aspects of methodology have been reviewed systematically to add further criteria, specific to this field.

However, in the absence of an independently validated appraisal checklist, quality assessment could be seen as a source of bias. The reviewers have, therefore, concentrated on matter-of-fact descriptions of the studies together with a qualitative assessment of the place of each study within the corpus of the modelling literature.

The criteria for the assessment of the value of a methodology fall into two broad categories of theoretical soundness and practicality, which comprise the following subsections.

Time and cost

Clearly, methods to plan and prioritise research studies will take time in terms of person resources to undertake the modelling work. This will involve costs. Who should undertake the work and bear the costs is also an issue.

Delays in research

Timescales are a major issue in getting research off the ground. Investigators complain about the length of time it takes to go through the outline and full proposals process. The implications of a modelling approach for introducing further delays need to be considered.

Data availability

Some methods may attempt to use a very limited amount of data in order to keep the process manageable and time limited. Others can consider the development of extensive models particular to each different research topic. In either case, the availability of the data to populate the model is a crucial factor in the practical feasibility of an approach.

Timing of the use of the method

It may be possible to use some modelling approaches very early in the research priorities decision-making process: for example, to select broad topics. Other approaches may require detailed description of the intended health technology assessment before they can evaluate its likely value and priority.

Evidence of successful use

Clearly, where published evidence of the successful use of a methodology is in existence, this is an important factor in the recommendations.

Feasibility of achieving economies of scale by applying a generic method or model

It may be feasible to produce a generic model or method that can be used on a large set of research studies that require prioritisation. Alternatively, a separate model may be required for each individual topic.

Acceptability to health technology assessment commissioners

This category considers whether the organisations involved in designing and prioritising health technology assessments would find the processes and methods acceptable given their current *modus operandi*, resources and timescales.

Acceptability to HTA researchers

This dimension aims to establish whether research investigators would be willing to subject their

proposals to this type of modelled review and assessment. Clearly, the modelling could be seen as an extra hurdle when bidding for research funds. However, it could also have benefits in helping investigators to decide on optimal design and anticipate the likely priority of their research through some early modelling themselves.

Theoretical validity

This dimension attempts to consider whether the underlying theory involved in the method of modelling is valid for the task at hand.

Reliability

The acid test of reliability would involve applying the modelling procedure to the same study or set of studies and producing the same results in terms of recommendations for design or assessment of priorities. One important practical method of testing reliability of a broad approach is to see whether different modellers given the same potential problem would produce results of the same order.

Empirical validity

This dimension considers whether the predictions produced by a model as to the likely outcomes of research and its potential value actually come to fruition. For example, in the payback approach, retrospective analyses of the value of research involve comparing what happened as a consequence of research done with what might have happened without the research. There are difficulties involved in testing the empirical validity of such modelled predictions.

Value added, improved decisions

The essential test is whether a modelling approach can improve planning and prioritisation of research studies. In a scientific approach, one should compare the decisions without a modelling approach with those when a modelling approach is incorporated.

Presentation of results

A major finding from the analysis of the initial literature was that the literature around planning trials is quite different from that around prioritisation of trials. This was independently confirmed initial advice provided by members of the Project Advisory Group. As a result of this observation this report will be presented with separate sections for planning and for prioritisation. Within each section the literature is further divided into methodology and case

studies. It was intended that the final report would contain both qualitative textual data and tabular presentation of results. Examination of the products of the data extraction process revealed great heterogeneity between articles, and opportunities for tabular presentation were very limited. The report has, however, remained true to its original intentions in indicating the epidemiology of modelling studies and assessing the quality of items found.

Conclusions and recommendations from the systematic review process

Given the prematurity of techniques in the conduct of systematic reviews of methodology topics, this review has provided a valuable opportunity to build on previous experience of such reviews and to demonstrate important issues that continue to require further exploration.

1. Methodological reviews require an iterative approach that is informed by the extent and success of retrieval from each round of the literature searching process. To this extent they contrast with the optimal techniques required for a more classic effectiveness review where the scope is determined *a priori* by the review protocol and where modifications to the searching process will necessarily be confined to fine-tuning and refinement. The methodological review process is more akin to a Delphi technique where each successive round increases consensus and certainty with the panel of experts, in this case, being the published literature. It is important, however, to guard against 'scope-creep' and the introduction of systematic biases by recording a clear and focused question at the beginning of the process and then documenting any stages whereby an enhanced understanding leads to additional avenues for investigation.
2. The poor yield of materials from subject searching of bibliographic databases (no more than 1–2% in most instances) questions the cost–benefit of such techniques in methodology reviews and emphasises the importance of supportive techniques such as citation searching and follow-up of references. In this review the pearl-growing technique was much more productive than systematic subject searching. Nevertheless, this finding needs to be set against the recognition that pearl growing can be a source of systematic citation bias and so subject searching can perhaps be seen as a means of constructing an independent frame of reference against which the findings of the pearl growing can be evaluated.
3. The level of informativeness of abstracts as a basis for making initial judgements of relevance was very poor, requiring that a large number of potentially relevant articles had to be obtained with little eventual yield. Structured abstracts, now widely regarded as essential in biomedical journals, are still fairly rare in the methodological literature. The only database that has a field for the specific concept of modelling is the Office of Health Economic Evaluations' HEED database.
4. A related point to that in 3 above is that subject terms for modelling [in controlled vocabularies such as MEDLINE's Medical Subject Headings (MeSH)] are infrequent, inconsistent and poorly defined. The application of some acknowledged classification scheme to the subject content of modelling articles where specific techniques can be identified easily and consistently is therefore a priority.
5. Outside the subject searching and citation searching mentioned above, the subsidiary techniques such as searching the grey literature, World Wide Web searching and the survey of health technology assessment organisations yielded very little. Nevertheless, it is probable that such techniques are an important method for the independent verification of techniques such as citation searching that are prone to citation or database coverage biases.
6. Despite tremendous advances in review methodology, none of the methodological reviews so far commissioned by the NHS Health Technology Assessment Programme has been able to handle qualitative statements from source papers in a suitably rigorous and systematic manner. For example, they discriminate poorly between isolated statements from an individual that may be given undue prominence and views that are held in common across a large number of researchers. Techniques such as 'vote-counting' from the literature on a topic are clearly crude and unsuitable for a systematic and unbiased approach, while bibliometric techniques may merely serve to perpetuate existing biases. In this review the investigators began to explore use of the technique from the social sciences known as meta-ethnography. However, systematic application of these principles was constrained by the resources available and the requirement for the investigators to learn such experimental techniques. Nevertheless, it is possible that meta-ethnography may be able to assume an importance in methodological

reviews that becomes comparable to that of meta-analysis in effectiveness reviews. This certainly requires further investigation.

Conclusions regarding systematic review methodology

Study identification

The study identification strategy was iterative and multifaceted. Twelve electronic databases were searched using intersects of four key concepts (modelling, health technology assessment, clinical trials, priorities and prioritisation). The strategy also used citation pearl growing, which is appropriate for identifying a corpus of knowledge where there are known deficiencies in indexing or terminology. Contact was made with experts in 258 international and UK health technology assessment agencies. In addition, handsearching of eight specified target journals was conducted and 12 separate Internet engines were used. This search generated a database of over 8000 references.

Study selection

There was a very low yield of useful articles from the abstracts obtained from the search of electronic databases. Following repeated kappa tests and refinement of study selection rules, selection was done by one of two independent reviewers. A very forgiving standard was used for identification of 'candidate' articles (if in doubt, a copy of the item was obtained). Tighter imposition of inclusion criteria was applied at the review of copies of articles themselves. In total, 601 articles were obtained, read and classified by the reviewers.

Data extraction and appraisal of studies

The data extraction form for case studies recorded the place of the model in the R&D cycle, its setting, the nature of the intervention and of the management decision, study objectives and the objectives of the modelling process. A detailed taxonomy was used for modelling methods including model structure and the outcomes modelled. The data extraction also recorded the implications for further research identified by the study. Use of the data extraction form for the methodology studies was less structured, with provision for more extensive qualitative extracts from each article.

Synthesis of results to form conclusions and recommendations

The analyses of individual items, methodological

papers and case studies were summarised. Specific themes were identified and explored. The modelling expertise and experience of the researchers further informed conclusions and recommendations based on this systematic analysis. Papers on good practice guidelines for modelling have been reviewed and guidance more specific to pretrial modelling is presented.

Results

Results: systematic search issues

- Methodological reviews require an iterative approach that is informed by the extent and success of retrieval from each round of the literature searching process.
- A comprehensive subject search of electronic databases is likely to yield about one-third of all relevant articles, but between 97 and 99% of references initially reviewed will prove irrelevant. The poor yield of materials from subject searching of bibliographic databases (no more than 1–2% in most instances) calls into question the cost–benefit of such techniques in methodology reviews.
- Subject terms for 'modelling' (in controlled vocabularies such as MEDLINE's MeSH) are infrequent, inconsistent and poorly defined. The application of an acknowledged classification scheme to the subject content of modelling articles, so that specific techniques can be identified easily and consistently, is therefore a priority.
- Citation searching is likely to yield the highest proportion of relevant articles, with references usually being placed in context within the citing article.
- Handsearching proved less significant, because of both the success of citation searching and the diffusion of case studies across a wide range of general medical journals.
- Contact with experts yielded little further material of direct relevance, but had the benefit of giving the review a higher profile among the international health technology assessment community.
- Internet searching yielded a low proportion of material of great relevance or high quality.
- In this review the investigators began to explore the use of the technique from the social sciences known as meta-ethnography. Meta-ethnography may be able to assume an importance in methodological reviews that becomes comparable to that of meta-analysis in effectiveness reviews. This certainly requires further investigation.

Chapter 3

The use of mathematical modelling in health technology assessment

This chapter provides a brief introduction to three of the most important modelling techniques used in health technology assessment. This should help the reader with the more advanced concepts and discussions that occur later in the report. This discussion is drawn from a review of methodological literature on modelling in health technology assessment and on a review of modelling case studies that claim value in informing research design. This case study review is included in Appendix 1. For readers interested in health technology assessment but new to modelling, useful background reading and a wide set of general case studies are provided.

Introduction to the three important modelling methods

The three main forms of modelling used in the evaluation of healthcare interventions are:

- **decision analysis:** used to determine optimal strategies when a decision-maker encounters several decision alternatives under conditions of uncertainty
- **state transition modelling:** takes the form of Markov chains or Markov process models, and is used to model the natural history of a disease and the effects of a technology. These models use a finite number of discrete health states to model the disease and estimate the flow of people through these states over the full time horizon of an evaluation
- **discrete event simulation (DES):** patient-level simulation commonly using Monte Carlo analysis to model specific events experienced by individual patients.

These methodologies are used for different elements of the disease modelling and analysis process and are not mutually exclusive; for example, DES and Monte Carlo simulation techniques may be used in conjunction with the two other modelling approaches.

Decision analysis

Decision trees are the simplest of the commonly

used decision modelling techniques. As a tool for modelling relatively uncomplicated scenarios, decision trees provide a means of structuring a problem, and an effective method for combining data from various sources.

The underlying principle is the use of a decision tree to represent the available options (or decision nodes), possible probabilistic events (chance nodes) and outcomes (terminal nodes). *Figure 2* presents a simple example demonstrating the key decision tree features based on the choice between two possible interventions faced by a decision-maker. The payoffs may represent the possible costs, health benefits or health economic measures associated with the possible outcomes following from the two choices. A chance node represents any uncertain occurrence, described probabilistically, that may occur following the treatment decisions. Events may include being cured by the treatment, not responding to treatment and possibly death or the occurrence of side-effects. When deciding between the two treatments one takes into account the probability of the events occurring and the costs and benefits of the event. This decision tree approach provides a formal method for quantitatively combining these probabilities and outcomes.

Decision trees are most appropriate for modelling programmes in which the relevant events occur over a short period, or evaluations that use an intermediate outcome measure. Decision trees are especially convenient for capturing a range of unidimensional outcomes. Costs and effects are typically incorporated into a decision tree in different ways. The outcome measures of interest are generally attached to the end-points of a tree, and the proportions of patients completing the tree at the respective end-points are summed to give a measure of effect. Costs, however, may be attached to events within the tree, as well as to end-points. To calculate total costs for each intervention, the costs associated with each unique pathway in the relevant section of the tree are summed. At each chance node, probabilities, conditional on the previous event, determine the proportion of patients progressing along each

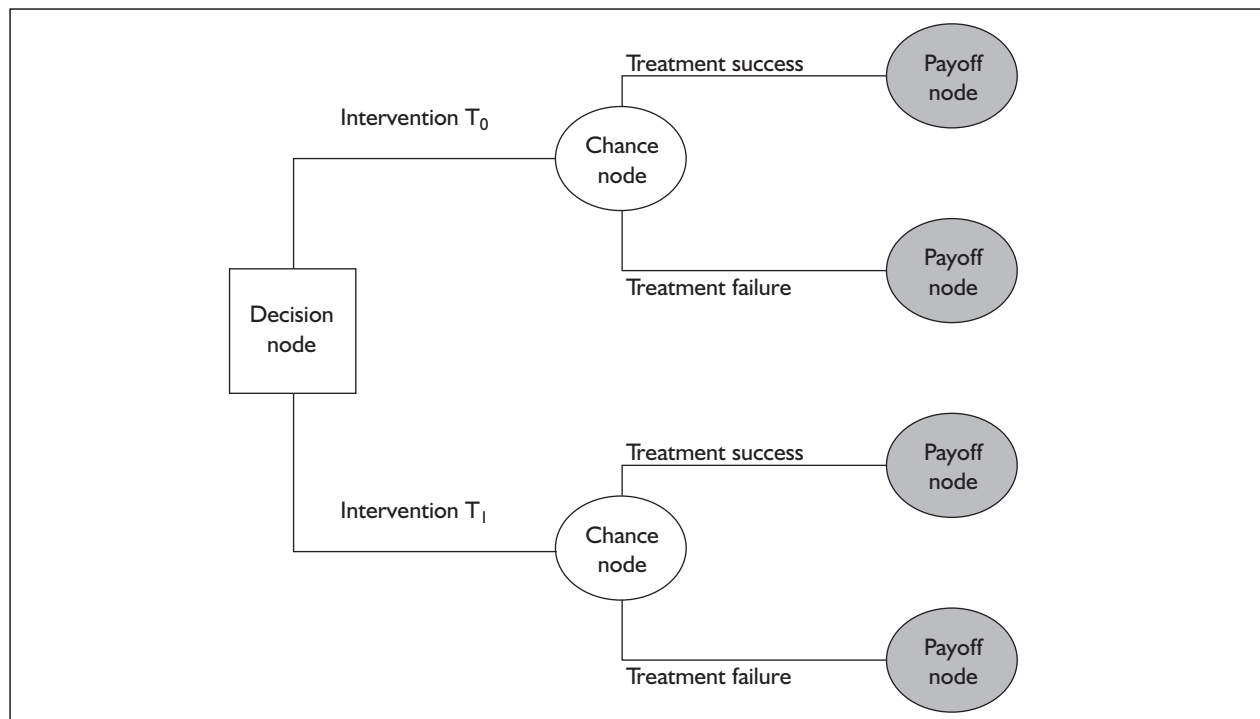


FIGURE 2 Demonstrative decision tree model

unique pathway in the tree. The process of calculating the payoffs of each possible treatment strategy, along with their associated costs and health benefits is known as folding back.

Decision analysis has a long, if not prominent, history of use within the health field, with a specific focus on aiding clinical decision-making. Clinical decision modelling focuses on identifying the key determinants of clinical effectiveness and choosing or supporting the choice between available treatment options based on clinical outcome. While the objective function, that is, the criterion for evaluating model outcomes, concerns clinical effectiveness as opposed to cost-effectiveness, the modelling methodologies used are the same as those used within modelling to support health technology assessment. Key texts describing the underlying methodology have been available for some years.^{5,32} Other authors provide useful reviews of the use of such models in health-related studies.^{33–36}

State transition modelling, including Markov models

State transition models are suited to estimating the long-term outcomes or payoffs associated with different treatment options. Within state transition models, events are modelled as transitions from one health state to another. The underlying principle of state transition models is that at any

point in time, a patient must exist within one of a finite number of predefined health states. The time horizon covered by the model is divided into cycles of equal length. At the end of each cycle a patient may move to a consequent health state, or remain in the same state. This process of moving between states continues until a patient enters an absorbing state, such as a ‘death’ or ‘mortality’ state, where they remain until the time horizon is reached. Transitions between certain health states may be restricted, and only one transition is allowed per patient per model cycle. *Figure 3* demonstrates the movements between health states.

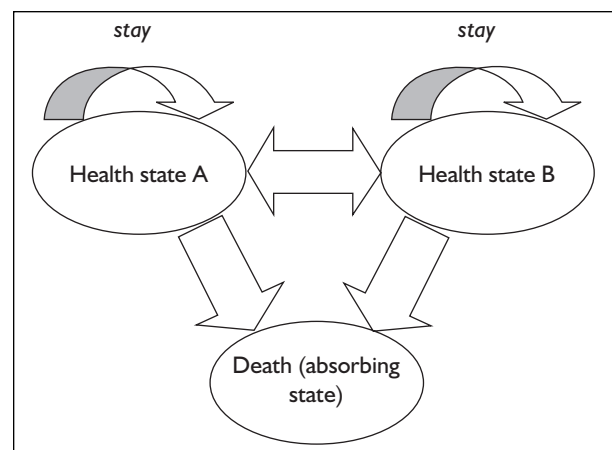


FIGURE 3 Example of possible transitions in a demonstrative state transition model

The parameters required for developing a state transition model are the probabilities of being in various states at the start of the model, and the probabilities of moving between states. Most commonly, these probabilities are grouped in matrices known as transition matrices. Utility values can be attached to each health state modelled, reflecting the severity of the state. Similarly, costs are assigned to individual health states to reflect the cost of remaining in a particular health state for the length of one cycle. The outputs of a state transition model are estimated by multiplying the respective costs and utility values by the time spent in each health state, and then summing across all possible states.

It is a common mistake for authors of economic modelling studies to describe models as Markov when, in fact, they are describing a state transition model (otherwise known as a Markov process). The distinction between these concepts is a relatively simple one: Markov models, more correctly termed Markov chains, have constant probabilities of transiting between states, whereas in the case of state transition models or Markov processes, transition probabilities may be allowed to vary according to another model variable, for example, increasing the probability of transiting to a death state as time increases. In the spirit of ensuring clarity within this review, Markov modelling exists within the broader category of state transition modelling. Although it is not inappropriate to refer to state transition models as Markov processes, state transition modelling does not use the equilibrium results that the Markov assumption enables.³⁷

An often-cited drawback of Markov models is their lack of memory, which means that the probability of moving from a particular state is not influenced by the route taken to arrive in the state (the Markovian assumption). Technically, the Markovian assumption may be overcome by splitting health states so as to describe the path taken to reach the present state, for example, state C could become 'state C after state A' and 'state C after state B'.

Using a simple Markov model to predict long-term survival from a non-start state, one is implicitly assuming a constant risk which relates to an exponentially distributed survival function. In contrast, many diseases will be characterised by either an increasing risk, in the case of a progressive disease, or a decreasing risk, in the case of an acute disease. Furthermore, the hazard

for an underlying healthy population is known not to be constant over all ages. While it is possible to include time-dependent transition probabilities, in which case the model should more correctly be termed a state transition model rather than a Markov model, this option is seldom highlighted and this aspect of the Markov model is seldom justified.

The most common approach to evaluate Markov models in the economic evaluation of healthcare technologies is the cohort method, which is an analytic approach that follows a cohort of patients through the model. At the end of each cycle the proportion of patients remaining in a state is multiplied by the relevant transition probabilities to determine how many patients move to each consequent state during the following cycle. It should be noted that the choice of evaluative method does not affect the characteristics of the modelling technique.

State transition models are often used within a decision-analytic framework when cost-effectiveness analysis is the aim. Further complete descriptions of state transition methods and their application within the health environment can be found in several useful references.³⁸⁻⁴⁰

Discrete event simulation

Both DES models and Markov processes are forms of simulation, although DES allows more complicated representations of the system being modelled. DES is event orientated, whereby the model asks what and when is the next event for every patient at each event, rather than a Markov model, which asks what events are occurring at regular intervals. DES involves entities with attributes undergoing processes, which take time and resources. For example, patients (entities) with certain age and gender and disease severity characteristics (attributes) may undergo treatments (processes) and clinical consequences (events), which take time to achieve and require physician time, drugs, etc. (resources).

DES has two main advantages over Markov models, namely increased flexibility over data requirements and an ability to overcome the restrictions of the Markovian assumption. Two potential disadvantages include the dangers of overspecifying models and the need for increased analytical input (because DES models can be analysed stochastically using first-order Monte Carlo simulation). The specification of modelling approach should be justified with relation to the decision problem being modelled.

There are dozens of excellent textbooks on general simulation approaches. The authors recommend *Computer Simulation in Management Science* by Pidd.⁴¹

The roles of modelling in health technology assessment

There is a general consensus in the methodological literature on the role of mathematical models in structuring a decision problem, in marshalling the available evidence and in providing a structure for linking the available evidence to the decision problem under concern. There is good agreement that mathematical modelling is superior to mental modelling (where an individual thinks about the pertinent information and mentally estimates the consequences of using the technology in the circumstances of interest), the only identified alternative in this context. Modelling may also support communication between those involved in policy-making, through the need to reach explicit agreement on issues such as the specified objectives, relevant interventions, model structure and the input parameter values.^{1,42}

Concerns about the conduct and reporting of modelling studies have been expressed,^{43–46} particularly with respect to the black-box nature of the working of models and the scope for manipulation. However, the fact that there have been weaknesses in published analyses is not a methodological problem with modelling, but rather a practical problem concerning the implementation, reporting and peer reviewing of such studies. These issues are very important, are strongly related to the validation and critical appraisal of models, and are discussed more thoroughly in Chapter 4.

Modelling in health technology assessment can operate in three areas (*Figure 4*):

- to assist in the actual assessment or economic evaluation of a technology
- to aid in the planning of future trials/assessments of a technology
- to assist in the prioritising of the different health technology assessments.

These three areas are not mutually exclusive and modelling can be used in an iterative sequence to inform decision-making. Models to assess the cost-effectiveness of a technology can also be used to

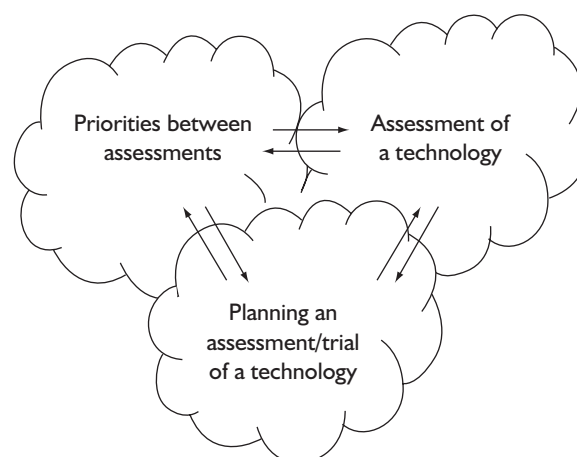


FIGURE 4 The three roles of modelling

plan and analyse the value of specific research projects such as clinical trials. Similarly, the trial results, once available, can be put into a revised version of the model to give a more up-to-date technology assessment.

The use of mathematical models to inform treatment allocation decisions in the present is the area of the most frequent application of modelling in healthcare. While the primary subject of the current review is the use of models to plan and prioritise future research, the iterative nature of the modelling process requires an understanding of the general issues around the modelling of healthcare interventions. These issues can be broadly categorised as relating either to the decision to model:

- in areas where a health technology has not been assessed directly in primary research
- where the results of experimental studies are not sufficiently generalisable to the patient population of interest to the decision-maker

or to methods used within a modelling study to improve the validity of the analysis:

- extrapolating trial results from surrogate to final outcomes
- extrapolating trial results beyond the duration of the trial.

The remainder of this section describes these issues, and methods used to address them based on the methodological literature identified as part of the systematic review and case studies included in Appendix 1.

Modelling technologies that have not been studied directly

The use of models to evaluate decision problems that have not been addressed directly has been advocated widely,⁴⁵⁻⁴⁹ and may arise from a variety of scenarios:

- early assessment problems before full-scale experimental investigation
- assessment problems not practically amenable to experimental analysis
- assessment problems where further experimental investigation is ethically unjustified.

Mathematical models can integrate data from different studies addressing alternative aspects of the disease and treatment process, and so estimate outcomes that have not been observed in any study. A great number of medical technologies present assessment problems of this type and have been successfully addressed using modelling techniques. One such example is the case of evaluating screening for cervical cancer, where the duration and reversibility of carcinoma *in situ* of the cervix are important determinants of the effectiveness of screening. While neither the duration nor reversibility can be directly observed, a mathematical model can estimate parameters for these variables from observable data.¹

Extending the generalisability of trial results to the relevant patient population

The major strength of RCTs in assessing clinical differences between treatments – high internal consistency – means that RCT results have a low generalisability to typical practice populations. In situations where experimental investigations have been undertaken, it is highly unlikely that these investigations would include all the factors relevant to a particular resource allocation decision. This leads to a number of biases in the trial population that may give rise to a biased estimate of cost-effectiveness in practice.^{48,50}

In such cases, some method for generalising the information obtained from experimental investigation is still required. Models can help to integrate the results of experimental and epidemiological studies to estimate the impact of applying a technology in a particular setting. Drummond and co-workers⁵¹ compared the impact of variation in management strategies and health delivery systems in different countries on the health economics of misoprostol therapy. In this case study modelling was used to estimate the

implications of economic assessments of this drug in the USA for the economics of treatment in a sample of European countries. The modelling yielded markedly different results concerning the potential value for money of this therapy, indicating that simply extrapolating from the original US-based study directly to the other healthcare settings, without modelling, would give very misleading economic results for this drug.

While a range of examples exist where adjustments to clinical trial results have been undertaken through modelling (Hillman and Bloom⁵² and Rittenhouse⁵³) the only complete theoretical handling has been by Eddy,⁵⁴ who classifies the potential biases as either additive or intensity based.

- Additive bias relates to patient selection bias or confounding factors, not to the effectiveness of the intervention.
- Intensity bias does affect the effectiveness of the intervention, for example, through the chosen dose of a drug, frequency of an examination, skill of a provider, type of equipment or susceptibility of a patient to a treatment.

Eddy⁵⁴ proposes a quantified approach to synthesising available evidence regarding health outcomes from a technology of interest, which he calls the confidence profile method.⁵⁵⁻⁵⁷ In principle, the approach is to define a chain of evidence that links the available data to the outcome of interest. This chain can be defined in terms of an influence diagram and may be used to link evidence together in series (e.g. linking surrogate to final outcome measures) or in parallel (e.g. constructing a meta-analysis of data from varied sources on a single outcome measure). The available evidence is then drawn together using a Bayesian updating algorithm that allows the additive and multiplicative biases to be incorporated.

The mathematical modelling of internal trial biases provides an explicit framework for the quantification of the impact of a bias. However, information on the magnitude and even direction of potential biases is unlikely to be available and there is particular need for transparency in the definition of the bias adjustments.

In summary, while there are strong cautions within the literature about avoiding potential hidden bias within model-based studies, there appears to be a consensus in the literature, ranging from strong

advocacy to an acceptance of its necessity, regarding the role of modelling in this area.

Extrapolating results from surrogate to final outcomes

In areas where the patients have a relatively long life expectancy, it is common for clinical trials to assess effectiveness in terms of an intermediate end-point,⁴⁵ such as reductions in cholesterol level.⁵⁸ Surrogate outcomes may be difficult to assess in terms of their value (the maximum value a decision-maker would be willing to pay to gain an additional unit of the outcome). Moreover, such data will not be sufficient to inform resource allocation decisions that cover a budget broader than a single disease area, where generic measure of outcome, such as the quality-adjusted life year (QALY), will be required. In addition, the relative long-term costs of the alternative interventions may significantly affect the cost-effectiveness results. For example, the outcome measure proportion surviving an episode of septic shock is likely to be an inadequate measure as the further prognosis of patients is likely to have a significant impact on the cost-effectiveness of an expensive therapy for the treatment of septic shock.^{59,60}

A range of other case studies exist, including studies in coronary heart disease⁶¹ and osteoporosis,⁶² which examine the potential effect of a technology on final outcomes through modelling from surrogate outcomes. Eddy's confidence profile method⁵⁴ involves the construction of a chain that links the technology to the short-term outcome, and the short-term outcome to the long-term outcome. The chain is populated using data from previous research relating the short-term outcome to the long-term outcome.

Particular caution is required in cases where the surrogate end-point has not been validated and does not predict the final outcomes with accuracy.⁴⁶ If the assumed relationship is critical to the marginal cost-effectiveness of a treatment then this may be examined and highlighted through modelling.

An alternative form of extrapolating non-ideal outcomes measures involves cases where the trial duration describes life expectancy, but has collected only disease-specific outcomes measures, which can be converted to a utility-based outcome measure in the form of QALYs. The Wessex Development and Evaluation Committee (DEC) report on donepezil⁶³ undertakes this type of modelling within the context of a health technology assessment exercise.

Extrapolating results beyond the trial duration

The extrapolation of short-term outcomes has much in common with the previous discussion on estimating the impact of surrogate outcomes on final outcomes, although in linking surrogate outcomes to final outcomes, while there may be an element of extrapolation over time, the focus is on transforming the unit of measurement of benefit. The extrapolation of results beyond the trial duration refers to cases in which an appropriate outcome measure can be derived from the trial (quality-adjusted or unadjusted life years gained), but the length of the trial precludes the direct estimation of lifetime survival.

The inadequacy of much empirical RCT evidence in terms of its short duration in relation to the potential impact of interventions on health and economic outcomes is a common theme across the literature. Unless it is reasonable to assume that an intervention will not have long-term effects, it is necessary to extrapolate beyond the period observed in the trial.⁴⁵

Many different modelling methods have been used in extrapolating long-term impacts and no exhaustive review has been presented within the literature. A brief discussion of three methods identified within this review is presented.

Long-term benefits associated with treatment of acute diseases have been predicted either using published life tables obtained from national statistics⁶⁴ or through estimations using the declining exponential approximation of life expectancy (DEALE).⁶⁵

This approach has been further extended to estimate potential long-term effects on mortality by including additive or multiplicative adjustments to annual risks associated with treatment. An example case study of this type of modelling is presented in an assessment of the cost-effectiveness of 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitor (statin) treatment.⁶⁶

State transition modelling

A closely related method for implementing such long-term assessments is state transition modelling, including Markov modelling. As described earlier in this chapter, such models use a matrix representation to describe the probabilities of transiting between a set of health states within defined time cycles. By tracking the flow of a group of patients through these health

states, estimates of the long-term health outcomes, resource usage and costs associated with different treatment strategies can be made.^{37,38,40}

Survival analysis

The Markov methods are particularly useful where a large number of health states have to be considered, in the simpler two- or three-state scenario, for example where patients are either alive or dead, extensions to traditional survival analysis have been used in extrapolating the long-term implications of short-term outcomes. Two papers have been identified that propose closely related methodologies for extrapolating long-term survival from restricted survival data. These are 'Parametric extrapolation of survival estimates with applications to quality-of-life evaluation of treatments'⁶⁷ and 'Survival curve fitting using the Gompertz function: a methodology for conducting cost-effectiveness analyses on mortality data'.⁶⁸ In brief, the methodology consists of fitting a parametric survival model to the tail of the survival distribution using the formal survival analysis techniques for fitting parametric models. The parametric model thus obtained is then used to project forwards beyond the limits of the available data.

Two key issues should be at the forefront of modelling in this area. First, in the absence of long-term empirical data, expert clinical judgement should be involved in defining assumptions regarding long-term risk patterns. For example, expectations concerning constant, increasing and decreasing risks should be matched to the characteristics of the specific model being used. Secondly, limitations on the analysis imposed by the assumption that long-term risks will follow the pattern suggested by the short-term data cannot be circumvented by more detailed and more complex analysis. This limitation should be borne in mind when interpreting the results of any analysis using this modelling approach.

Use of standard cost-effectiveness modelling and sensitivity analysis to inform the design of clinical research

Appendix 1 describes a review of health economic modelling case studies that specifically claim to inform the design of proposed research. The studies concentrate on the application of standard cost-effectiveness modelling and sensitivity analysis as an input to the research design process.

What aspects of trial design can modelling feasibly inform?

The review of case studies concludes that standard cost-effectiveness modelling and sensitivity analysis can inform four aspects of research design in particular.

- **Identifying key parameters for further investigation:** once a cost-effectiveness model has been developed, one-way sensitivity analysis and scenario sensitivity analyses indicate the importance of input parameters by examining the stability of the model results when the value of the parameter is varied across its plausible range. Although such analyses do not estimate the value of additional information on alternative parameters, they can usefully inform changes in trial design that do not impact greatly on the research costs. One note of caution is required, however, as a recent comparative study of alternative approaches to sensitivity analysis⁶⁹ found that these approaches tended to overestimate the sensitivity of model results and can give misleading rankings for key parameters.
- **Specifying the minimum clinical difference required for sample size calculations for a proposed trial:** a cost-effectiveness model can help to inform the minimum significant clinical difference for a trial. The analyst asks the question, 'what value would the clinical effectiveness of the new intervention have to achieve for the new intervention to be proven cost-effective versus current interventions?' The model is used to determine a threshold clinical difference and then traditional statistical methods are used to calculate the implied study sample size. For new interventions where cost is an order of magnitude greater than current care this issue can be fundamentally important because proof of clinical effectiveness alone is likely to require a lower sample size than proof of cost-effectiveness.
- **Deciding on the required duration of a proposed trial:** the estimation of the threshold clinical difference can also influence the duration of a trial because it has implications for the period required to observe the necessary differences in the number of events between interventions. A further use of modelling to determine trial duration is when a validated model enables the confident extrapolation of surrogate end-points. In this case only the surrogate end-points may be required, which reduces the duration of a trial.
- **Defining the population characteristics for a proposed trial:** models can also be used to

explore potential differential effects of a new technology across different population subgroups and thus to define target population groups for a trial. This is particularly useful in considering treatments with potentially small effects. A common mistake in these situations is to increase the recruitment and sample size by relaxing the entry criteria, when this may simply dilute the treatment effect. The model can also be used to examine the threshold clinical difference required for subgroups.

The case studies presented in Appendix 1 demonstrate the feasibility of cost-effectiveness models across a broad spectrum of disease areas, and across a broad range of intervention types. Again, there are certain forms of health technology for which modelling may offer greater benefits as an evaluative tool, including screening, diagnostics and other areas where the impact of the technology occurs over a long duration. The case studies presented also demonstrate that standard cost-effectiveness modelling and sensitivity analysis do not, however, enable the relative valuation of alternative research designs. Further methods are needed to address this problem.

Conclusions and recommendations

Disease treatment pathway modelling has been defined as the use of a mathematical model to describe the natural history of a medical condition and how the natural history is affected by different health technologies. Such models can be used to support comparative health economic evaluation of alternative healthcare technologies.

There is general agreement within the literature that modelling provides a formal structure for addressing a decision problem and bringing together available information from varied sources. This enables assumptions to be made explicit. Modelling therefore has a strong advantage over the alternative of implicit mental comparisons.

The following roles of modelling have been identified in the literature reviewed:

- structuring the decision problem
- studying assessment problems that have not previously been addressed or are not amenable to empirical evaluation

- synthesising head-to-head comparisons
- extending the generalisability of trial results to the relevant patient population
- extrapolating results from surrogate to final end-points
- extrapolating results beyond the duration of a trial.

There is a clear consensus that these roles are valid and useful in terms of performing health technology assessments, provided that guidance on good practice is followed.

Case studies of general health economic models incorporating standard one-way and multiway sensitivity analyses have been identified that claim value in informing the design of proposed research. Specifically:

- identifying key parameters for further investigation
- specifying the minimum clinical difference required for sample size calculations for a proposed trial
- deciding on the required duration of a proposed trial
- defining the population characteristics for a proposed trial.

The methods used, however, do not enable the relative valuation of uncertainty and can give misleading rankings for key parameters. Further methods are needed to address these problems.

The recognition of the potential for bias within models, for the manipulation of models, and for the need to handle uncertainty in a thorough manner has led to the development of critical appraisal guidelines for the undertaking and reporting of model-based preliminary evaluations.^{1,36,70} For instance, in using evidence from small case series as opposed to large RCTs, the results of modelling may be prone to bias. Similarly, where the link between surrogate and final endpoints is unvalidated, uncertainty concerning this link must be incorporated within the model. In the same way, where subjective judgements are used within a model, the representation of uncertainty around such judgements needs to account for potential bias in subjective estimation.⁷¹

Chapter 4

Good practice and critical appraisal of modelling studies in health technology assessment

Introduction

A key issue in the use of modelling to inform the design and prioritisation of clinical research is the reliance that can be placed on results of modelling studies. If the modelling framework used to prioritise a trial is reliable and the model successfully captures the relevant factors of a decision problem, then modelling is more likely to be accepted; if not, then the modelling approach will be undermined. The ability to distinguish 'good' modelling from 'bad' is crucial to all of the research questions under the scope of this review.

One of the major criticisms of modelling studies has been the lack of transparency in many peer-reviewed publications reporting modelling studies. This, together with the potential for bias, intended or accidental, in the model development process has led to much debate over the value of modelling studies. There is a need, therefore, for guidelines or standards in the critical appraisal of modelling studies.

Various studies have addressed alternative aspects of the modelling process. Weinstein and Fineberg's guidance for clinical decision analysis is a seminal publication,³² although earlier papers had applied the techniques of decision analysis to clinical research. In particular, Kassirer published an introduction to decision analysis in a clinical context in 1976.⁷² The main limitation of both these sources was that they concentrated on the principles of decision analysis using decision tree methodologies, only providing a brief summary of other aspects of the process, such as the derivation and assignment of probabilities. Indeed, the majority of the identified texts covering the use of decision models within the evaluation of healthcare technologies focus on the principles of the use of specific modelling techniques rather than other aspects of the modelling process.^{37,38,73}

The most useful publications relating to the process of modelling projects in healthcare were those that did not focus solely on the principles of

the modelling techniques, but rather on a framework for modelling studies. Two such studies aimed at improving the general process of undertaking and reporting medical decision-analytic models,^{1,36} while another paper aimed to increase standardisation with regard to modelling practices.⁷⁰ Each study produced good practice recommendations; synopses of these frameworks are presented in Appendix 2.

In 1999, a consensus conference on Guidelines on Economic Modelling was held at the University of Sheffield, from which a signed consensus statement was generated.⁷⁴ Papers presented at the conference were published in *Pharmacoeconomics* in May 2000, including a suggested framework for assessing the quality of modelling studies,⁷⁵ and a review of the existing literature on quality assessment in modelling.⁷⁶ There is also a more specific literature on guidelines for undertaking decision-analysis modelling in the health services domain; this details good and bad practice in the development of models and application of this specific modelling methodology. Key references in this literature are briefly discussed during this chapter.

Buxton and colleagues present five recommendations for good practice in modelling which provide a broad description of the main issues surrounding the use of modelling.⁴⁵

- The model should be kept as simple as possible to aid understanding by decision-makers.
- The presentation of results should be as transparent as possible (including submission of model and data for thorough scrutiny by reviewers).
- The quality of all the data used in the model should be made explicit.
- Uncertainty within the model should be explored thoroughly using sensitivity analysis, not compensated for.
- The model should be validated against the results of other models and/or the results of intervention studies.

Review of good modelling practice

The following sections describe suggested approaches to good modelling practice, differentiating between three broad areas: model structure and modelling technique, populating the model, and model analysis.

Model structure and modelling technique

The decision problem should be clearly stated, including the condition(s), interventions, specific study populations and study perspective, as this information will form the basis for the development of the model.⁷⁰

The primary act in the development of a model is to develop a model structure, which forms a framework for the rest of the modelling process. The structure may change as the disease pathways become better understood, or if there are inadequate quantitative data to populate the model in its initial form, although the extent to which model structure should be based on the available data is debatable. Sonnenberg and co-workers³⁶ define the practical model as “the most detailed model that can be constructed given the limitations of available data”, reflecting that changes to the structure of a model are “necessary and useful compromises”. Sculpher and co-workers⁷⁵ warn that structuring models on the basis of the quality of data available could cause the loss of important clinical events. In their view, expert opinion is always a valid means of estimating parameter values because the sensitivity of the results to changes in the parameter values can be assessed.

A general rule might state that subtle modifications to the structure of the model may be enacted that rearrange the relationships relating to the underpopulated parameters to reconcile the format of any available data. If such modifications fail to accommodate the available data then expert opinion may be sought to fill the void.

The choice of modelling technique has been identified as a potential area of divergence between alternative modellers, owing to the different types of models that are available. The choice of modelling technique will depend on whether a model is discrete or continuous, deterministic or probabilistic, or static or dynamic, as well as other characteristics such as the appropriate number of dimensions or distributional assumptions.¹

Sonnenberg and co-workers³⁶ restricted their definition of alternative modelling techniques to simple decision trees, Markov models and simulation methods. They state that decision trees are appropriate in circumstances where events occur only once and at some prespecified time. State transition models are recommended in most cases where the time horizon of a model is too long to be comfortably handled by a decision tree. The recommendation for the use of simulation methods as an alternative to state transition models is restricted to evaluations that cover complex model interactions, such as areas where patients interact or where resource availability is a relevant issue.

Although simulation models (specifically, discrete event simulation) generally enable a more flexible interpretation of disease pathways, which could provide advantages over Markov models in a number of disease areas,⁷⁷ no evidence of any important differences between the two techniques (when applied to the same evaluation) has been reported.⁷⁸

There is general agreement that the simplest modelling approach that adequately captures the necessary characteristics of the evaluation should be used.^{45,75} However, the degree to which the characteristics of an evaluation should be precisely described is not always clear, which is where the judgements of the modeller and the art of mathematical modelling may differ between analysts.¹

The scope for alternative interpretation of an appropriate modelling approach has been highlighted as a potential for the introduction of bias in model development.^{45,46} It is also important for modelling studies to justify the approaches adopted in areas in which subjective judgement has been used, for example, the estimation of uncertain parameters and their impact on final outcomes and recommendations.

Populating the model

Much of the criticism concerning the use of modelling in technology assessment has focused on the presence of bias in the data used to populate the model and its effect on the reliability of the outcomes obtained from modelling.^{45,46} In this respect, it is important to differentiate between cases in which data sources are manipulated to influence the results of the modelling process, and instances in which the data quality is simply poor. The former scenario is to be condemned, but the lack of good-quality data

to populate a model is not a criticism of the model.⁷⁵

Eddy¹ requires transparency and explicitness in the reporting of data sources for modelling studies, including a description of the data sources and a discussion of the strengths and weaknesses of each source. The parameter values, including base case, ranges or distributions for each parameter, should be reported. In reporting the results, the direction and potential magnitude of bias in the underlying data should be discussed, together with its impact on outcomes.

Halpern and co-workers⁷⁰ also require that the quality of the data used to populate a model should be explicitly assessed, including the following factors:

- availability of data (e.g. whether they have been published in a peer-reviewed journal)
- sample size
- duration and frequency of data collection
- degree of patient follow-up [i.e. all patients/intention to-treat (ITT) versus only a subset of patients]
- patient population characteristics (e.g. inclusion/exclusion criteria)
- methods of data collection (e.g. physician case-report forms, patient self-administered survey, structured interview and information from proxy)
- analysis of data (e.g. the use of ITT versus on-therapy results).

In addition, Nuijten⁷⁹ calls for information regarding the cost of access to database and data abstraction to be reported, as well as the justification for the final 'yes' or 'no' decision to use a data source, based on the advantages and disadvantages of the specific source. Nuijten⁷⁹ also makes a number of recommendations for good practice for the population of models that operationalise the concept of transparency identified by many other authors.

- Sources of study data should be recommended and explained in sufficient detail.
- A general rule should be used whereby clinical outcomes data are assumed not to be country specific. For each study this assumption has to be controlled.
- Economic data and therapeutic choices should be assumed to be country specific, requiring separate data collection.
- For each location in the model (e.g. a Markov state), the patient subpopulation has to

correspond as much as possible with the population in the data source(s) being used.

If no primary data can be identified, and expert opinion is required to populate parts of the model, the methods used to elicit expert opinion should be fully detailed.⁷⁵

Model analysis

Model analysis consists of two main issues: the need to validate the model, and the analysis of the sensitivity of the evaluation results to the modelling process adopted.

Sensitivity analysis

The appropriate form of sensitivity analysis is dependent on the characteristics of the model and the purposes of the model. Halpern and co-workers⁷⁰ identify four of the simpler approaches to sensitivity analysis: unidimensional (or one-way), multidimensional, best/worst case and threshold analysis. Felli and Hazen⁶⁹ believe that the role of sensitivity analysis has expanded beyond describing the possible impact of variations from the baseline value of an input parameter, to predicting the likelihood and effect of the uncertain outcomes.

Felli and Hazen⁶⁹ compare four more complex methodologies for assessing uncertainty in decision-making. The threshold proximity method (which plots how close baseline input parameter values are to the threshold for a decision, i.e. by how much would one parameter need to change to affect the decision made) and entropy-based measures (which describes the expected information that an input parameter yields about the whole of the model through the estimation of the mutual information between the parameter and the model) have not been applied widely, and become unwieldy when multiple parameters are tested simultaneously.

Multiway Monte Carlo (probabilistic) sensitivity analysis is assessed, where all inputs are described by parametric distributions that generate a probability distribution of the model's outputs. Individual parameter sensitivity is assessed by estimating the probability of decision change (from the baseline allocation decision) on the basis of varying a single parameter during the Monte Carlo analysis. The fourth approach is labelled an information-value-based measure, which uses the expected value of perfect information (EVPI) (for the full model and individual parameters) as a measure of uncertainty. The EVPI approach is described in detail in Chapter 7.

The application of the alternative methods to three case study evaluations found that the results of probabilistic and the EVPI sensitivity analyses were broadly similar, although the former tended to overestimate sensitivity relative to the EVPI approach. Felli and Hazen⁶⁹ believe that the EVPI approach is a natural extension to probabilistic sensitivity analysis, which provides additional benefits owing to the simultaneous assessment of the probability of making the optimal decision, and the change in payoff allied to an alternative decision.

In addition to analysing the impact of parameter uncertainty within a specified model structure, there is a need to assess the sensitivity of the allocation decision to possible variation in the structure of the model.

Model validation

Validation is a key element of the modelling process, as it is the chance for the modeller to satisfy the audience that the model is suitable for use within its defined experimental framework. Unfortunately, validation is often the most difficult phase of a modelling project. Eddy¹ describes four sequential orders of validation, which are detailed below.

Expert concurrence

The first test for a model is that the model approach is acceptable to people with a good knowledge of the area being modelled. The model structure should include factors they consider to be important, the relationships described in the model should be recognisable and data sources should seem reasonable.

Internal validity

Internal validity is also referred to as verification.^{70,80} This second validation level tests the technical accuracy of the model and should identify errors in model syntax, data entry errors and logical inconsistencies in the model specification. Mandelblatt and colleagues⁸⁰ recommend that the performance of the model is tested under hypothetical conditions, such as 100% and 0% efficacy, which should produce easily predictable results.

Estimates made by the model should also be compared with actual observations, such as a comparison of intermediate outputs from the model with the data entered into the model to check consistency. It is reasonable to define a second-order validation hurdle: any model should be able to match the data used to estimate

parameters. Failure to pass this test strongly suggests that the structure of the model is faulty.

However, it may be possible to vary some model parameters drastically and still have the model generate results that are always close to some observations. A close fit in such instances can be meaningless, and the weight to be placed on a first- or second-order validation will depend not only on the number of observations that the model can predict and the accuracy of the predictions, but also on the sensitivity of predictions to the model parameters about which there is the greatest uncertainty.

Predictions agree with non-source data

In theory, a model can be constructed using one set of existing data and tested against a different set of existing data (e.g. Schwartz⁸¹). A third-order validation compares the predictions of a model with observations that were not used to populate the model.

To avoid the trade-off between increasing the accuracy of a model by using all the available data, and using only part of the data to populate the model and enabling this form of validation, the whole data set can be used in the model when the validation assessment is completed.

Predict-experiment-compare

A fourth-order validation could be defined by comparing the outcomes predicted by a model for a new and previously unobserved research programme with the actual outcomes of that programme when it eventually is conducted. This highest level of validation is a useful approach where it is possible to undertake the experiment. One problem is that the actual conditions being experimented upon can be different from those assumed within the model. However, if the model is well structured then the important experimental conditions should be key parameters within the model. Changes in the technology itself; the age, risk and behaviour of the patient; the institutional setting; and many other factors can make such comparisons difficult. Beyond this, the random component to the outcomes of any clinical trial can prevent the predicted and observed outcomes from matching, even if a model is perfect.

To summarise, there is no simple and universally applicable procedure for validating a model. Each case must be considered by itself. In many cases only a first-order validation will be possible, and only in very rare cases will a fourth-order validation be possible. The decision to use a

model should be based on a comparison with the validity of the other techniques that may be used to assess the technology.

In some areas it may also be possible to compare the structure, inputs and results of a model to existing models in similar disease and/or intervention areas. Any observed differences between alternative models should be capable of being explained and justified. Such validation is more likely to be applicable for disease areas or technologies that have been subject to significant economic evaluation effort and hence where the modelling work may be considered relatively mature.

Consensus on principles

The broad set of principles formulated during the 1999 consensus conference for Guidelines on Economic Modelling, agreed in the form of a consensus statement, is reproduced in Box 1.⁷⁴ The principles reflect previous modelling advice

(e.g. Eddy¹ and Halpern and co-workers⁷⁰) and the views of a group of researchers with considerable experience in the use of decision-analytic modelling in healthcare.

Conclusions

- There is considerable consistency between all the identified studies that published guidelines for critically appraising modelling studies in technology assessment. This is especially noteworthy considering the time lapse of 15 years between the early Eddy study¹ and the consensus conference.⁷⁴
- All of the guidelines recognise the problems of generalisability when considering critical appraisal of modelling studies. They refrain from setting down detailed prescriptive checklists such as might be found in checklists for reporting statistical meta-analyses or reporting clinical trials. Their approach is to identify a set of principles of good practice in undertaking and reporting modelling studies.

BOX 1 Properties of good decision-analytic models⁷⁴

A good decision-analytic model for the economic evaluation of health technologies is one that:

- is tailored to the purposes for which it is to be used
- is useful for informing the decisions at which it is aimed
- is readily communicated.

A good decision-analytic model must, therefore, have certain characteristics which are summarised here.

Transparency

Transparency enables a user to examine the structure of a decision-analytic model and any incorporated data without obstacle. The analyst must make the sources of these elements (i.e. the structure and the data) clear, including any underlying theory and assumptions, and justify the choices that are made.

Internal consistency

A good decision-analytic model must be mathematically well defined for all combinations of parameter values specified as feasible. No such values should result in inconsistency in the mathematical logic of the model.

Reproducibility

Reproducibility, i.e. being able to be replicable by an independent competent analyst, thereby leading to the same result subject only to expected random variations, is an important characteristic of a good decision-analytic model.

Interpretability

A good decision-analytic model, and its results, must be clear and interpretable for the decision that it is being used to inform.

Exploration of uncertainty

The implications of all forms of uncertainty, including methodological, structural and parameter uncertainty, must be explored appropriately.

Other characteristics of good decision-analytic models

Statement of scope

The scope of a decision-analytic model should be clearly specified, including the health technologies involved, the populations addressed and the time-frame to which it relates.

External consistency

The structures used within the model, and data used to populate the model, should be consistent with the most appropriate information. The outputs of the model should be assessed in comparison with the best available relevant empirical evidence or evidence from other models.

Parsimony

A good decision-analytic model should avoid unnecessary complexity and introduce only such variables and structural components as are important to the scope of the evaluation.

Inferential soundness

The causal relationships included in the model should be explained and substantiated by the best available evidence.

- A wide range of modelling methodologies may be appropriate under different circumstances to address different problems. It is not feasible to construct a prescriptive all-purpose toolkit that strictly defines appropriate methodologies for use in any given circumstances, although Sonnenberg and co-workers³⁶ have defined general conditions where decision trees, Markov models and simulation models may be appropriate.
- In the area of sensitivity analysis, much methodological development has been undertaken, and the use of stochastic (probabilistic) sensitivity analysis is gaining popularity. However, no consensus on a prescriptive approach to sensitivity analysis has been reached. The recent health technology assessment review presents an excellent starting guide,⁸² although it is not comprehensive.
- Eddy's four levels of validation¹ provide the benchmark for validation of modelling studies. The four levels are:
 - expert concurrence
 - internal validity
 - predictions agree with non-source data
 - predict–experiment–compare.Again, however, the precise implementation of these validation levels is not defined owing to differences in the context, policy question and data availability for alternative models.
- The focus of the critical appraisal guidelines is towards transparency and explicitness in reporting of modelling studies. There are specific domains where this transparency is essential, in particular:
 - the modelling methodology used
 - the structure of the model
 - the sources of data, including subjective judgement, used to populate the model
 - validation of the model
 - analysis of uncertainty or sensitivity analysis of key outcomes.
- The need for transparency and explicitness may be compromised by the space limitations of published articles. There are often too few words, tables and figures to present a sufficiently complete picture of a typical modelling study. It is recommended that sufficient information to support peer review should be made available to reviewers. If necessary, this should be included in a supplementary report or technical appendix.
- In reviewing a modelling study it is necessary to review both the technical application of the modelling methods used and how well the underlying structure of the model reflects or incorporates known disease- or technology-specific factors. For this reason it is necessary to have both clinical and modelling input into the peer-review process and advisable for this to be coordinated.

Chapter 5

Background to the research prioritisation process

Introduction

This chapter provides an introduction to the issue of research prioritisation within the healthcare field through analyses of previous work that has assessed alternative approaches to the prioritisation of research. The aim of the chapter is to describe a range of criteria that should be accounted for in a research prioritisation process, and to establish the arguments that have previously been made for and against alternative prioritisation processes.

Two main studies that assessed alternative research prioritisation from a European perspective³ and a US perspective⁸³ are reviewed to establish the criteria that should be accounted for in the prioritisation process, which are followed by a general representation of currently applied approaches to setting research priorities.

General assessments of modelling approaches are then reviewed, followed by the various recommended approaches to prioritisation.

Criteria for prioritisation

A European Union funded research project (EURASSESS), aimed at informing the research prioritisation process, identified seven prioritisation criteria:³

- uncertainty around the impact of the technology (the plausible range of possible answers and the difference between the best and worst health impact)
- uncertainty around the financial impact of the technology
- number of potential target recipients
- timing of the impact
- trend in the technologies used during the assessment time
- improvement in equity or other ethical dimensions
- general relevance to health policy.

The authors point out that the approach taken to each of these elements depends crucially on the context in which priorities are being set. This

includes the goals of the programme, the types of assessment and technologies it covers, the nature of the organisations' funding and the allocation of responsibilities between different organisations' involved. They also point out that the general approach to priority setting currently varies across Europe and indeed the world.

A similar project commissioned in the USA reports a similar set of criteria:⁸³

- prevalence of the specific condition
- unit cost of the technologies commonly used to manage the condition (or the unit cost of a technology and its alternative)
- variation in the rate of use of a technology for managing the condition (or variations in the rates of use of the technology and its alternatives)
- burden of illness imposed by the clinical condition
- potential of the results of the assessment to change health outcomes
- potential of the results of the assessment to change costs
- potential of the results of the assessment to inform ethical, legal or social issues.

Current approaches

Figure 5 presents a schema of the steps in the prioritisation process, which has gained acceptance across Europe³ and the USA.⁸³

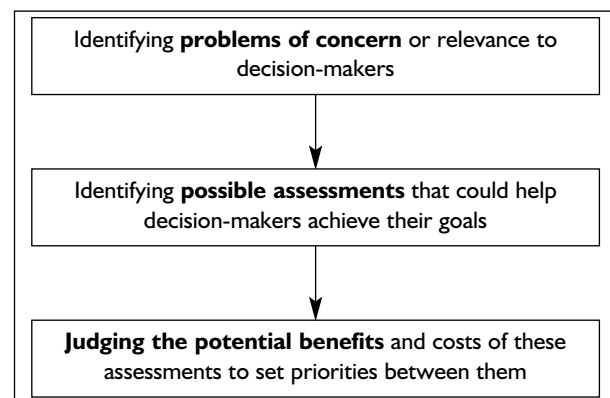


FIGURE 5 Schema of the research prioritisation process (adapted from Henshall et al.³)

As an example, *Figure 6* describes the well-defined process used by the NCCHTA in the UK to move from a broad range of possible topics to a decision on which assessments should be funded.

At present, explicit criteria for the allocation of research funds are applied only qualitatively. This general approach consists of listing a range of factors that are relevant to the prioritisation process, and then assessing the relative costs and benefits of alternative research proposals, with respect to the defined factors, through discussion among a panel of decision-makers (as described in *Figure 6*).

At the time of the EURASSESS review it is clear that the application of modelling in the formal prioritisation of health technology assessment is almost non-existent across Europe.³ Sculpher and co-workers⁸⁴ believe that, at best, research decisions take some account of the clinical and economic burden of a condition, but despite the development of various analytical frameworks there are few instances in which economic analysis of new interventions have been used to inform the development process.

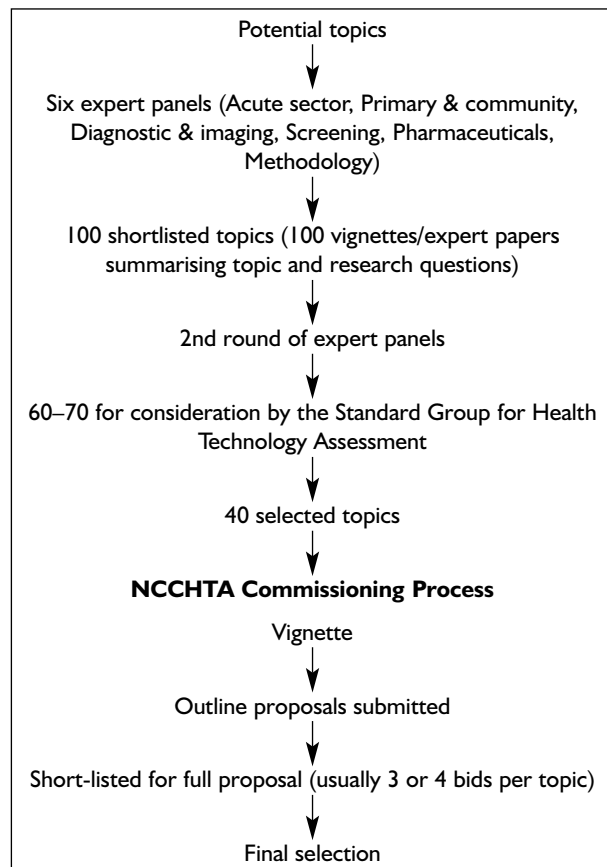


FIGURE 6 NCCHTA topic prioritisation process (adapted from Davies et al.⁸⁵)

Critique of general modelling approach

Henshall and co-workers³ discuss the role of modelling in assessment, noting that there is a complementary relationship between primary research (which can provide evidence for the input parameters of the evaluation) and the model itself (which can inform on the importance of the various parameters for future research). However, in the absence of further research into methodologies for prioritising research, the EURASSESS team is sceptical about the value of a modelling-based approach owing to the difficulties involved in quantifying the potential benefits of an intervention. A particular concern is raised over quantitative estimates based on predictions regarding the evolution of a technology and the wider implications it may have for healthcare provision. The explicit estimation of the costs of possible assessments is also thought to be subject to too much uncertainty, and so should be incorporated pragmatically as a separate criterion within the decision-making process.

They also believe that the implementation of an explicit and systematic prioritisation process would divert too many resources away from primary research, and that such a process may disengage decision-makers, especially if the process required additional time inputs from these people.

The US review of prioritisation approaches recognises the benefits of modelling, for example, data synthesis and explicitness, as well as weaknesses, for example, the use of data and ratings, even though subjectively derived, can appear more precise and authoritative than is warranted, and models can be perceived as mechanistic and insensitive to human concerns. In particular, they also raised concerns over the complexity and cost of modelling to inform the prioritisation process.⁸³

Both Donaldson and Sox⁸³ and Lilford and Royston⁸⁶ are explicit in their view that the results of any modelling evaluation should only be used to inform the prioritisation process, rather than encompass the final product of the process itself. Lilford and Royston⁸⁶ stress the interpretation of decision analysis as being part of the assessment process that informs the appraisal process, whereby the utility-based research priorities are modified according to their moral and political implications. However, they believe that the systematic nature of decision analysis provides a

defensible approach to prioritisation, citing a model of antenatal tests for foetal well-being that shows further research in this area is unlikely to detect any health benefits. Harper and co-workers⁸⁷ accept the relevance of factors other than cost-effectiveness in the prioritisation process (such as equity implications), but they are not clear as to how these factors can be included in the process.

Recommended prioritisation approaches

In their assessment of various theoretical models and practical systems that have been developed for setting priorities in health technology assessment, Henshall and colleagues³ found that few had been formally evaluated. Their recommended prioritisation approach involves the identification of priorities by decision-makers who are informed by data from various sources, including analyses of health and healthcare statistics, and the opinions of other experts to identify particular areas where health technology assessment may be of value. Preliminary reviews of existing data, to inform decision-makers regarding the existing knowledge, and the extent of uncertainty around a particular intervention, are also recommended.

Donaldson and Sox⁸³ describe their own proposed priority scoring system that combines objective and subjective criteria. Three objective criteria are defined:

- prevalence of the relevant condition
- the cost of the alternative technologies
- variation in the rates of use of the technology and its alternatives (measured by the coefficient of variation; a high coefficient of variation frequently implies a low level of consensus about clinical management).

Four criteria are subjectively defined:

- burden of illness imposed by the clinical condition (estimated as the QALY difference between those with and without the disease)
- potential of the results of the assessment to change health outcomes
- potential of the results of the assessment to change costs
- potential of the results of the assessment to inform ethical, legal or social issues.

The mechanism for synthesising these seven criteria calculates a single priority score (PS) based

on a weighted summation of each criterion score using the formula:

$$PS = W_1 \ln S_1 + W_2 \ln S_2 + \dots + W_7 \ln S_7$$

where W is the criterion weight and $\ln S$ is the natural logarithm of the criterion score. It is recommended that a panel with a broad spectrum of healthcare interests should set the criterion weights, although a subpanel with expertise in clinical epidemiology and statistics determines the objective criterion scores.

Buxton and Hanney⁸⁸ raised concerns regarding the conceptual validity of the proposed priority ranking process, while Harper and co-workers⁸⁷ regard the process as a positive step towards the quantification of the prioritisation process, but guard against the level of subjectivity required for the process.

Sculpher and co-workers⁸⁴ argue for the role of economic data in informing decisions over future research, primarily using decision-analytic modelling at different phases of health technology assessment to ensure that research adequately informs the cost-effectiveness of the interventions under comparison. The defined 'early development stage' describes the use of systematic reviews of the treatment area and informed clinical opinion to assess the potential economic impact of the proposed intervention, which identifies whether a technology is, or is not, a promising area for further research and development.

At subsequent stages, Sculpher and colleagues⁸⁴ concentrate on the use of modelling to inform trial design, although the process encompasses the decision to discontinue research if further research is unlikely to prove an intervention's cost-effectiveness. Lilford and Royston⁸⁶ implicitly raise the issue of the combination of the research design and prioritisation processes through their description of a decision analysis that shows that a trial investigating the mortality effects of screening for prostate cancer would be unlikely to be cost-effective because of the disutility effects of treatment for presymptomatic disease.

Henshall and co-workers³ realise that it is difficult to value a health technology assessment topic without some idea of the assessment design, and describe the development of a potential assessment as requiring input from decision-makers, technical experts and potential assessors in order to develop a feasible and sound methodology.

Conclusions

This chapter provides several benchmarks against which the identified modelling approaches to research prioritisation (as described in the following chapters) can be judged. First, a general consensus around the criteria that should be considered within the research prioritisation process was described, which can be reduced to three broad questions:

- How big is the problem?
- How likely is the assessment to make a difference?
- Are there any pertinent ethical, legal and social issues?

Ideally, therefore, any modelling-based prioritisation approach should provide answers to each of these questions. However, it is also noted that some authors believe that decision-analytic models can only inform the prioritisation process, and any ethical, legal and social issues could only be assessed separately from clinical and economic components.

Modelling-based approaches may also be compared to applied approaches to prioritisation which, as described in this chapter, are based on the informed views of expert panels of decision-makers. Direct comparisons between this principally qualitative approach and a quantitative modelling approach are difficult: therefore, an alternative approach would be to address stated objections to an analytical prioritisation process. Such objections include difficulties in obtaining reliable data to populate a model and, in particular, the quantification of the evolution of a technology. Other major concerns relate to the cost and complexity of modelling all alternative options for research funding.

The final conclusion derived from this chapter relates to the proposed advantages of the combination of the research design and prioritisation processes, which emphasises that competing areas of research should be represented by proposed research that is designed to answer most fully the relevant policy question(s).

Chapter 6

The direct assessment of the cost-effectiveness or 'payback' of research

Introduction

The focus of the chapter is on the evaluation of methods for the direct assessment of the cost-benefit or 'payback' of research, through the review of relevant methodological papers and case studies. The literature on methods for the direct assessment of the cost-effectiveness or likely payback of a research study addresses the question:

"Given a particular proposal for research, what are its likely estimated costs and benefits?"

The approach is analogous to that for assessing the economics of a technology itself, where the decision is whether or not to commission the research, rather than whether or not to commission the technology. This formulation has an intuitive appeal, and is exactly the question facing funding bodies as they examine each potential research bid. The questions include, "how would a certain research result change the use of the technology?" and "what would be the economic impact to the healthcare system or society (or the impact on profits to a commercial organisation)?"

The following section reviews a series of case studies that present adaptations of the basic approach to the direct assessment of the economics of proposed research projects. The final section discusses the advantages and disadvantages of payback approach.

Review of case studies

In this section, the basic payback approach, as developed in the Technology Assessment Priority Scoring System (TAPSS)⁴ is described, followed by alternative applications of the general approach.

As discussed in the methods chapter (see Chapter 2), the yield of useful papers per abstract examined was very low indeed. Following the detailed review of all of the papers obtained, the research team agreed that just seven papers addressed the direct assessment of the economics of proposed research projects. Three of these were retrospective analyses of the payback of particular research projects, but raised significant methodological issues, which contribute to the debate. *Table 6* sets out the seven papers, which are reviewed in chronological order of publication.

TABLE 6 Papers reviewed relating to the direct assessment of the cost-effectiveness of research

| Reference | Prospective/retrospective | Summary |
|--------------------------------------|---------------------------|---|
| Eddy ⁴ | Prospective | Sets out the theoretical approach of the TAPSS |
| Eddy ⁴ | Prospective | Eddy case study: maple syrup urine disease (MSUD) |
| Detsky ⁸⁹ | Both | Assessment of seven large RCTs and discussion of the relationship between sample size and the cost-effectiveness of the research |
| Drummond <i>et al.</i> ⁹⁰ | Retrospective | Case study: a post-trial analysis of the value of a diabetic retinopathy study |
| Buxton <i>et al.</i> ⁹¹ | Retrospective | Discussion paper and a series of examples assessing the payback of previous research |
| Townsend and Buxton ²² | Prospective | Case study: evaluation of a proposed trial of hormone replacement therapy (HRT) |
| Davies <i>et al.</i> ⁸⁵ | Prospective | Discussion paper on a pilot project to prioritise 25 health technology assessment pharmaceutical topics using the cost-benefit approach |

The TAPSS approach

The TAPSS approach, developed in the late 1980s,⁴ traces the pathway from technology assessment to the impact of the research results on the relevant patient population.

Figure 7 illustrates the approach. The first important component of the payback approach is the set of delta results. A delta result is a potential outcome from a proposed health technology assessment. Eddy⁴ states that the set of delta results should have four properties. They should:

- be meaningful to practitioners
- inform all health outcomes of interest
- be mutually exclusive
- be exhaustive, that is, cover all possibilities.

The second component of the approach is a framework to analyse the economic consequences of each delta result. This involves the concept of 'change in practice' or uptake of the new technology; that is, the extent to which policy-makers and clinicians will respond to different published results in terms of adopting the technology. The effect of the research will apply to a number of potential candidates for receiving the new technology and for a number of years before another technology may emerge as a policy option.

The final component of the approach is to make a prior estimate of the likelihood of each delta result. These probabilities, combined with the costs and benefits of each different delta result, are used to calculate the expected benefit of the research. Eddy⁴ is clear that these probabilities should be estimated using the existing knowledge base in whatever form. Factors to be considered include results of previous assessments, existing data and potentially subjective judgements about the true effect of the technology or biases inherent in available quantified evidence.

The level of change in use of a technology following the assessment is also a fundamental driver on the costs and benefits of each delta result. Eddy⁴ identifies a set of driving forces that impact on the scale of change in use. These include:

- results of other assessments, which are dependent on various factors, such as the credibility of the research organisations, how widely their assessments are disseminated, whether they draw the same or differing conclusions, and whether the studies are competent and unbiased;
- other policies, such as the potential inability or refusal of policy makers to fund technologies that one may rationally assume should be funded;
- subgroup analyses that may conclude that a technology is useful only for people with a specific indication;
- variations in current practice across regions, or different initial beliefs or policies relating to the technology concerned;
- the timing of uptake; for example, an assessment may expect to have little effect for the first 2 years after publication, a moderate effect in years 3, 4 and 5, and diminishing effect in years 6 and 7, with no further effect after year 7;
- lifetime of the assessment, which is affected by changes in the technology itself, development of new competing technologies, changes in the frequency or severity of the disease, and changes in the cost of the technology or its competitors.

Each of these driving forces has an impact on the likely implementation consequences of the research. The TAPSS framework is intuitively appealing in this regard because research commissioners naturally consider these driving forces as part of their conceptual list of criteria in the absence of any formal modelling.

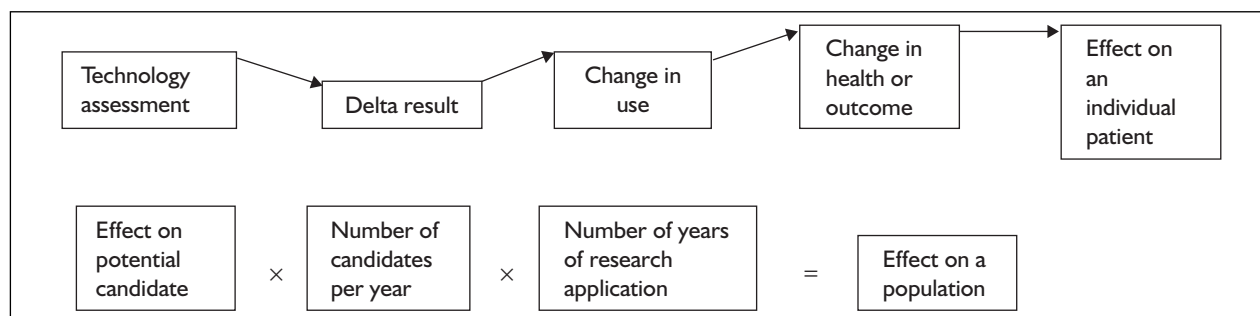


FIGURE 7 Schema representing the TAPSS approach to prioritisation

The TAPSS case study: maple syrup urine disease⁴

Eddy's original paper⁴ includes a case study to illustrate the method. The study attempts to calculate the potential impact of a technology assessment of screening for MSUD, a very rare condition that affects approximately 1 in 200,000 births and can cause death or severe mental retardation in children.

The illustrative example considers the situation in the USA, whereby one set of states had a policy to screen and another set had a policy of no screening. It considers two delta results for the research: (1) screening is cost-effective, and (2) screening is not cost-effective. If the outcome were a delta result of cost-effective, then organisations would shift to a policy of screening. This change in use of the technology would bring health benefits, which are calculated in terms of the number of severe retardations avoided. If the alternative delta result (not cost-effective) were to occur, then states would shift to a policy of no screening. Finally, the example calculates the expected impact of the assessment by assuming that there is a 50:50 chance of each delta result. The summary suggests that the assessment is expected to prevent severe retardation in approximately one additional child every 6 years (0.17 per year).

Despite Eddy's discussion paper⁴ being very clear that a range of different factors will affect these probabilities, the illustrative example opts for a very simple formulation of the problem, assuming equipoise between the two delta results. With the assumption of 50:50 probabilities for the delta results, the expected number of children with severe retardation avoided per year equals 0.17. However, if the probabilities were defined differently, then a different expected outcome would be calculated. For example, setting the probabilities to 38.5% (screening cost-effective) and 61.5% (screening not cost-effective), the expected impact would be zero severe retardations per annum; that is, the study would be expected to provide no additional health benefit.

The interpretation of the results is not explicitly addressed. The implication is that the expected impact of the assessment (i.e. 0.17 severe retardations avoided per annum) should be weighed against the costs of the policy change and the costs of the trial. Clearly, this implies that the value of further research is critically associated with the prior probabilities of the delta results.

The Detsky formulation: the impact of sample size on the cost-effectiveness of clinical trials⁸⁹

Detsky published an important development of the TAPSS approach in 1990 which considered sample size and design of trials.⁸⁹ In particular, the paper focuses on the choice of the minimum clinically significant difference (δ) between two comparators in an RCT, which has a direct consequence for the sample size required (sample size is inversely related to the square of δ). Detsky set out to examine the costs and benefits of trials, and extended this to consider the relationship between marginal changes in δ and the associated costs and benefits of research.

The paper examined seven trials, each involving treatments that could decrease the relative risk (RR) of death, measuring the cost per life year saved as a consequence of conducting each trial. The Detsky study⁸⁹ therefore covers both trial design issues (how might changing δ and the sample size influence the cost-benefit of the trial?) and prioritisation issues (what is the potential value of undertaking a specified trial?).

The estimation of the costs of the trial assumes that trial costs are directly proportional to the sample size required; however, the proportionality assumption is recognised as a limitation of the applied approach. There is no discussion of the other cost consequences of the trial results, for example, savings from reductions in adverse events, or cost increases due to the use of a more expensive technology.

However, the methods for assessing benefits of the trial of a given sample size are more sophisticated. The first key difference is that the expected difference in efficacy between the experimental treatment and the control is considered as an uncertain variable, and hence a frequency distribution of all possible risk reductions is required [denoted $P(RR)$]. This is equivalent to defining exhaustive and mutually exclusive delta results and their likelihood. It is important to note that Detsky,⁸⁹ like Eddy,⁴ considers the use of prior knowledge in making these estimates to be essential.

Secondly, the power of the trial to detect the size of the risk reduction is considered. The power of the trial is the probability that it will detect as statistically significant a real difference of a given magnitude, that is:

$$\text{Power} = p(\text{test shows up as significant} \mid \text{a real difference of a given size})$$

The smaller the expected risk reduction, the lower the probability that the trial of a given size will detect it as significant.

The third step involves calculating a particular trial's overall expected risk reduction, estimated using the formula:

$$E(RR) = \int_0^{100} \text{power}(RR) p(RR) RR \, dRR$$

the integral over all the possible risk reductions, accounting not only for their prior probabilities $p(RR)$, but also for the probability that the trial will detect the risk reduction as significant; that is, $\text{power}(RR)$.

Beyond these direct links between the size of the delta result and the sample size/power of the trial, the formulation used by Detsky⁸⁹ to assess the health benefit consequences of the trial parallels that used by Eddy.⁴ Detsky⁸⁹ estimates the life years gained as a consequence of the trial using the baseline risk, the expected risk reduction (equation above) and a life-tables method. He considers an implementation factor, equivalent to Eddy's 'change in use' idea.⁴ However, in the actual calculations he assumes that implementation will be complete. Harper and co-workers⁸⁷ strongly criticise this assumption as it may be a critical variable and could usefully be informed by expert opinion and sensitivity analysis.

Finally, Detsky⁸⁹ estimates the likely target population of new recipients per year, but unlike Eddy,⁴ he declines to predict the length of time for which the interventions would be used, owing to a lack of credible evidence.

The Detsky results⁸⁹ show that each of the trials was very cost-effective when set against the context of treatment interventions. The highest cost trial had a cost-effectiveness ratio of US\$685 per life year saved and some had a cost per life year saved of less than US\$10. However, these results underestimate the cost consequences of a trial significantly in that they include only the direct trial costs and not the costs of ongoing service delivery. To a large extent, therefore, Detsky's⁸⁹ discussion of the relative cost-effectiveness of the trials versus delivery is flawed. The Detsky sensitivity analyses of cost-effectiveness to changes in the minimum clinically significant difference show interesting results. For six of the seven trials the impact of having a larger trial (i.e. a smaller

minimal clinically important difference) was to increase the cost per life year gained. In addition, for five of the trials, if they were made smaller, the cost per life year gained would also increase. This suggests that the actual sample size and minimal clinically important difference chosen for five of the trials were close to the optimal point in terms of trial cost-effectiveness.

Detsky⁸⁹ recognises that the approach has significant data requirements, and a large number of assumptions, but believes most of the data are already contained in most trial planning documents.

In summary, the Detsky methodology⁸⁹ makes a formal link between the direct assessment of the cost-effectiveness of a trial and its sample size. The main development beyond the TAPSS approach is the focus on the relationship between the power and sample size of the study and the likelihood of changing practice.

The impact of this study has been surprisingly small. A search on the Science Citation Index found just eight papers citing Detsky⁸⁹ as a reference. Only one was a case study, the others being fairly unrelated broader modelling discussions. This combination of classical statistics in trial design and health economics has clearly not had significant uptake. However, the ideas expressed, while not covering all potential aspects of the approach fully, nevertheless represent the direct assessment approach very well and provide a credible approach to the use of modelling in planning and prioritising clinical trials.

Retrospective cost-benefit of diabetic retinopathy trial⁹⁰

This study on photocoagulation for diabetic retinopathy provides a retrospective cost-benefit analysis of a US\$10.5 million clinical trial. The trial had a complex result; a beneficial effect was seen for patients with 'proliferative' disease and high-risk characteristics, but the results were inconclusive for non-proliferative disease and those without high-risk characteristics.

The methodology used is almost identical to that described by Eddy,⁴ examining the net impact of the trial on health outcomes and costs over a lifespan for the new treatment. The paper sets up the clinical pathways in the form of a decision tree, showing the probabilities of the various events (type of disease, treatment, vision loss, etc.). In addition to the methodology, Drummond and

colleagues⁹⁰ analyse the trial from four different perspectives: the government, the healthcare sector, the patient and the community at large.

The study is useful in providing more of an insight into the feasibility and difficulties of the operational use of the Eddy⁴ framework. Even in this retrospective case study, the analysts found that examination of the scale of change in practice was difficult. No hard data were available on practice before and after the trial, and so expert opinion, the trial data and data from an epidemiological study were used to estimate change in use for patients with proliferative disease (20–46%).

Drummond and co-workers⁹⁰ stress the importance of a rigorous sensitivity analysis. The central estimate showed substantial positive health gain (279,000 vision years over 22 years), net societal savings (US\$2800 million at 1982 prices) and a direct cost in terms of healthcare investment (US\$772 million). The sensitivity analysis shows that assumptions on change in use and the inclusion/exclusion of production costs were particularly important. (A pessimistic scenario gave just 24,000 vision years, net societal cost US\$382 million and a cost per QALY of US\$53,000 excluding lost production.)

It is important to note that the retrospective analysis of the trial's costs and benefits is not necessarily a guide to its prospective value. In this example, a significant positive benefit was found for a large patient subgroup within the trial. On a prospective basis, it is likely that some delta results might show little or no net benefit. In such a case, the prospective costs per QALY would be much higher than the retrospective result.

Drummond and co-workers⁹⁰ discuss the implementation of the methodology on a prospective basis, suggesting that some of the necessary data are relatively easy to obtain:

- the population likely to benefit from implementation of therapies (i.e. the number of candidates)
- the costs of the existing treatment
- the economic consequences of the untreated disease
- the likely increase in effectiveness from the new intervention, which should be evident from the clinically significant difference being sought in the trial
- the impact of the trial on treatment practice: clinical practitioners could be asked whether

they would change their practice as a result of different outcomes from the trial

- the likely time-span of the benefits: this will relate most strongly to the development of competing technologies and a reasonable guess can be made together with upper and lower bounds
- the costs of the new therapy in practice: it is usually the case that the treatment is being undertaken in a few centres on an experimental basis and that this should give a guide to the costs in the future.

Although the case study does not explicitly analyse the probabilities of prospective delta results, the authors' experience led them to believe that using the approach prospectively is possible. They explicitly recommend its prospective use for setting research funding priorities (alongside criteria on scientific merit and investigators' capability) and suggest that the costs of this preliminary modelling would be small in relation to clinical trial costs. The authors also comment that its prospective application to more basic scientific research is much more difficult, given its potential relevance to a wide range of practical applications. However, they do suggest that the methodology could be used to quantify the potential payoff from investing in promoting the dissemination and adoption of new cost-effective clinical practices.

This case study is useful in highlighting the complexities of a single analysis of one delta result. It further serves to identify the great importance of attempting to define and analyse the prior probabilities of alternative delta results if the method is to be used prospectively.

Early research on payback in the UK^{88,91}

A retrospective assessment of the costs and benefits of cardiac transplantation research undertaken during the late 1970s and early 1980s is presented, and is another example of the approach developed by Eddy.⁴

The authors assume a slow growth in transplant surgery over a 10 year period, estimating 1056 extra heart transplants providing 4400 QALYs at a service cost of £29.3 million and a research cost of £200,000 (incremental cost per QALY £6700). There is no assessment of uncertainty; instead, there is simply a calculation of the *post hoc* change in health benefit that might have occurred as a result of the research. The main problem cited was the estimation of the counterfactual state of heart

transplant activity in the UK without the study, which required a series of subjective judgements that were assisted only slightly by the retrospective nature of the analysis.

Buxton and Hanney⁸⁸ state that the counterfactual transplant policy, level of activity and cost per transplant are fundamental variables to the estimation of the payback of the evaluation. Furthermore, they believe that attempts to quantify such variables prospectively are "probably near impossible and so contentious as to be potentially counter productive, particularly if the answer is to be used in a quantitative way to compare with other proposed research".

However, they do recommend the use of such retrospective analysis "to ask questions [on how] the likely impact of a proposed piece of research might be maximised ... and ... to indicate possible ways of improving the research and dissemination plans".⁸⁸

Prospective analysis of hormone replacement therapy trial²²

This paper examines a proposed UK trial of hormone replacement therapy (HRT). HRT provides relief of menopausal symptoms, with possible long-term protection against osteoporosis and cardiovascular disease, balanced against possible increased risk of breast and endometrial cancer. It is estimated that a UK trial, powered to identify a 25% reduction in cardiovascular events over 10 years (power 80%, significance 5%), will cost £21 million, while an international trial powered to detect differences in osteoporosis and cancer risks would cost £47 million.

The cost-effectiveness (CE) of the assessment for a particular exemplar outcome is defined as:

$$CE = (C_T + C_{PC} - C_{NPC}) / (E_{PC} - E_{NPC})$$

where C_T is the discounted cost of the trial, C_{PC} is the discounted cost of the policy change, C_{NPC} is the discounted cost given no policy change, E_{PC} are the benefits from the policy change, and E_{NPC} are the benefits given no policy change.

While the TAPSS approach requires an exhaustive set of possible delta results,⁴ Townsend and Buxton²² select three likely exemplar outcomes, each of which is assessed in turn. These exemplar outcomes are 'positive outcomes' (benefits in cardiovascular events and fractures outweigh increased risk in breast cancer), 'inconclusive' (short-term benefits exist but no clear evidence of

long-term benefit is found) and 'negative long-term' (benefits in cardiovascular events and fractures are outweighed by increased risk in breast cancer). The existing literature was used to quantify each of these possible outcomes in terms of expected risk reductions, events and QALYs gained or lost.

In terms of change in use for each exemplar outcome, the authors assume alternative scenarios. For a positive outcome the policy might be 50% of women aged 50–64 years using HRT long term. The inconclusive outcome may give HRT usage continuing along present trends rising to 30% of women (the same as one might expect without undertaking the proposed trial). The negative outcome would probably mean HRT used only for short-term symptom relief.

The analysis built on existing economic studies of HRT to examine the cost-effectiveness of the HRT trial. The results for a positive outcome show increased health service costs (£598 million), net health gain (350,000 discounted QALYs) and a cost-effective outcome (net cost per QALY of £1709). For the inconclusive outcome, there would be no changes in policy or service cost and no health improvements, but there would be a trial cost of £47 million (hence an infinite cost per QALY). For the negative outcome there would be net savings, from reduced use of HRT, alongside net health benefits since the long-term risks of HRT would be avoided.

Alternative sets of probabilities of the exemplar outcomes are tested using a scenario approach: one-third probability for each outcome (£772 per QALY); 0.5 probability for positive and 0.25 for the others (£1153); 0.5 probability for long-term negative and 0.25 for the others (£261). It is concluded that the trial is cost-effective since the costs per QALY for these three probability scenarios are lower than many common treatments. The authors also comment on the additional useful knowledge, such as profiles of health states and health service use through middle and into old age for a sizeable cohort of women taking HRT compared with those not taking HRT.

The authors suggest that this methodology could be applied elsewhere, although less evidence may often be available for forecasting potential costs and benefits. It is interesting to note Buxton's⁹¹ development in confidence concerning the application of this method. As noted above, Buxton and co-workers⁹¹ thought that the

prediction of the uptake of an intervention was virtually impossible, although by the time of this 1997 paper, Townsend and Buxton²² conclude that this methodology could be a good response to the growing demands on the research community to show that there is a good return on the investment in health services research.

A potential criticism of the approach used here concerns the simplicity of the exemplar outcomes. There is no separate analysis of different patient groups (e.g. women with/without a hysterectomy), for whom the delta results may be different. Even though the overall cost-effectiveness results appear wholly conclusive, further detailed work might have brought an additional understanding.

Perhaps the most important criticism concerns the limited sensitivity analysis on the probabilities of these three exemplar outcomes. The 'most realistic' scenario gives higher weight to the positive outcomes and results in £1153 per QALY. However, the inconclusive outcome is risky from a health service perspective. The present investigators undertook a simple threshold-style sensitivity analysis, which shows that even with an 80% probability of 'inconclusive', the trial would have an expected cost per QALY of £140. Indeed, the probability of achieving an inconclusive result would have to be over 98% for the cost per QALY of the proposed trial to rise above £10,000.

In summary, the approach of Townsend and Buxton²² is fairly intensive. It builds successfully on previous health economic evaluations of HRT to examine the potential cost-benefit of a trial and concludes that the proposed trial would be worthwhile.

UK pilot study on prioritising research topics⁸⁵

This recent UK paper presents the methodology and results of a pilot study conducted to assess the feasibility of estimating broad cost and outcome data to inform the first round of the prioritisation process; for example, the NCCHTA's attempts to reduce a shortlist of 100 research topics to around 40 priorities. The objectives of the study were to analyse comparative information on the value for money of health technology assessment questions/topics, provide useable information for decision-making, and identify critical factors that determine the value for money of specific assessments. The pilot work was undertaken on the pharmaceutical panel topics for 1997 and 1998, a total of 25 topics.

The following variables were included:

- the probability that new intervention will be proven effective or ineffective by the health technology assessment
- the likely utilisation of new technology in future
- the probability of other competing new technologies emerging and their rates of utilisation
- the maximum lifetime for the new technology
- the transition costs of adopting the new intervention
- the costs of the health technology assessment research itself
- the intervention costs for 1 year incidence or prevalence cohorts
- the benefits of the intervention.

The methodology used in this pilot study has several defining features; in particular, the delta results were divided into just two possibilities: (1) the technology is proven effective, or (2) the technology is proven ineffective. Most importantly, the probability of each technology being proven effective was assumed to be the same, at 67%. This value of 67% came from a study examining the rate of new pharmaceutical compounds that were successful in Phase III clinical trials during the 1980s.⁹²

The change in use or level of adoption for each technology was related to whether the health technology assessment was exploratory, whether coherent implementation was planned, and the likely scale of transitional costs of implementing the particular technology. To inform the counterfactual, in the absence of data describing current utilisation, low rates of existing utilisation were assumed for new technologies (3–5%). If proven effective, the rate of utilisation was assumed to increase by 16% per annum. If proven ineffective, the annual rate of utilisation was assumed to reduce to 2% per annum.^{93–96}

Transition costs of adopting the new intervention were assumed to be zero unless significant infrastructure investment was required in staff, equipment or facilities. Data describing the health benefits of the different interventions could not inform a consistent measure of patient benefit, and so different measures were used, including cases/deaths averted, life years gained, and people with improved symptom control.

The results of implementing the method on 25 topics are shown in *Table 7*. In terms of feasibility, the method was operationalised for the majority of

TABLE 7 Expected costs and benefits of health technology assessment (HTA): results of payback analyses, pharmaceutical panel, 1997–8 (from Davies et al.⁸⁵)

| Topic no. | Net expected cost (range) (£ million) | Net expected benefit of HTA (improved symptom control) | ICER (£ per unit of outcome) | Critical determinants of cost/benefits of HTA | | |
|-----------|---------------------------------------|--|------------------------------|---|--------------------------------|------------------------------|
| | | | | Cost/benefit of interventions | Effectiveness of interventions | Utilisation of interventions |
| 98-A | +307 (–695, +1678) | 0.6 m | People with ISC | 494–3020 | Y | Y |
| 98-B | +434 (–88, +956) | 5800 | People with ISC | 2917–16,457 | Y | Y |
| 98-C | –315 (–823, +844) | 0 | | NA | Y | Y |
| 98-D | +2 (–45, +50) | 0 | Years with ISC | 0–746 | Y | Y |
| 98-E | +104 (–118, +339) | 2518 | Cases averted | 41,000–224,133 | Y | Y |
| 98-F | –9 (–409, +652) | 7030 | People with ISC | 0–92,745 | Y | Y |
| 98-G | +1122 (+357, +1887) | 13,432 | People with ISC | 26,600–140,500 | Y | Y |
| 98-H | +136 (–775, +600) | 1819 | People with ISC | 74,876–0.3 m | Y | Y |
| 98-I | +27,309 (+10,769, +43,849) | 290,484 | People with ISC | 37,072–150,954 | Y | Y |
| 98-J | –59 (–28, –117) | 0 | | NA | Y | Y |
| 98-K | +364 (–549, +15,079) | 4378 | Cases averted | 83,082–3.4 m | Y | Y |
| 98-L | –860 (–239, –1238) | 0 | | 0 | | |
| 98-M | +181 (–647, +1005) | 5781 | Cases averted | 27,568–173,804 | Y | Y |
| 97-A | –1005 (–1947, +223) | 0 | | NA | Y | Y |
| 97-B | +42 (+11, +50) | 1160 | Life years gained | 9353–53,323 | Y | Y |
| 97-C | +55 (–41, +84) | 448 | Deaths averted | 0–187,500 | Y | Y |
| 97-D | +182 (+2, +337) | 13,019 | Deaths averted | 1313–25,892 | Y | Y |
| 97-E | –167 (–167, +118) | 0 | | NA | Y | |
| 97-F | +333 (+318, 358) | 3414 | Deaths averted | 48,990–104,730 | | Y |
| 97-G | +9 (+8, +10) | 81 | Deaths averted | 94,174–117,949 | Y | Y |
| 97-H | –0.2 (–2, +3.8) | 0 | Cases averted | 0–4445 | Y | Y |
| 97-I | +68 (–51, +70) | 1252 | Life years gained | 0–55,910 | | Y |
| 97-J | +1924 (–75, +1924) | 3380 | Life years gained | 0–569,183 | | Y |
| 97-K | +200 (+19, +259) | 8740 | Life years gained | 0–29,558 | | |
| 97-L | +15 (–203, +15) | 0 | | NA | Y | Y |

ISC: improved symptom control; ICER: incremental cost-effectiveness ratio; NA: not applicable.

topics studied (approximately 70% of the topics considered). Some analyses were not possible owing to uncertainties about the proposed research, which could mean that the topic description (vignette) was not clear enough to specify particular interventions and research questions.

Because of time constraints, several parameters had to take default values, for example annual utilisation rates and probabilities of effectiveness, which were based on a limited number of published sources. The authors concluded that it is feasible to conduct *ex ante* assessments of the cost-effectiveness of health technology assessment for some topic areas, and they called for further research into the value of providing decision-makers with quantitative estimates of payback of health technology assessments, versus softer qualitative approaches for the prioritisation of research portfolios.

The most important issue arising from this pilot study is how to incorporate uncertainty into the decision-making process, as sensitivity analyses found that variations in the assumed values of variables such as intervention costs, likely levels of utilisation, and the probability that the technology would be found effective, had large effects on the observed results.

Davies and co-workers⁸⁵ suggest that projects with net costs and negative benefits, or projects with excessive cost per QALY would not be priorities. For remaining projects, they propose categorising projects as, for example, those always net saving with positive benefits, those with costs per QALY within a defined range, and those with results sensitive to changes in input parameters. Here, the authors are heading towards a categorisation of the potential research on the basis of uncertainty, which is based on the recognition that uncertainty determines the need for health technology assessment and that accounting for uncertainty should also underpin methods for prioritising health technology assessment.

Methodological debates

This section reviews the case study material in the payback literature to compare systematically the methods used and examine the level of consensus.

Perspective of the analysis

All of the case studies took the view that health and social services costs and measures of health benefit should be included. Two studies took a wider perspective. Drummond and colleagues⁹⁰

examined cost and benefits of the research to the government, the healthcare sector, the patient and the community at large, and sensitivity analysis including lost work time. Buxton and Hanney⁸⁸ present an even broader conceptual model of the forms of payback from health services research. These include knowledge, benefits to future research (targeting future research, developing skills, etc.), political benefits (improved information or decision-making), health sector benefits and broader economic benefits (including commercial exploitation of innovation and economic benefits from a healthy workforce). The research prioritisation literature does not discuss perspective in any depth and the debate is equivalent to that for broader health technology assessment. However, almost all of the case studies focus exclusively on health sector benefits and costs, ignoring both production losses and indirect costs.

Identifying the possible outcomes of a trial: 'delta results'

There is some disagreement within the literature regarding the specification of the delta results. Eddy⁴ states that delta results should be meaningful to practitioners, carry information on the actual effect of the technology on the outcomes of interest, be mutually exclusive; and cover all possible outcomes. The latter requirement, for delta results to cover all possible outcomes, is the most contentious. A simpler alternative is to include only two delta results, namely, the technology is proven cost-effective or not cost-effective. Indeed, Eddy's MSUD case study⁴ was of this limited form. Some studies incorporate the very important possibility of inconclusive research results. Townsend and Buxton²² use three "exemplar outcomes" (positive, negative and inconclusive).

Three of the case studies were retrospective, and these only consider the actual trial because the trial was already complete at the time of the analysis of the payback from research.⁸⁸⁻⁹⁰ These retrospective analyses show the importance of covering all possibilities, for example, the diabetic retinopathy study⁹⁰ has a relatively complex result, where the technology is effective for one patient subgroup but results are inconclusive for two other subgroups. The Detsky formulation,⁸⁹ undertaken on seven different trials, is the most sophisticated in its assessment of delta results as it allows for different sizes of delta result (in this case mortality risk reduction) rather than just positive or negative.

In contrast, the largest attempt at a prospective analysis has also used the simplest formulation of

delta results.⁸⁵ The UK pilot study on 25 topics simply used two options: technology is proven effective or not effective. It is clear that there is a tension between the sophistication of approach and the time available to undertake the study. However, the loss of a coherent logic, if too many simplifications are made, is a serious risk, and may well invalidate any results.

Identifying the probability of each delta result

Identifying the probability of each delta result is often the most arbitrary component of the existing case studies. Eddy⁴ assumed a 50:50 chance of each delta result (ignoring the possibility of inconclusive research). Townsend and Buxton²² specified three equally likely alternative sets of probabilities for positive, inconclusive and negative outcomes, and undertook some scenario analysis on different likelihoods. In the UK pilot study on 25 pharmaceutical topics,⁸⁵ the probability of proving each technology effective was assumed the same at 67%, based on a study of success rates for new pharmaceutical compounds in Phase III clinical trials.⁹² Only Detsky⁸⁹ considered the difference in efficacy between the experimental treatment and the control as an uncertain variable, and specified a probability distribution to describe the range of possible delta results. However, even here the sources of evidence for these probabilities were not discussed in detail.

None of the case studies based their prospective estimation of the trial results on data specific to the trial being assessed. However, most well-designed RCTs are informed by prior research and it is recommended that where knowledge exists regarding the most likely outcome of a prospective trial, this should be used. Arbitrary allocation of these probabilities within the cost-benefit framework is a major issue for the validity and robustness of this approach. In fact, identifying the probability of each delta result is a connecting point between the direct assessment of cost-effectiveness approach and the value of information approach. This is discussed in more detail in Chapter 7.

Change in health outcomes for each delta result

There is a clear consensus in the case studies that the health outcomes associated with each possible delta result and indication (i.e. patient subgroup) should be described. The depth of the analysis varies substantially and the prime determinant of this variation appears to be availability of data.

The method of estimation is sometimes made easier by the way in which the delta results are defined. For example, the HRT delta results are defined as exemplar outcomes with a certain scale of effect in cardiovascular events, fractures, risk of breast cancer, etc.²² For the retrospective analyses, the methods are based on the clinical trial's actual results. Thus, in the early payback work,⁸⁸ it was possible to estimate the QALYs gained for each cardiac transplant.

Each particular topic presents different challenges to a modeller. Detsky⁸⁹ estimated life years gained using the baseline mortality risk, the expected risk reduction and a life-tables method. For the diabetic retinopathy trial, the health benefits were estimated using a decision tree approach.⁹⁰ The sensitivity analysis in the diabetic retinopathy trial showed that the value of the trial was significantly affected by its assumed effectiveness in subgroups where inconclusive results had been discovered. Even in this retrospective analysis, a significant factor in the calculation of the trial's value (i.e. the 'true underlying effectiveness' in the 'inconclusive' study groups) is still an unknown following the research.

The approaches and methods (discussed in Chapters 3 and 4) for the use of modelling in health technology assessment in general are the tools of the trade for estimating the health outcome consequences for each delta result. In each case, a well-informed estimate of the benefits will increase the validity and robustness of the final result concerning the potential cost-effectiveness of the research.

Which measure of health benefit to use?

Townsend and Buxton²³ examined QALYs as the measure of health outcome for each of their three delta results on the HRT trial. Of the 25 pharmaceutical topics examined by Davies and co-workers,⁸⁵ they found that the available data were not sufficient to generate the same measure of benefit for all topics. However, in order to prioritise research across different disease areas, the measure of health benefit needs to be generic rather than disease specific. It is recommended that health outcomes should be described generically wherever possible, using informed opinion if necessary.

Change in use: uptake following technology assessment

Eddy⁴ suggests that change in use (uptake) can depend on the following: results of previous

assessments, policies and policy-makers' perceptions, patient types, geographical regions and time periods. The MSUD case study⁴ presents assumptions for the likely quantified impact of each of these issues. Townsend and Buxton²² used broadly the same approach, designing a likely scenario of implementation for each delta result. Detsky⁸⁹ used an even simpler approach, assuming 100% implementation of trial results in practice. Davies and co-workers⁸⁵ assumed the change in use to be the same for each technology, which was based on a generic estimate of the uptake of trial results.

Retrospective studies of research payback also have to estimate technology uptake. Drummond and colleagues⁹⁰ formally sought expert opinion to estimate the difference in practice with or without the trial, owing to a lack of hard data. Sensitivity analysis showed these assumptions to be very important. Buxton and Hannay's retrospective study⁸⁸ also found change in use (in cardiac transplantation) to be their main problem. Clearly, this is a concern for the validity and robustness of the payback approach (and also for the value of information approach).

Sensitivity analysis and uncertainty

Each component of the payback approach may be subject to uncertainty and sensitivity analysis could be useful in:

- identifying the possible outcomes of a trial (delta results)
- identifying the probability of each delta result
- change in health outcomes for each delta result
- change in use: uptake following technology assessment.

Some of the case studies consciously avoid considering uncertainty, for example, Eddy⁴ does not describe the effects of uncertainty at all. Others undertake limited sensitivity analyses on certain components, the most common being the probability of each delta result. Townsend and Buxton²² undertook a partial sensitivity analysis concerning differing values for the probabilities of their exemplar outcomes, but did not consider any uncertainty analysis in the change in health outcomes or the change in use. Detsky⁸⁹ undertook a probabilistic sensitivity analysis around the likelihood of each outcome, but did not examine any uncertainty in change in use. Some authors use sensitivity analysis across the range of components in the method; for example, the Drummond study⁹⁰ undertook a more

rigorous analysis of uncertainty, including one-way sensitivity analyses on the uptake of treatment, the costs of treatment, rehabilitation and societal costs. In addition, multiway sensitivity analysis was undertaken, making conservative assumptions about several variables at the same time.

How to do the sensitivity analysis is a problem, but even more pressing is the interpretation of the results. All of the authors struggle when interpreting uncertain results in terms of the trial's cost-effectiveness. What if the estimated cost per QALY of the trial ranges from say £2000 to £2 million? Should the trial be funded? Fundamental decision theory states that the action that gives the maximum expected benefit should be chosen. However, the methodological inadequacy of ignoring both the uncertainty around the expected benefit and the opportunity loss of making the wrong policy decisions remains.

Suggested approaches for categorising projects (e.g. those always resulting in a net saving with positive benefits, those with costs per QALY within a predefined range, those with results sensitive to changes in input parameters) do not resolve this issue of interpretation.

What really matters is that the trial itself provides extra information that can help researchers to avoid making wrong decisions into the future. It is the impact on changed future decisions, combined with the probabilities of making the wrong decision given current information, that is the key to the value of a trial. This is not addressed at all through this approach to the direct assessment of the cost-effectiveness of the trial. It is the subject of the value of information approach covered in detail in Chapter 7.

Uncertainty in the potential value of the payback approach

Eddy,⁴ Drummond and co-workers⁹⁰ and Townsend and Buxton²² all found the same problem, which can be expressed as follows:

“How do we interpret estimates of the expected costs and benefits of health technology assessments? and what do we do if they are uncertain?”

However, this is often exactly the situation in which research is needed. If it were possible to estimate the consequence of the health technology assessment with narrow confidence intervals then it would be likely that further research should be unnecessary. It is the uncertainty that determines the need for further health technology assessment.

Many of the practical difficulties raised by the case studies presented in this chapter centre on the difficulty of characterising prior uncertainty in the key variables. At the initial topic screening phase, resources for this process are limited in terms of both time and money. At the later stage it would be possible to insert a pretrial modelling phase into the R&D commissioning process where resources are put into the critical assessment. The value of doing this pretrial modelling would arise from changes to the trials and trial designs commissioned.

Evaluation against criteria

The criteria used for assessing the value of a methodology have been developed during the course of this review and were discussed in Chapter 2. They fall into two broad categories: theoretical soundness and practicality.

Time and cost

Undertaking modelling work to plan and prioritise research studies will involve significant resources, and this is one of the main factors that has been used to argue against the implementation of a systematic and explicit prioritisation process.³ At present, however, the evidence for the scale of resource required does not exist.

In the case of direct assessment of the cost-effectiveness of a trial, the question of who should undertake the work and bear the costs is also an issue.

One option is for the research commissioning body to undertake or commission specific pretrial modelling work; the Davies study⁸⁵ represents a pilot study of such a mechanism. Alternatively, the investigators submitting the proposed research could follow some template method to estimate the cost-effectiveness of the proposed study. Given the rare implementation of the approach in practice so far, this is unlikely to happen initially.

Delays in research

Time-scales are a major issue in getting research off the ground. The implications of the payback approach for delays are unclear and no evidence on this is presented in the literature.

Data availability

Some case studies have attempted to use a very limited amount of data in order to keep the process manageable and time limited. Others have

considered the development of extensive models particular to each different research topic. In either case, the availability of the data to populate the model is cited by the authors as a crucial factor in the practical feasibility of approach.

The main criticisms of the payback approach cited in the EURASSESS report³ relate to the availability of data, including the lack of agreement on the range of likely outcomes and implications of most technologies, and the difficulty in predicting the evolution in the use of a technology and the associated implications.

However, the payback approach fundamentally requires a health economic model for the technology of interest versus its comparators, and the growth in health technology appraisals across the world in the past 5 years clearly shows that, in most cases, model construction is possible. Most issues of data availability can be overcome by the use of subjective expert opinion and by appropriate and wide estimates of uncertainty. Although the use of expert opinion to populate decision-analytic models has been criticised generally, the three alternative approaches have severe limitations:

- Exclude elements of the model for which no data exist. This is completely unacceptable and exceedingly dangerous since potentially key structural effects may be excluded from the evaluation.
- Operate a two-level prioritisation process, where decision problems for which complete and good-quality evidence are available are evaluated explicitly, while decisions for which observed data are even partially unavailable are evaluated implicitly. This again is dangerous since it may result in modelling only being applied where there is relatively little uncertainty, which would obviate the key benefit of modelling work.
- Solve all decision problems without the use of modelling techniques, relying on the implicit judgements of the decision-makers.

The incorporation of expert opinion as a data source for modelling evaluations exposes the key uncertainties to debate, allows the impact of alternative assumptions to be assessed, and indicates where more evidence is required.⁹⁷

Timing of the use of the method

Henshall and co-workers³ point out that it is difficult to value a health technology assessment

topic without some idea of the assessment design. They are clear that prioritisation using this formal method can only be done once a clear description of the intended assessment exists. It is impossible to judge the costs of the trial needed at the stage of identifying problems of concern. The payback method clearly requires a description of the intended trial as its starting point.

Evidence of successful use

There is no conclusive evidence of the success of the payback approach to prioritising health technology assessments, as the technique has not been applied to the decision-making process. However, there is evidence of partial success in the production of coherent estimates of the cost-effectiveness of clinical trials.

Feasibility of achieving economies of scale by applying a generic method or model

It may be feasible to produce a generic method, which can be used on a large set of research studies that require prioritisation. Alternatively, a separate model may be required for each individual topic.

Acceptability to health technology assessment commissioners

Various authors have raised concerns over the presentation of subjectively derived data in a form that implies a greater degree of scientific authority, which could have an adverse impact on decision-makers,⁸⁸ as well as being perceived as mechanistic.⁸³

Acceptability to health technology assessment researchers

Modelling could be seen as an extra hurdle when bidding for research funds. However, it could also have benefits in helping investigators to decide on optimal design and anticipate the likely priority of their research through conducting some early modelling themselves.

Theoretical validity

The strength of the cost-benefit approach lies in the formulation of the impact of further research within a decision-analytic framework that estimates the expected net benefits of research. The theoretical validity of this framework is well established in the annals of welfare economics.

There is, however, a fundamental issue regarding the theoretical validity of the applied payback approaches, which separate the prioritisation process from the planning (trial design) process.

The starting point for the cost-benefit approach follows the specification of a proposed research project, a position that implicitly assumes the proposed research has been optimally designed. However, the optimality of a research design can only be proven following the comparison of the costs and benefits associated with, ideally, the full range of alternative trial designs.

The optimality of purely clinical trials, aimed at establishing a narrow definition of clinical effectiveness, may be approximated without recourse to direct assessments of their cost-effectiveness. However, if the ultimate purpose of health technology assessment is to inform commissioning based on the economics of alternative technologies, then optimal research design is significantly more complex. Cost-benefit analyses will be required for alternative combinations of sample size, length of follow-up and the range of variables to be assessed. In this latter case, if the objective of the prioritisation process is to optimise the allocation of research funds, it may be less likely that trials designed outside the prioritisation process will approximate an optimal design aimed at addressing all uncertainties.

In addition, a number of theoretical issues relating to the identified application of the payback approach can be questioned, including the assumptions made with respect to the uptake of a new technology and the methods used to handle uncertainty. The two main issues, however, are the implementation of a clinical effectiveness decision rule that drives the impact of the trial results, and the use of arbitrary probabilities to define the likelihood of the defined set of delta results.

The majority of the papers reviewed in this chapter^{4,51,85,89} assume that research priorities can be set on the basis of cost-effectiveness, but that the decisions that the trials aim to inform will be made on the basis of proven effectiveness alone. This assumption may be related to the age of the analyses, but in recent years the impact of health economics on the allocation of healthcare resources has increased substantially; Canada, Australia and the UK have founded institutions that include cost-effectiveness as a significant factor in the guidance issued. More significantly, the guidance issued in each of these countries is now mandatory. Indeed, the current status of economic evaluation appears to be greater with respect to informing clinical decisions than it does with respect to informing research decisions.

The use of arbitrary probabilities, such as Eddy's assignment of a 50:50 chance,⁴ to represent the likelihood of each of the specified delta results occurring within the trial also undermines the theoretical validity of the payback approach. The only circumstance in which the use of arbitrary probabilities can be justified is to demonstrate that the choice of probabilities has no impact on the cost-effectiveness of a proposed trial, which may apply in a limited number of cases. Otherwise, if the cost-effectiveness of a trial is sensitive to the probability of success, the use of arbitrary probabilities will invalidate the outcome of the payback analysis.

However, it should rarely be the case that analysts are completely uninformed about the likely outcome of a trial, as data are usually available from pretrial research phases, and a well-designed (frequentist or Bayesian) trial should be informed by sample size calculations that include estimates of the anticipated effectiveness of new technologies. Such data should always be used to inform the relevant parameter values within the payback approach.

Reliability

The acid test of reliability would involve applying the modelling procedure to the same study or set of studies and producing the same results in terms of recommendations for design or assessment of priorities. One important practical test of reliability is to see whether different modellers given the same potential problem would produce results of the same order. No evidence of this sort is available for the payback approach. However, differences in economic assessments undertaken by different authors have been well documented, and the key issues in good practice in undertaking and reporting economic modelling exercises are discussed in Chapter 4.

Empirical validity

No analysis of the empirical validity of any of the applied cost-benefit approaches was identified. Empirical validity can be established if the value of research predicted before the research is similar to the value implied by the actual outcomes of the research. This comparison is difficult in the context of the cost-benefit approach because the decision-analytic framework for both the pre- and postresearch estimation of the value of research changes from that represented in *Figure 8a* to that in *Figure 8b*. Namely, the preresearch analysis predicts the probability of alternative research outcomes, whereas the actual research only has one outcome.

Although the preresearch cost-benefit analysis is unlikely to be accurately validated, it is possible to compare a preresearch decision to implement research (made on the basis of positive expected incremental net benefits) with the estimated incremental net benefits associated with the conduct of the research. However, the decision-analytic framework for estimating the postresearch value of research also requires the prediction of the course of events in the absence of research, a task that has been found to be particularly difficult in the retrospectively applied analyses of the value of research (payback studies).

It should be noted that in so far as a model is an accurate reflection of current knowledge, if a trial results in an actual observation close to the model predicted outcome, then although this may empirically validate the model, the implied information content of the additional trial will be small. That is, it simply confirms what was already known. In contrast, a strongly divergent trial result should lead to a re-evaluation of current knowledge and would imply a greater information content of the trial.

The issue of empirical validation therefore is not a good indication of the value of a model. The key issue, once again, is that the model accurately reflects current understanding, as discussed in Chapter 4. Modelling is one of the few activities in life where it is good to be proved right, but potentially more useful to be proved wrong.

Value added, improved decisions

There is substantial discussion within the literature concerning the benefits to decision-making that a modelling approach could bring, such as structuring thinking and improving transparency,^{83,84} but these benefits are not specific to the direct assessment of the cost-effectiveness of a trial approach.

To date, there is no evidence that the cost-benefit analyses described in this chapter have been used to inform the research prioritisation process, which precludes the estimation of the impact of the approach on the decision-making process.

Retrospective applications of the payback approach, which estimate the likely parameter values that would have been specified before the commencement of the trial, in combination with estimates of the likely course of uptake in the absence of a trial, could give some indication of the potential value of the approach.

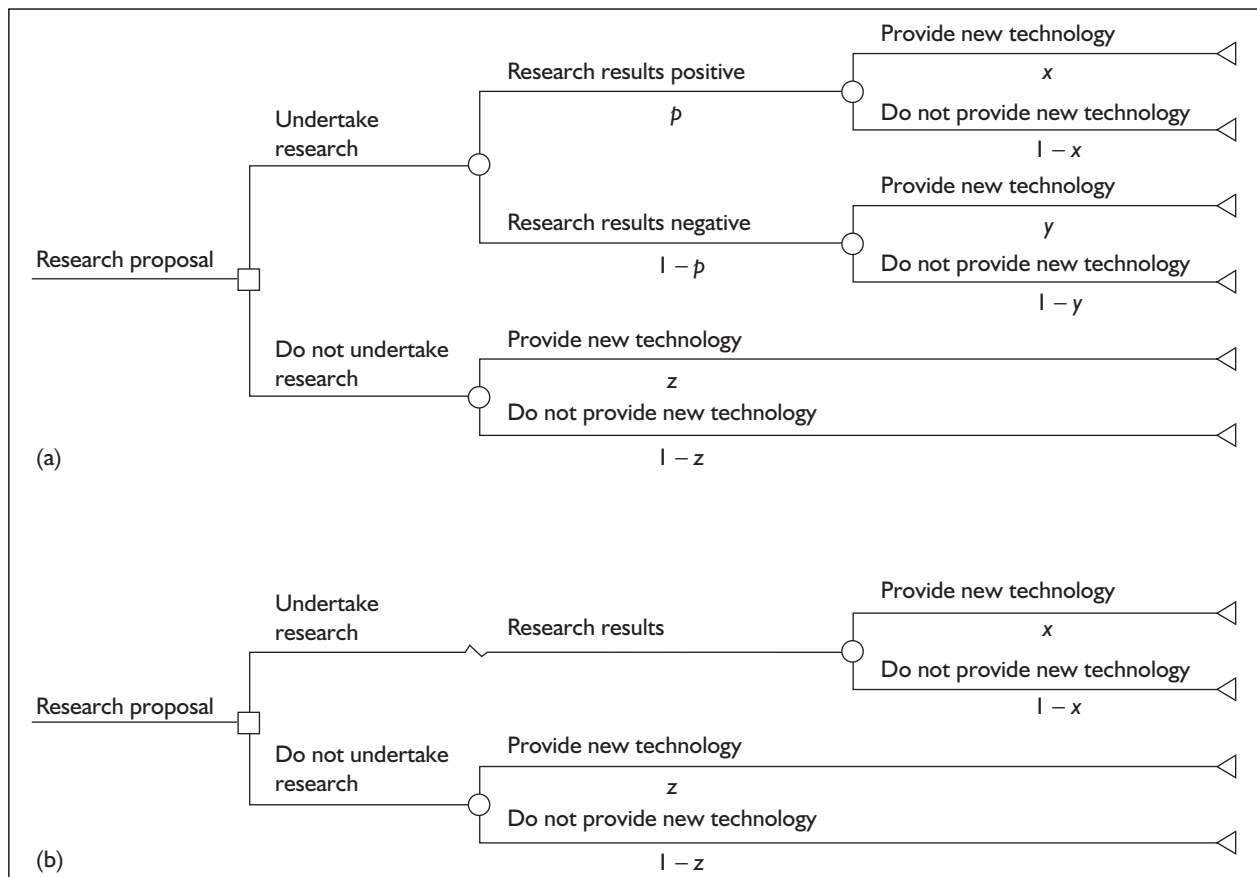


FIGURE 8 (a) Preresearch and (b) postresearch representation of research valuation framework

Conclusions and recommendations

A range of applied studies have been identified that present estimates of the payback, or costs and benefits of a specific trial. There is broad agreement on the basic approach, which involves listing the possible results of a particular technology assessment (i.e. the delta results) and then describing clearly the benefit and cost consequences of each of these results. By also making a prior estimate of the likelihood of each delta result, the expected value of a technology assessment, in terms of its net benefits, can be calculated.

In the literature concerning the cost-benefit approach there is no consensus on the appropriate methods for defining the possible outcomes of a trial, with suggested approaches including the two outcomes (effective/cost-effective or not), a range of discrete scenarios, or the specification of a probability distribution of clinical effectiveness results. There is also uncertainty regarding appropriate methods for identifying the probability of each delta result occurring, which is

often the most arbitrary component of the existing case studies.

The explicit definition of possible research outcomes, and their likelihood of occurrence, is a necessary component to any modelling approach to prioritisation. The specification of probability distributions around clinical effectiveness parameters moves towards the simultaneous description of the possible research outcomes and their likelihood of occurrence. An extension to this approach is to base the cost-benefit approach around a stochastic decision-analytic model with full parametric characterisation of all uncertainty in random values. In this way, the total uncertainty around the resource allocation decision could be accounted for.

A major shortcoming of the applied cost-benefit approach is the evaluation of a single specified research proposal, which implicitly assumes that the specified research is optimally designed. More informative applications of the cost-benefit approach should compare the expected net benefits of alternative research designs and sample sizes, thus integrating the research design and prioritisation processes.

The cost–benefit approach to research prioritisation applies a well-founded methodology (cost–benefit analysis) to the issue of research prioritisation. To date, the approach has mainly been applied using inadequate modelling methodologies, though the potential to incorporate improved sensitivity analysis techniques does exist. The benefits of such a state-of-the-art application of the cost–benefit approach would include the following.

- It is a natural extension of early economic evaluation and takes proper account of uncertain parameters.
- It is focused on the decision of the particular healthcare system (e.g. between two technologies in the NHS).

- It follows the same decision-making framework used in general modelling assessments of health technology assessments.
- It informs the optimal choice of research design required through the comparison of alternative forms of research.

As noted above, state-of-the-art modelling techniques have not been used within the cost–benefit framework. The primary requirement for further research into payback methods is the implementation of stochastic sensitivity analysis methods within exemplar case studies.

Chapter 7

The expected value of information approach

Introduction

The expected value of information (EVI) approach uses a decision-analytic framework to prioritise research by analysing the impact of existing uncertainty in parameters on the net benefit of alternative interventions.

Broadly, the EVI approach describes the costs of the current uncertainty regarding the provision of one intervention in terms of the probability that an alternative intervention should be provided, and the benefits that are foregone as a result of providing the 'wrong' intervention. The approach then estimates the impact of collecting more primary data to inform the resource allocation decision, and re-estimates the costs of uncertainty given the predicted sample data. The difference between the estimated costs of uncertainty is then compared with the relevant costs of collecting the stated amount of sample data in order to calculate the net benefits of prospective research.

The following sections in this chapter introduce the basic concepts and steps in the EVI approach in more detail, assess the approach against the criteria developed in Chapter 2, and report the conclusions and recommendations. Chapter 8 reviews case studies to demonstrate the key methodological issues and debates.

The concepts involved

This section provides a summary of the concepts underlying the value of information approach to sensitivity analysis, including:

- the decision variable: incremental net benefit (INB)
- uncertainty in the decision
- consequences of a wrong decision = opportunity cost
- the value of perfect information
- the value of sample information
- the expected net benefits of sampling.

The decision variable: incremental net benefit

The starting point of the analysis is the need to allocate resources between alternative health technologies. The approach can actually be used to compare any number of competing technologies but an illustration using two technologies is presented for simplicity. The decision between two technologies is based on the costs and health benefits of each.

Health benefits are measured using, for example, QALYs, but in order to value the net benefit of an intervention the health benefits need to be valued in monetary terms. The INB of treatment 1 versus treatment 0 is:

$$\text{INB} = \lambda (B_1 - B_0) - (C_1 - C_0)$$

where λ is the (societal) willingness to pay for a unit of health benefit (e.g. a QALY), B_i are the health benefits associated with each intervention, and C_i are the respective costs.

The following example is used to demonstrate the concepts presented in this section: if the societal value of a QALY is £30,000, then a treatment providing an additional 2 QALYs at an additional cost of £15,000 has incremental net benefit:

$$\text{INB} = (£30,000 \times 2) - £15,000 = £45,000$$

The adoption decision rule, that is, the rule by which we decide whether to adopt a technology into practice, is to fund a technology if it has a positive INB. It is important to note that the decision rule is based on the expected value of the INB, which indicates that a technology should be adopted regardless of the uncertainty around the INB as long as the expected or mean INB is greater than zero. That is, the baseline decision concerns the choice of technology given current available data.

Uncertainty in the decision

The data available to calculate the costs and benefits of alternative technologies will be

uncertain, and the uncertainty in parameter values may affect the adoption decision rule, that is, the INB may shift from positive to negative over the range of feasible values for one or more parameters, tested in isolation or combination. With this uncertainty comes the possibility that the baseline decision could be wrong. That is, if the true value of all parameters were known, the true INB would shift from being positive to negative or vice versa.

In the case of the example study, the expected additional QALYs provided by treatment 1 is 2. However, this may have a confidence interval of say 0.2 to 3.8 QALYs. If the true health benefits were just 0.2 QALYs, then the true INB would be:

$$\text{INB} = (£30,000 \times 0.2) - £15,000 = -£9000$$

and the baseline decision would be incorrect.

It is important to emphasise that the uncertainty being considered here is the uncertainty in mean parameter values. It is not the individual patient level variance that matters for the policy decision, it is the expected net benefit received by the relevant patient population.

Consequences of a wrong decision = opportunity cost

Investing resources in an intervention that does not maximise expected INBs means that potential health benefits have been foregone, and the value of that loss, namely, the number of units of health benefits multiplied by the willingness to pay for such benefits, is the opportunity cost of making the wrong decision.

The effect on the baseline decision (e.g. a shift in the INB from positive to negative) can be calculated for any possible combination of parameter values, alternative to the baseline set of parameter values. The opportunity cost of making a wrong baseline decision (i.e. the loss in INBs) can also be calculated. In the circumstances where the baseline decision is correct, there is no opportunity cost.

In the example study, if the true value of the QALYs gained by treatment 1 was only 0.2 QALYs, and the INB was negative, the baseline decision should have been treatment 0. The opportunity cost of not choosing treatment 0 is:

$$\begin{aligned} \text{Opportunity cost} &= (£30,000 \times -0.2) - (-£15,000) \\ &= £9000 \end{aligned}$$

The expected opportunity cost combines the likelihood of making the wrong baseline decision, that is the probability of each set of parameter values that indicate a shift in the baseline decision, with the associated opportunity costs of each erroneous baseline decision. In mathematical terms, this is the integral of the loss function over the possible outcomes.

The value of perfect information

Further data collection (research) is valuable, therefore, if it reduces the uncertainty in the baseline decision, and therefore reduces the likelihood of making the wrong decision and reduces the expected opportunity loss. The potential value of further research can be quantified by the associated reduction in the expected opportunity cost. Perfect information about all parameters in the decision problem would eliminate uncertainty altogether and reduce the expected opportunity cost to zero.

The per person expected value of perfect information (EVPI) is equal to the expected opportunity cost. The population expected value of perfect information (PEVPI) is estimated by multiplying the per person EVPI by the number of people who will receive the chosen intervention over the assumed time horizon for the intervention (i.e. the period over which the intervention will be offered to patients).

The EVPI for single parameters or sets of parameters within the decision problem is called a partial EVPI; for example, the EVPI can be estimated for further research on treatment effect, or utility weights for a set of health states, or the costs of a particular side-effect. It should be noted that the EVPI for a set of parameters is not additive, as there are likely to be joint effects within the model.⁹⁷

The value of sample information

The collection of perfect information, and the elimination of uncertainty, is an impossible target. The expected value of sample information (EVSI) predicts, and values, the reduction in uncertainty due to the collection of additional data from a sample of the relevant population:

$$\text{EVSI} = \text{EVPI}_{\text{current data}} - \text{EVPI}_{\text{current data} + \text{predicted sample data}}$$

The expected net benefits of sampling

Collecting further information takes time and money, and the value of sample information must be balanced against the costs of sampling.

Moreover, as sample size increases, the marginal value of additional information diminishes; for example, the reduction in uncertainty from the first 300 cases in a trial is greater than the value gained from moving from a sample size of 10,000 to 10,300. Assuming constant marginal costs of collecting sample information, the expected net benefits of sampling (ENBS)

$$\text{ENBS} = \text{EVSI} - \text{Cost of sampling}$$

will reach a maximum at some finite sample size. This is the optimal sample size for data collection on the parameter(s) concerned. The EVSI and the ENBS can be estimated for the overall decision problem or partially, for different parameter sets.

Summary of concepts

- **Uncertainty:** there is always uncertainty about the health and cost consequences of interventions; for example, the confidence interval for the risk reduction effect of a new drug for angina.
- **Adoption decision rule based on cost-effectiveness:** if the INB of a new technology versus current care is positive then it should be funded.
- **Incremental net benefit** = λ * (QALY difference) – cost difference, where λ = maximum acceptable incremental cost-effectiveness ratio (ICER).
- **Baseline decision:** the choice of treatment made when current data are used to estimate cost-effectiveness.
- **Opportunity cost:** the cost of making the wrong baseline decision: function of incremental net benefits foregone and the probability of making a wrong decision.
- **Value of (further) information:** further data collection (research) has value in reducing the uncertainty around the baseline decision, which is quantified using the reduction in opportunity cost.
- **Expected value of perfect information (EVPI):** the value of eliminating uncertainty around a parameter(s) or in the whole decision problem.
- **Expected value of sample information (EVSI):** the value of reducing uncertainty around a parameter(s) or in the whole decision problem through the collection of data from an additional finite sample.
- **Expected net benefit of sampling (ENBS):** EVSI minus the cost of sampling. For a health economic perspective the optimal sample size for further research is obtained when the ENBS is maximised.

Steps in the EVI method

The first step towards the estimation of the ENBS is to estimate the EVPI. The EVPI can be estimated for individual parameters, sets of parameters or the model as a whole, where the benefits associated with eliminating uncertainty around a parameter, a set of parameters or all the parameters within the model, respectively, are estimated.

The objective in estimating the EVPI for alternative sets of model parameters (including the full set) is to identify circumstances in which it may be more efficient to concentrate research efforts on the collection of a limited set of variables, rather than attempting to inform all possible variables.

Where the costs of further research will clearly exceed a low overall EVPI this would indicate that no further research is merited.

Decisions over the extent of the research required should only be made when the EVSI and the research costs associated with alternative sampling strategies have been incorporated. The specification of the alternative sampling strategies should be based on an informed assessment of the possible research studies that could be undertaken. *Table 8* describes possible sets of parameters and possible types of research study that could inform their values.

Building on details provided in the literature, the following sections describe full step-by-step processes for estimating the EVPI for the full set of model parameters, as well as methods of the estimation of the partial EVPI for individual or sets of parameters. Three alternative processes for the estimation of the partial EVPI are presented, along with a discussion regarding the correct method.

Method to calculate overall EVPI

1. Select the maximum acceptable ICER (λ) to be used or range to be analysed.

TABLE 8 Possible parameter sets for partial EVPI analyses

| Parameters | Possible research vehicle |
|------------------------|---------------------------|
| Full set | Complex RCT |
| Clinical effectiveness | Simple RCT |
| Natural history | Observational study |
| Costs | Observational study |
| Utility values | Benefit valuation study |

2. Identify all of the parameters within the model and assign probability distributions to each parameter. As discussed above, these distributions should capture the existing uncertainty in the mean value of each parameter, not the patient level variation.
3. Undertake a Monte Carlo simulation of the model sampling, for example, 10,000 sets of parameter values in the model.
4. Calculate the mean net benefits for each technology across all iterations, and identify the decision-maker's preferred baseline decision, that is, the strategy with the highest mean net benefits.
5. Record the optimal strategy, and calculate the opportunity cost, for each iteration as follows:

$$\text{Opportunity cost} = (\text{Net benefits for the optimal strategy}) - (\text{Net benefit for the baseline decision})$$

Note that when the baseline decision is the optimal strategy for an individual iteration, the opportunity cost is zero.
6. Calculate the mean value of the opportunity loss over all the iterations. This is equivalent to the expected cost of uncertainty and provides an estimate of the per patient EVPI.
7. The PEVPI for current and future patient populations is based on the incidence of presenting patients over the effective lifetime of the technology, including discounting.

Some authors express the above process analytically without using the Monte Carlo

method, instead using integration over the possible uncertain values of parameters. However, all agree on the underlying approach to the calculation of the EVPI, which is based on the original formulation of decision analysis.

Special case: estimation of EVPI if net benefits are normally distributed

The early health economics literature describing the estimation of EVPI concentrates on the special case when the uncertainty in the prior estimate of INBs is normally distributed.^{69,98} If the uncertainty in the INBs is normally distributed, then there is a simple mathematical formula for the overall expected value of information.

Claxton and Posnett⁹⁸ presented a hypothetical application to demonstrate the technique that defined net benefit in a QALY currency (i.e. QALY difference – Cost difference/ λ). An analogous example is used to illustrate the concepts and the formulae, but it is presented in the more common form of monetary net benefits. It is assumed that the prior estimate of INBs is normally distributed with a mean δ_0 of £1500 and a standard deviation σ_0 of £1000 (see *Figure 9*). The allocation decision will affect 100,000 patients over the relevant period.

The mean net benefit is positive and the baseline decision should be to adopt the technology in question. As shown in *Figure 9*, there is a chance that the true INB of the technology is below zero. If the INB was truly negative then the baseline decision would be wrong, and there would be

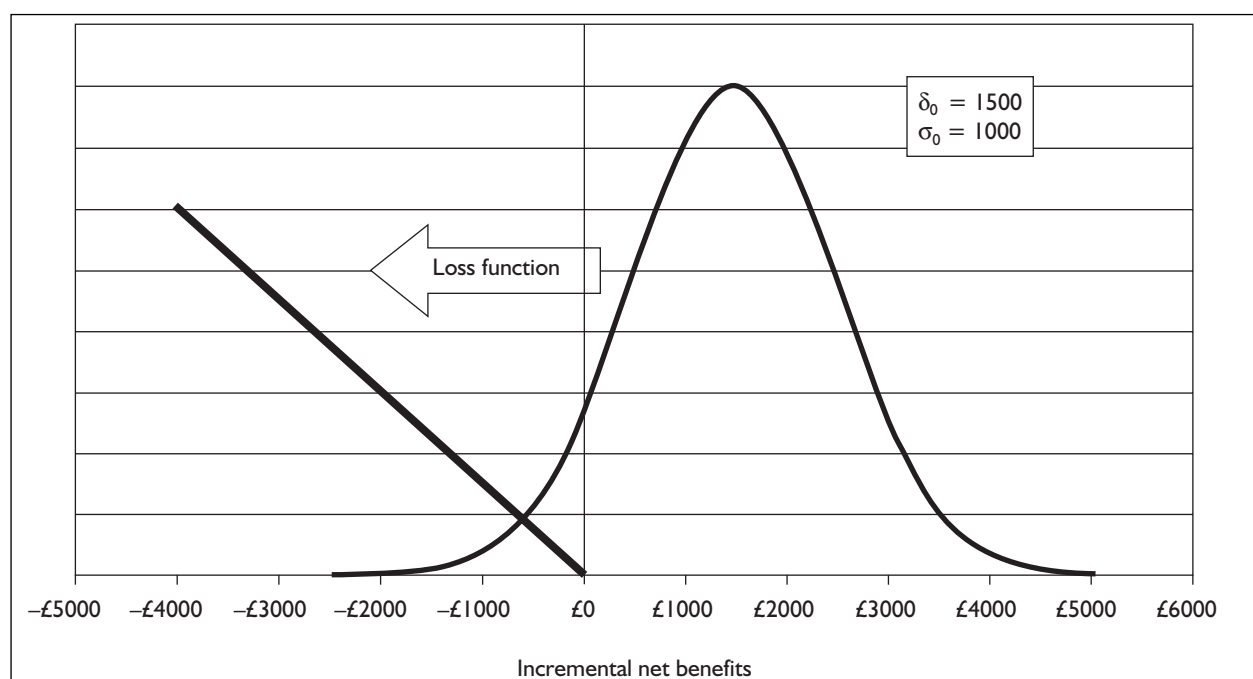


FIGURE 9 Representation of the distribution of INBs and the associated loss function

opportunity losses equal to the distance between the true INB and the break-even value (zero). This is illustrated in *Figure 9* with the line known as the loss function showing the scale of the opportunity loss for each possible value of the INB.

The expected opportunity loss is given by multiplying the value of the loss at each possible INB by the probability of that INB occurring. In mathematical terms, this is the integral of the loss function multiplied by the probability distribution for INB. Fortunately, a version of the integral of the loss function was tabulated over 50 years ago (just like log tables), which requires only the mean and standard deviation of the INBs in order to calculate the opportunity loss via a simple formula.

In general decision theory, just as the payoffs can be measured in non-monetary terms, so the break-even point in deciding between alternative strategies does not have to be zero. In the general case there can be a break-even value for the payoff that is denoted by δ_b in the formula. To utilise the tables for the integral of the normal loss function, we first have to calculate the standardised distance between the prior mean value (δ_0) and the break-even value of the incremental net benefit. This is defined as:

$$D_0 = |\delta_0 - \delta_b| / \sigma_0$$

In the general case, the formula for the expected opportunity loss (EVPI) is:

$$EVPI = |\text{slope}| * \sigma_0 * L(D_0)$$

where $|\text{slope}|$ is the monetary value of the unit of payoff (= 1 when payoff is defined in monetary

terms), σ_0 is the standard deviation of the prior distribution of the INBs, and $L(D_0)$ is the unit normal loss integral for the standardised distance D_0 . In the example, the slope is 1, the break-even value is 0 and σ_0 is 1000. D_0 is $(1500 - 0)/1000 = 1.5$, and $L(1.5)$ equals 0.02931, as defined in the unit normal loss integral tables. The per person EVPI for the illustrative example is:

$$\begin{aligned} EVPI &= |\text{slope}| * \sigma_0 * L(D_0) \\ &= 1 * 1000 * 0.02931 \\ &= \text{£}29.31 \end{aligned}$$

The population EVPI for 100,000 relevant patients is therefore £2.9 million. This represents a ceiling for the investment in further research. If it is possible to improve the certainty around the INB of the technologies concerned for less than £2.9 million, then such research may be worthwhile.

To illustrate the power of the formula, three further examples are shown in the *Figure 10*. The first has the same mean but wider uncertainty ($\delta_0 = 1500, \sigma_0 = 2000$). The second has a mean closer to the break-even point ($\delta_0 = 500, \sigma_0 = 1000$) and the third has a mean much farther away ($\delta_0 = 2500, \sigma_0 = 1000$).

The results show the key principles of the estimation of the EVPI. First, if the uncertainty is wider (curve 1 compared with the original example), then the value of further research is likely to be greater, EVPI is £262.40 per person compared with £29.31 in the original example. Secondly, if the prior mean INB is closer to the break-even point (curve 2 compared with curve 3),

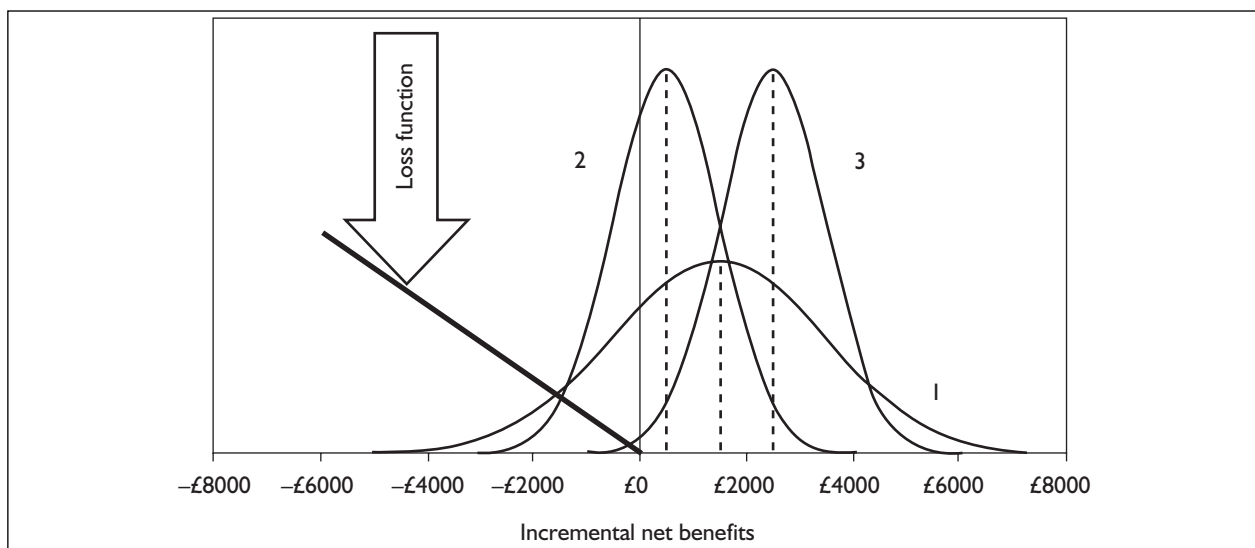


FIGURE 10 Further examples of the estimation of the EVPI using the unit normal loss integral

then the value of further research is likely to be greater, EVPI is £197.80 per person for curve 2, and £2.00 for curve 3.

Methods to calculate partial EVPI

Three alternative approaches to estimating the partial EVPI have been defined, which are described below.

Method 1: fix parameter values of interest at prior mean

The EVPI for a set of parameters less than the full set is estimated as:⁹⁹

$$\text{Partial EVPI} = \text{Overall EVPI} - \text{Opportunity cost with perfect information about the specified parameter set}$$

The opportunity cost associated with uncertainty existing only around the parameter values outside the specified set is estimated by repeating the process for estimating the overall EVPI, but holding the value(s) of the specified parameter set constant (at their prior mean) during the Monte Carlo simulations.

Specifically, the steps involved are as follows.

1. Calculate the overall EVPI as described in the previous section.
2. Repeat steps 1–7 (from the overall EVPI process), but hold the values of the specified parameter set at their prior mean value during the simulations.
3. The EVPI for the parameter set is the difference between the values estimated in steps 1 and 2 (Partial EVPI = Overall EVPI – Opportunity cost with perfect information about the specified parameter set).

Method 2: fix all other parameter values at prior mean

Felli and Hazen⁶⁹ describe a similar but inverted procedure for calculating the EVPI for parameters, whereby the values for the ‘non-specified’ parameter set are fixed at their prior mean values while allowing the specified parameter set to vary. Exactly the same steps as described for method 1 are then followed to estimate the partial EVPI.

Method 3: two-level sampling (the correct method)

The algebraic representation of the EVPI process provided by Felli and Hazen⁶⁹ implies two-level sampling, although the process is not described clearly in the narrative. The only identified studies describing the practical application of method 3

are in the case studies evaluating screening for cervical cancer,¹⁰⁰ kidney preservation systems¹⁰¹ and screening for inborn errors of metabolism.¹⁰²

Method 3 accounts for the fact that, in advance of further ‘perfect’ research, the true values of the specified parameter set are not known. The true value of the parameter could be discovered at any point within the current uncertain range, and finding the true parameters value at different points in the uncertain range affects the probability and scale of the opportunity cost. For instance, finding the true parameter value at an extreme point may make the baseline decision certain (no matter what the values of the other parameters in the model), or finding the true parameter value at a particular break-even point may mean that the baseline decision would become more uncertain and the opportunity cost may even increase.

In the general case, the procedure should estimate EVPI using the full range of potential values for the specified parameter set. This requires a two-level Monte Carlo simulation, where the steps set out above in method 1 are repeated for a sequence of different sampled set values for the parameter of interest.

1. Calculate the overall EVPI as described in the previous section.
2. Repeat steps 1–7 but this time hold the parameter (set) of interest constant during the simulations; revise the baseline decision if required and calculate the opportunity loss.
3. Sample a single value for each of the specified parameter set, then repeat steps 1–7 (from the overall EVPI process), holding the specified parameter set constant at their sampled values during the simulations.
4. Repeat step 2 for samples of the specified parameter set (the issue of the number of samples of the parameter of interest is discussed in the section on method debates later).
5. Calculate the mean opportunity cost across all the step 4 iterations.
6. The EVPI for the parameter set is the difference between the values estimated in steps 1 and 5 (Partial EVPI = Overall EVPI – Opportunity cost with perfect information about the specified parameter set).

Parameter EVPI in the normal case

The investigators could find no literature or textbook examples where partial EVPIs for subsets

of parameters had been established via an equivalent simple formula.

Comparison of methods for the estimation of partial EVPI

The key difference between methods 1 and 2 is that method 1 answers the question “how valuable would it be to know about the parameter of interest with certainty?”, while method 2 answers a slightly different question, “If we knew all of the other parameters were at their mean value, then how important would it be that we do not know about the parameter of interest?” These distinctions may appear semantic and researchers may expect the different methods to give the same answers. There is no published comparison, but investigations conducted during the course of this review show that the two methods result in different answers.

However, there is a conceptual problem with both of these simpler methods, as they both fix the values of the specified parameter set at their prior mean value, and the opportunity cost can vary according to the assumed true value of a parameter. The correct approach, therefore, is to take all the possible true values of the parameter, calculating the resulting remaining opportunity cost for each, and then calculating the mean opportunity cost across all possible true parameter values (i.e. method 3).

The two-level Monte Carlo simulation method is the correct generalised approach but the simpler one-level approaches provide much quicker calculation. There remains a need for case studies to resolve which of the two simpler methods produces a better approximation to the two-level approach. At this stage it can be hypothesised that when there is a linear relationship between input parameters and the INB, and potentially when the probability distributions representing the specified parameter set are not skewed, and when the prior mean value is close to the break-even value for the decision problem, then method 2 may provide a reasonable approximation.

Expected value of sample information

In the absence of an infinite sample, perfect information will be impossible to obtain. The value of sample information predicts, and values, the reduction in uncertainty from the collection of a specified sample of additional information.

The basis for the estimation of the EVSI is that the original probability distributions around the input parameters can be updated, reducing the variation

described by the prior distribution to reflect improved precision due to the collection of more data. The key assumptions in updating the probability distributions relate to the choice of data used to update the prior distributions. This aspect of the EVI process is a key element in ongoing research.

There are no step-by-step descriptions of processes to calculate the EVSI within the literature. Indeed, only three studies were identified that discussed the mathematics in relation to actual case studies.^{97,98,103} Thompson¹⁰³ uses an approach that is similar to the cost-benefit of a trial approach examined in Chapter 6, where the probabilities of different sizes of trials giving type I and type II errors are estimated. The case study characterises current uncertainty in ‘lives saved per electronic foetal monitor used’ with a uniform distribution, and incorporates a threshold value (in this case \$240,000 for a foetal life saved). A formula is developed to estimate the expected net benefit gained from the trial, which uses (unspecified) numerical methods to perform the integration.

The Claxton studies^{97,98} estimate EVSI under the assumption that the INBs are normally distributed. In addition to the requirement that the uncertainty in the INBs is normally distributed with a known standard deviation (σ_0), the estimation of the EVSI requires that the INBs at individual patient level are normally distributed with a known standard deviation (σ_1). The algebraic expression for the relationship between EVSI and n is taken from Raiffa’s⁵ seminal work on decision analysis, where:

$$\text{EVSI}(n) = |\text{slope}| * \sqrt{V_n} * \sigma_0 * L(D_n)$$

where $\sqrt{V_n} = \sqrt{\sigma_0^2 / (\sigma_0^2 + \sigma_1^2/n)}$, σ_1 = the standard deviation of the prior distribution of the INB at the individual patient level and $D_n = |\delta_0 - \delta_b| / (\sigma_0 \sqrt{V_n})$.

It can be seen from the formula that V_n is always less than 1, and that as n increases V_n approaches 1. As V_n is always less than 1, then D_n is always greater than D_0 , and $L(D_n)$ is always less than $L(D_0)$, and hence $\text{EVSI}(n)$ is always less than EVPI.

The V_n term requires an estimate of the patient-level variance in the INB. Patient-level uncertainty in treatment effect can be estimated from any existing early trial or case series data. In order to estimate the associated uncertainty around INBs, the uncertainty in patient-level costs and outcomes will also need to be factored in. This presents a

practical difficulty that is not discussed in the identified literature.

Claxton and co-workers⁹⁷ apply the method described by Claxton and Posnett⁹⁸ to an evaluation of donepezil, a drug for Alzheimer's disease. Most of the study relates to the estimation of the EVPI and, in particular, the EVPI for various model parameters.

The necessary assumption for the applied analyses of the EVSI, that the net benefits of research are normally distributed, has been shown to be flawed.¹⁰⁴ Monte Carlo methods for EVPI may be able to be adopted for sample information, although the key to implementing such an idea will be the methods used to estimate the expected posterior variance in the parameters of interest following additional sample information. Methodological work in this area is ongoing, focusing on the use of specialised software, such as WinBUGS, and the use of Bayesian distribution theory and the properties of conjugate families of probability distributions.

Expected net benefits of sampling

The ENBS is equal to the EVSI minus the trial costs. Claxton and Posnett⁹⁸ characterise the cost of a prospective trial of two treatments as the sum of a fixed cost plus a marginal cost per patient. As the sample increases from zero, additional value will be gained from the trial information. Beyond a certain trial size, however, the information derived from the trial will be subject to diminishing returns, while the marginal costs of research will continue to rise and so the ENBS will fall. The sample size at which the ENBS peaks is the optimal sample size.

Issues around the application of the EVI approach

A limited number of case studies were identified that applied various elements of the EVI approach. These case studies were examined in detail on their methodological approaches as well as for issues of the feasibility and the value of the EVI approach in practice. A detailed review of each individual case study is presented in Chapter 8. The content of the identified studies is presented in the context of the following issues around the implementation of the EVI approach:

- populating the EVI model
- estimating the size of the relevant patient population

- determining the number of simulations required
- selecting a decision threshold.

Populating the EVI model

A fundamental issue for the EVI approach concerns the collection of data to populate the model. The central estimate of the mean (population) values, and their plausible range (in the form of a prior probability distribution) for each model parameter are central to the calculation of the EVI. A wider range for a particular parameter signifies that there is more uncertainty and will usually produce a greater EVI, and hence a greater priority for the research to reduce the uncertainty.

The issues around the appropriateness of alternative data sources to inform the prior distributions are similar to those faced during the conduct of a standard modelling economic evaluation, where data describing the feasible range of parameters values should be collected to inform sensitivity analyses. Hence, both Phelps and Mushlin¹⁰⁵ and Fenwick and colleagues⁹⁹ describe the use of expert judgement and the available clinical literature to describe plausible ranges for each parameter.

The use of subjective expert opinion to populate decision-analytic models is also relevant to the payback approach and was discussed in Chapter 6.

More specific to the implementation of the EVI approach is the need to describe uncertainty in the form of a prior distribution, and the choice of different mathematical forms for the probability density function. Various forms of probability distribution have been used. The earliest work in this area, by Thompson,¹⁰³ used uniform distributions to describe uncertain clinical parameters (costs and utility values were not assessed). More recent applications of the full economic models have used simple triangular distributions to describe most parameters,¹⁰⁰ or specified alternative forms of distribution for different categories of model parameters. For example, Fenwick and co-workers⁹⁹ assigned log-normal distributions to unit costs, utility and event time parameters (as they are all bounded by zero and tend to have skewed distributions), and triangular distributions to probability parameters.

Other studies have provided more detailed descriptions of the rationales behind the choice of distributional form for different types of model inputs, which are summarised in *Table 9*.

TABLE 9 Rationales for choices of probability distributions presented in the literature

| Variable | Distributional form | Rationale |
|---|---------------------|---|
| Fenwick et al.⁹⁹ | | |
| Unit costs, utility | Log normal | Bounded by zero and tend to have skewed distributions values, event times |
| Probabilities | Triangular | – |
| Claxton et al.⁹⁷ | | |
| Baseline transition probabilities | Beta | Beta distribution takes values between 0 and 1, and the distribution parameters (alpha, beta) represent the number of 'successes' and 'failures', directly available from longitudinal database |
| Mean utility for each health state | Normal | Standard deviations based on the standard errors from cross-sectional study: assumption of normality is not unreasonable to use sample size ranges from 55 to 191 |
| Direct and indirect costs for each health state | Log normal | Standard deviations based on standard errors from observational cost data: cannot be negative and is positively skewed |
| Relative risk ratios | Log normal | Mean and confidence intervals based on Cox proportional hazards regression |
| Dropout or discontinuation rates | Log normal | Mean and standard deviations based on trial data and an open-label follow-up study |
| Additional utilisation and prices | Constant | Difference in utilisation and price across decision-makers/settings represents variability rather than second order uncertainty |
| Duration of drug effect | Log normal | Based on a survey of clinical experts about the expected efficacy duration |
| Karnon and Brown¹⁰⁶ | | |
| Proportions | Beta | Beta is bounded by 0 and 1, and provides the most realistic representation of proportions |
| Survival times | Gamma | Gamma is bounded by zero and approximates the normal distribution at large samples |
| Costs | Gamma | Same reasons as for survival times |
| Utility values | Beta | Utility values portray similar properties to a proportion (if not bounded by 0 and 1 a scale parameter can be fitted to the beta distribution) |

Clearly each model is different, but the rationales underpinning the choice of prior distributions provided by the above case studies provide a focus for discussion.

Estimating the relevant patient population

Both the payback approach and the EVI approach require the estimation of the relevant patient population that will be affected by the allocation decision (to be informed by the prospective research). There is a need to define the numbers of people per period (e.g. per year) who will be affected and the likely period before the intervention is superseded. Thompson¹⁰³ estimated the population to benefit from foetal monitoring research, assuming 3.5 million births per year, that the research would be applicable for a minimum of 4 years, with a discount rate of 7%,

and calculated that the study would affect 9.5 million low risk-births. This is exactly the same series of assumptions used in the TAPSS payback framework.⁴ Both Fenwick and co-workers⁹⁹ and Claxton and co-workers⁷⁹ defined the expected time horizon for the specified interventions without providing a reason for their choices.

In general, the identified EVI literature does not add any methodological rigour to the process of estimating the relevant patient population.

How many simulations?

Adopting the two-level simulation approach to the estimation of partial EVPI and EVSI involves simulating the data that might be obtained from further research and, for each of these possibilities, also simulating the possible values of remaining uncertain parameters.

There is no detailed discussion in the EVI literature concerning the number of simulations required for convergence to stable results. Fenwick *et al.*⁹⁹ and Payne *et al.*¹⁰⁰ used 1000 iterations, while Claxton *et al.*⁹⁷ used 10,000 iterations to establish the probability distribution for the incremental net benefit of treatment. There is a substantial literature on convergence in simulation modelling outside the health economic context. This has not been the subject of this review. In general the approach is to examine alternative numbers of simulations and compare the results, though there are many variations on this theme.

Within healthcare modelling, there are some individual patient level simulation models, which have examined the relationship between numbers of simulations and convergence. Karnon¹⁰⁷ assesses alternative simulation sizes in a model evaluating adjuvant therapies for early breast cancer. Holding input parameter values constant at their mean values, repeated replications of the model were undertaken using different random number seeds for replications 100, 1000 and 10,000 simulations. The cost-effectiveness plane results showed wide dispersion using 100 patients, but much tighter concentration with 10,000 simulations. While this is a different context to the simulation of second-order uncertainty in EVI calculations, an analogous approach would be of benefit.

In summary, measurement of the scale of result changes when more or fewer simulations are used is good simulation modelling practice. It would be useful if EVI studies showed some analysis of the convergence of the simulation results for different numbers of iterations, in order to establish evidence for the robustness and stability of the results.

Selecting a decision threshold

The QALY is the most common measure of health benefit against which a monetary threshold [the maximum amount that the decision-maker is willing to pay in order to gain an additional unit of the defined benefit – the maximum acceptable incremental cost-effective ratio (MAICER)] has been set to determine the EVI. A higher decision threshold values health benefits at a higher rate, but the relationship between the threshold and the EVI is not simple or linear because the threshold level also affects the choice of intervention given current information.

Two theoretical options are available for the specification of the appropriate threshold:

- Establish the societal willingness to pay for health benefits.
- Estimate the ICER of the least efficient intervention to be funded within a healthcare system with a fixed budget (excluding those interventions funded for reasons other than efficiency).

At present, there is no explicit consensus or policy defining the relevant MAICER for the NHS. The identified case studies have generally tested the impact of alternative levels for the threshold. Fenwick and co-workers⁹⁹ used three values: £5000, £10,000 and £20,000, while Claxton and colleagues⁹⁷ used a central estimate of US\$50,000 and conducted sensitivity analysis on values ranging from \$1000 to \$100,000.

The National Institute for Clinical Excellence (NICE) is beginning to produce more clarity on the acceptable level of cost per QALY for NHS funding, and some commentators have speculated that decisions made by NICE imply a decision threshold of £30,000 per QALY. As further evidence from the NICE decision-making process emerges, more clarity as to an appropriate threshold may emerge. It is important to realise in this respect that the purpose of this type of analysis is to support rather than replace commissioning decision-making. Thus, outcomes should be presented over a range of possible thresholds in order to allow other subjective factors, for instance equity, legal issues and perceived but unquantified biases in the evidence base, to be evaluated together with these quantified economic outcomes.

Evaluation against criteria

The criteria used for assessing the value of a methodology have been developed during the course of this review and were discussed in Chapter 2 on methods. They fall into two broad categories, theoretical soundness and practicality. The EVI approach is first reviewed in relation to issues of practicality; this is followed by an assessment of the approach's theoretical soundness.

Time and cost

As with the payback approach, the EVI approach to the prioritisation of research is more complex and requires more time and resources than an implicit process.⁸³ However, there is almost no discussion in the literature of the length of time needed to build, run and evaluate the results of an EVI model or

the approximate costs. The basis for the assessment of time and cost requirements for the EVI process is based, therefore, on the experience of the present authors' work on EVI and informal discussions with researchers in this area.

The development of a new disease treatment pathway model and analysis of uncertainties require anything from a week to several months, depending on the depth of evidence. The time taken to adapt a well-developed model to an EVI analysis is significantly lower than when starting from scratch, as the architecture to run the EVI simulations and record results is relatively generic and can be transformed from one model to the next. Within the authors' university department this has been implemented within Microsoft Excel[®] and VBA[®] (Visual Basic for Applications). Developing this took 3 or 4 weeks, but its adaptation to a new model takes 2 or 3 days. Updating an existing EVI model, as new evidence emerges from clinical trials or other forms of research, simply requires the updating of the prior probability distributions. The time required to run the model and evaluate the results is mainly determined by the number of parameter sets for which the EVI is to be estimated and the time required to generate a single estimate. In the authors' experience, even in relatively simple models this can range from 1 day to 2–3 weeks.

In the case of EVI methods, just as for payback methods, the question of who should undertake the work and bear the costs is also an issue. It is possible that the primary research commissioning body will do the work, since they will want to ensure that a fair comparison is made between potential research studies. The alternative is for investigators submitting proposed research to follow some template method to examine the EVPI (and even EVSI) of the proposed study.

Delays in research

While there are many case studies of secondary systematic reviews being commissioned that specifically request modelling work, there are only three that have been undertaken in a rapid time-frame and include EVI analyses. These are the review of liquid cytology screening for cervical cancer undertaken for NICE,¹⁰⁰ secondly, included as a pilot analysis, the review of screening for inborn errors of metabolism,¹⁰² and finally an evaluation of kidney preservation systems.¹⁰¹ This last is of specific interest here as the rapid review was initiated at the start of the NCCHTA process for commissioning primary research. Thus, the rapid review with EVI analysis was delivered in

time to be used to support the review of detailed proposals for primary research.

Data availability

The EVI approach requires very similar data to those required for the payback approach (i.e. costs, outcomes and effectiveness data, plus information to inform the estimation of the relevant patient population), so the previous discussion regarding data availability for payback (Chapter 6) is also relevant to the EVI approach. Namely, a wide range of health economic models, covering a broad spectrum of disease areas, have been populated and evaluated, and the alternatives to the use of weak data with appropriate high levels of uncertainty to populate decision models have severe limitations.

Timing of the use of the method

Unlike the direct assessment of the cost-effectiveness of a trial method, which clearly requires a description of the intended trial as its starting point, the EVI approach can be used without a definite trial design in mind as long as the relevant patient population and choice of comparators are clear. The EVPI and EVSI can be estimated without a particular research design in mind, although the estimation of the ENBS requires the specification of potential research vehicles.

Evidence of successful use

There is case study evidence of success in the use of EVI approaches. At this time, no research funding body has undertaken a comparative assessment of the use of the EVI approach as opposed to existing methods of prioritisation across a range of research projects and examined its potential impact on R&D decision-making.

Feasibility of achieving economies of scale by applying a generic method or model

The EVI approach is based on the development of a health economic model specific to the disease and treatments involved. A separate model is needed for each topic area. Such a model (e.g. for treatments of Alzheimer's disease or rheumatoid arthritis or lowering cholesterol) can certainly be used over and over again for different treatment options, as new evidence emerges, and for different proposed trials or research studies. The analytical model architecture to undertake the EVPI and EVSI analysis is relatively generic and it would be feasible to develop generic analytical enquiries for use with a limited number of modelling platforms. These could be bolted on to new treatment areas relatively easily.

Acceptability to health technology assessment commissioners

As noted with respect to the payback approach, the use of analytical methods can be perceived as mechanistic, while the use of imprecise data may be of concern.⁸³ There is likely to be some scepticism about the potential for modelling in prioritising research. A process of pilot evaluations of the utility of modelling rather than wholesale adoption is most likely.

Acceptability to health technology assessment researchers

Again, as with the payback approach, the need to estimate the EVI of a research proposal could be seen as an extra hurdle when bidding for research funds. However, it could also have benefits in helping investigators to decide on optimal design and anticipate the likely priority of their research.

Theoretical validity

The estimation of the EVI is grounded in Bayesian statistical decision theory, and has strong theoretical foundations.

The EVI approach to assessing the importance of current uncertainty (i.e. as a form of sensitivity analysis) has been commended owing to the combined assessment of the likelihood of alternative outcomes and their effects on the health benefits experienced by the relevant population. The presentation of remaining uncertainty in terms of the marginal value of perfect information is thought to provide a more intuitive approach than other forms of sensitivity analysis.⁶⁹

An important aspect of the application of the EVI approach to the full prioritisation process is that the explicit starting point is the decision problem, not a specified piece of research. On the basis of a complete analysis, that is, estimating an adequate set of partial EVIs, the approach leads the analyst to the optimum research design, including the optimal sample size.

There are two broad approaches to the EVI process. The approach that requires an assumption of normally distributed net benefits enables the analytical solution of the EVI, although no definitive approach to estimating the individual patient level variance (σ_1^2) in the INBs has been proposed. Moreover, the assumption that the net benefits of research are normally distributed has been shown to be potentially flawed.¹⁰⁴

The application of the EVI approach based on the use of stochastic decision-analytic models provides a theoretically sound representation of existing uncertainty around the decision problem. The applied analyses are based on an assumed adoption rule that states that the technology shown to generate the highest expected net benefits will be provided to the full population (given both current information and following further research). Such an adoption rule may be deemed unrealistic, although alternative objective functions (to the maximisation of health benefits) and different uptake rates (to the instantaneous uptake rate) could be incorporated into the EVI approach. The applied EVI studies do not, therefore, contribute to methods for establishing the uptake rate for technologies.

The principal area of methodological uncertainty around the application of the stochastic EVI approach concerns the estimation of the EVSI; in particular, the assignment of updated probability distributions to represent the data obtained through further research. No definitive approach to updating probability distributions, in the context of the EVI process, has been specified, although approaches based on Bayesian statistical methodology have been proposed. These approaches have the potential to provide a sound theoretical basis and standardised approach to solving the most difficult practical aspect of any prioritisation process.

Reliability

The acid test of reliability would involve applying a modelling process to the same study or set of studies and producing the same results in terms of recommendations for design or assessment of priorities. One important practical method of testing the reliability of a broad approach is to see whether different modellers given the same potential problem would produce results of the same order. No such approaches have yet been published in the literature. This again, however, refers to the reliability of the underlying disease model rather than the EVI analysis which, if implemented correctly, would be reproducible.

Empirical validity

No analysis of the empirical validity of any of the applied cost-benefit approaches was identified. An accurate analysis of the empirical validity of the EVI approach would compare the value of research predicted before the research with the value implied by the actual outcomes of the research. This comparison is facilitated by the EVI

approach (but is not possible using the cost–benefit approach), because the decision-analytic framework remains the same for both the pre- and postresearch estimation of the value of research; that is, the predicted reduction in the costs of uncertainty can be directly compared with the actual reduction in the costs of uncertainty.

However, it should be noted that the EVI approach is based on estimates of the current level of uncertainty around the allocation decision (including the estimation of the EVSI), and that additional research will not validate the estimated value of uncertainty at the time of the decision to undertake the research.

Value added using EVI

One of the main advantages of the EVPI approach is that it quantifies the maximum potential value of research on the subject concerned. This is done without the need for a specific design for the research. The EVPI simply quantifies the value of eliminating uncertainty from the decision-making process, which sets an upper limit on the value of research. If the research costs more than the EVPI then any further analysis is not required.

Conclusions and recommendations

The EVI approach is based on the economic evaluation of costs and benefits of health technologies in the form of a cost-effectiveness model. It requires the development of an economic model comparing the intervention of interest with its relevant comparators. The variables within the model must be assigned probability distributions to describe their uncertainty. This is followed by an analysis of the expected incremental net benefit of the interventions [$\lambda * (\text{QALY difference}) - \text{cost difference}$], and selection of the best to make a baseline decision. However, because there is uncertainty in the variables, the method also analyses the probability that the intervention not selected (given current information) could actually be the best. The benefits lost by not selecting the true best intervention given current uncertainties are calculated and known as the expected opportunity loss. Given expected numbers of people in the system and the likely life span of the interventions, the monetary value of perfect information is calculated. This is measured by reduction in expected opportunity loss if we had absolute certainty about the value(s) of the parameter(s) concerned. The method can estimate

the value of further research overall and on individual parameters.

A further extension is the calculation of the EVSI. There are fewer descriptions of the steps involved in calculating the EVSI within the literature and there is no consensus on the methods for implementation. The value of a trial of a specific size can be compared against its costs and a trade-off between sample size and increasing costs of research can be optimised. The estimation of the expected net benefit of sampling (ENBS = EVSI – Cost of sampling) leads to the specification of an optimal sample size for an optimally designed research project.

The EVI approach to research design and prioritisation is a systematic and well-founded methodology based on longstanding decision theory. The benefits of the EVI approach include:

- It is a natural extension of early economic evaluation and takes proper account of uncertain parameters.
- It is focused on the decision of the particular healthcare system (e.g. between two technologies in the NHS).
- It quantifies the value of perfect knowledge concerning the decision overall and thus provides an upper ceiling estimate of the value of research. If any proposed research would cost more than the overall expected value of perfect information (EVPI) then it cannot be deemed cost-effective. In particular, this allows comparison across disease areas.
- It quantifies the value of seeking information on specific parameters. It allows uncertain variables to be ranked as to the importance of their effect on the decision about the technology and quantifies the maximum potential value of research on the parameters concerned.
- It supports decisions on the form of research design required. EVI confirms the need for RCTs where the clinical efficacy of proposed new technologies is the key uncertainty, and suggests other forms of study where quality of life, utility, cost consequences, and so on, are most important in influencing the decision.

However, significant areas of uncertainty remain around the application of the EVI approach. The following sections describe recommendations for further research relating to different aspects of the EVI approach to research design and prioritisation.

General methods and assumptions

The basic steps in the process of calculating the overall expected value of perfect information are agreed upon by all the identified studies. Some express it analytically without using the Monte Carlo method (using integration over the possible uncertain values of parameters). However, all agree on the essence of calculating the overall EVPI, which itself is based on the original formulation of decision analysis.

There are some methodological differences in approach in the literature, concerning the precise implementation of the EVPI calculations. There are two simple versions of the method. The first calculates the EVPI for a parameter by holding the value of the parameter of interest constant (at its prior mean) during the simulations. The second suggests a similar but inverted procedure, which leaves all of the other parameters set at their prior mean values while allowing the parameter of interest to vary. There are conceptual problems with both of these simple methods.

This review has led to the development of recommended methods for the calculation of EVPI. This involves estimating the EVPI using the full range of potential values for the parameter of interest, using a two-level Monte Carlo simulation. This two-level method samples from the prior distribution for the mean of the parameter of interest, then holds that sampled value constant during a second level for a set of iterations where the other uncertain parameters are varied. The average opportunity loss over the two levels of simulations is calculated and the EVPI for the single parameter is then calculated as the overall EVPI minus the opportunity loss for the parameter of interest.

None of the papers has examined the differences in the resulting answers between the simplified methods and the more complete (and correct) approach in a case study.

A potential drawback of the methodologically correct procedure is the practical issue of computer processing time required, since this expands exponentially with the number of parameters. In complex models that require significant time to produce a single sample result, the final processing time can be prohibitive. Further research is required in the development of approximation methods, for instance meta-modelling, or efficient sampling algorithms to enable general application of EVI methods.

If the uncertainty in the INB is normally distributed, then the overall EVI can be calculated using a simple mathematical formula, which obviates the need for the Monte Carlo simulation approach. This is of some use but it is limited because there is no equivalent formula for the EVPI for specific parameters. There remains a question as to how often the uncertainty in real case studies can be adequately represented by a normal distribution, and this should be the subject of some further research.

There are no step-by-step descriptions within the literature concerning the process to calculate the EVSI. There are discussions of the analytical formula to calculate an overall EVSI in the case where the incremental net benefit has a normal distribution. The formula also requires the assumption that the patient-level uncertainty in INB can be represented by a normal distribution with a known variance. Again, this is of value in giving an insight into the potential value of a trial of a specific size. However, crucially there is no equivalent analytical formula for calculating the EVSI for specific parameters.

It should be possible to adapt the two-level simulation approach suggested for parameter EVPI to parameter EVSI. No work of this kind has yet been published and it should be a priority for methodological development.

Populating the model

This issue is not specific to the EVI methodology; it is an important concern in the validity of general disease treatment pathway models, but is of particular importance for the application of EVI analyses. The parameter estimates, both the central estimate and the plausible range, are the fundamental drivers for the calculation of the EVI. A wider range for a particular parameter signifies more uncertainty and potentially a greater expected value of information. One of the main issues therefore is the question “how should you estimate the plausible range for a specific parameter?” or, in Bayesian parlance, “How should you estimate the prior distribution for a parameter of interest?” Recent case studies discuss the process in depth and a body of evidence is emerging.

Different forms of probability distributions are commonly used for different types of variables within the model, for example, beta distributions for probabilities of events and log-normal distributions for costs. This is an emerging science, however, and further work to develop

methods in this area is likely as more analysts undertake the EVI approach.

In some models it will be necessary to obtain subjective judgements on the values of parameters and the uncertainty around them. The methods for obtaining subjective prior probability distributions are not generally discussed in the literature reviewed in this report. However, there is a body of literature on this general topic. Among others, the HTA report by Spiegelhalter and colleagues¹⁰⁸ covers some of these issues in detail. There is a need for researchers to examine the learning from other application areas.

Estimating the population to benefit from research results

An important point at which the EVI approach meets the payback approach is in calculating the population that will benefit from the research. There is a need to understand both the numbers of people who will benefit and the likely lifetime before the intervention is superseded. In the EVI approach, this is used to convert the per person expected value of research to a population EVPI. The EVI literature examined does not bring any additional methodological rigour to understanding these issues beyond the discussions from the payback work.

The objective function and selecting a decision threshold

The developments of this methodology to date have used the INB between alternative treatments as the objective function for decision-making. Furthermore, case studies have all focused purely on health service costs and benefits. Research is required on defining an objective function, including its scope and perspective, that reflects

decision-makers' concerns and is acceptable to key stakeholders.

Approaches to selecting a decision threshold vary, but most authors have used a generic value such as a cost per QALY as the currency. The decision threshold can be disease specific and the EVI approach will then allow uncertainties within that disease area to be ranked and key parameters to be identified. Given no explicit consensus regarding the decision-makers' threshold, the analysis is often undertaken for a range of distinct threshold values. It should be noted that as λ increases then the weight given to the uncertainty in the health benefit differences also increases. However, the relationship between λ and the EVPI is not simple or linear because the value of λ also affects which decision is taken given current information.

An area of opportunity

The methodologies and theory to utilise the value of information approach have existed for many years. There has been exponential growth in health economic analysis alongside the emergence of health economics as a key criterion in NHS funding policy. Recent years have seen an increase in computing power, health economic and modelling expertise, and the development of a body of case study material. Thus, there is a substantial number of existing economic evaluations for a wide range of disease areas that could be extended to incorporate new interventions and to support research prioritisation. In general, almost all economic evaluations can utilise the EVI approach. All of these factors together provide the context for a new and more systematic approach to research design and prioritisation.

Chapter 8

Expected value of information case studies

Introduction

This chapter describes and analyses the identified studies that applied or discussed some aspect of the EVI process. This work is intended to provide a background to the EVI process, and issues around that process, that were described in the preceding chapter. The studies are presented in an order that represents the methodological development of the EVI process in the healthcare field, which follows a broadly chronological order.

Decision-analytic determination of study size: the case of electronic foetal monitoring¹⁰³

Methods

This paper describes the use of decision analysis to determine study size for an RCT of electronic foetal monitoring (EFM), combining elements of the EVPI and payback approaches. Thompson¹⁰³ reports that there is considerable disparity among experts concerning the advisability of EFM. Four RCTs have been undertaken, covering a total of 2027 mothers, but only 500 mothers were in the low-risk group. The primary outcome measure for decision-makers is perinatal mortality, although brain damage is also an important outcome.

In relation to sample size calculations, Thompson dismisses haphazard (or pragmatic) methods, where sample sizes are based on factors such as resources remaining, or the number of potential

patients accessible by a possible principal investigator over the duration of the experiment. He cites an application of the classical methodology for setting sample size based on statistical significance, which estimated the appropriate sample size would be 63,000 low-risk mothers assigned to each of the two experimental groups.

Thompson¹⁰³ criticises the arbitrariness of the classical approach in specifying a minimum clinical significant difference, and the appropriate levels of statistical significance and power for the experiment, as well as the lack of account of the costs of the study. Instead, he recommends the use of decision analysis to compare the net expected benefits of different sized RCTs with the expected results if no trial takes place.

In line with the payback approach, two delta results for a potential trial are identified. The first is the case where EFM produces positive net benefits, and hence more births would be monitored in the future and improvements in perinatal mortality will ensue. Alternatively, EFM produces negative net benefits leading to reduced monitoring and costs. The net expected gain is calculated by combining the possible gains and losses with their probabilities of occurring, as well as accounting for the probability of the trial providing a false result. In essence, Thompson¹⁰³ formulates a decision tree that compares the options 'undertake a trial' and 'do not undertake a trial', as represented in *Figure 11*.

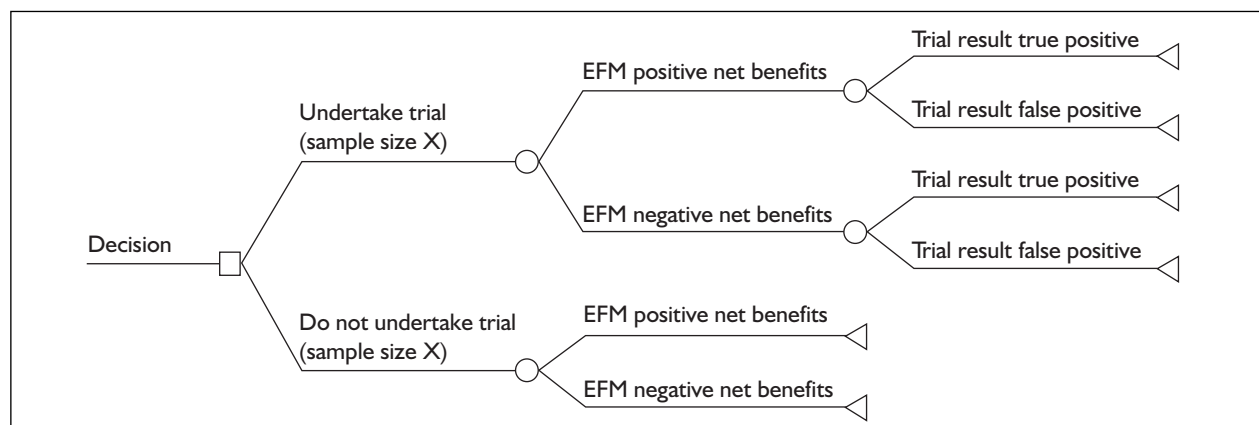


FIGURE 11 Representation of decision problem specified by Thompson¹⁰³

The most important assumptions in the model relate to the current estimates of life saved and mortality, and their possible range. Interviews with obstetricians showed that there was a range of prior estimate of the efficacy of monitoring from 0 to 1 in 2000 lives saved. There is also some uncertainty in the expected mortality in unmonitored low-risk births, ranging from 1 in 1300 to 1 in 700.

Thompson characterises both of these uncertain variables as a uniform distribution from the minimum to the maximum. The possible mortality rate among the monitored births is the difference between these two and could be positive or negative.

Thompson¹⁰³ sets out the population that stands to benefit from the study as 3.5 million births per year, assuming with the development of technology that the results of this study would be applicable for a minimum of 4 years. Using a discount rate of 7% it is estimated that the study has the power to affect 9.5 million low-risk births. He also assumes that current practice monitors around half of these births.

In order to balance the potential benefits of the trial with the costs of the trial size it is necessary to define a threshold level for the decision. This is the equivalent of g .⁹⁸ Thompson uses some data on the present value of average future earnings for a new-born baby and sets out a threshold level of US\$240,000 per perinatal life saved, which was estimated to equate to a break-even point for EFM as being one life saved per 4000 births monitored. The model also assumes fixed and variable (i.e. depending on patient numbers) study costs.

The optimal study size is that which yields the maximal positive net EVI. If the net EVI is negative, then society should not mount the study.

Results

Owing to the relative simplicity of the underlying decision model, Thompson¹⁰³ calculates the EVI analytically (in the appendix to his paper) rather than requiring a more general algorithm.

The EVPI would be US\$142.5 million, which provides a basic threshold that the study must be cheaper than US\$142.5 million in order for it to have any chance of producing net benefits. The analysis of sample information estimated the net benefits associated with alternative sample sizes, which indicated that the optimal study size is roughly 180,000 births in each of the EFM and control groups; a total of 360,000 births. While this study would cost US\$22 million, its net benefit would be US\$95 million.

However, the analysis shows that the estimated net benefits are not very sensitive to different sample sizes, as a sample of 100,000 per group would generate expected net benefits of US\$90.7 million, which is 95% of the maximal net EVI of US\$95.8 million.

Thompson¹⁰³ also undertook sensitivity analyses on a series of important variables within the decision tree, including study acceptance and take-up rate by obstetricians, costs of the study, value per life saved and the variance in the range of uncertainty. The results show that the optimum sample size varies from around 110,000 to 270,000 depending on the assumptions used (still orders of magnitude higher than 504). The net EVI calculation remains positive under all scenarios examined. The variable with the most effect on net EVI is the value of a saved life to society. If this is altered by $\pm 25\%$ then the net EVI varies from US\$61.9 million to US\$9.1 million.

The EVI is also calculated for an initial trial with a smaller sample size (100,000 per group), with the option of extending the trial (including an additional 80,000 per group) if the initial results are not statistically significant at a 95% confidence level. The expected additional gain in net benefits from following this strategy is roughly US\$1 million.

Arguments for the adoption of two or more stopping points include the possible savings (US\$8 million if the trial is not extended), as well as the ethical imperative of stopping the study as soon as it is clear that one form of treatment is significantly more beneficial than the other. Alternatively, there are practical difficulties in arranging for multiple stopping points. This work on stopping rules also relates to the work by Senn¹⁰⁹ and Grabowski¹¹⁰ in the commercial setting.

Thompson¹⁰³ also wonders whether net benefit is an appropriate decision criterion, or whether there should be a higher ratio; that is, research undertaken should have a much higher benefit-to-cost ratio.

Discussion

Thompson's paper¹⁰³ represents a synthesis of both EVI and the cost-benefit school. The model is concerned with optimal sample size once it has been decided that research on a specific variable is needed. In this case, the variable is the net mortality difference between low-risk babies who are monitored by EFM and those who are

unmonitored. The decision (that this variable is important) is made by calculating its EVPI, assuming all other variables are known. The EVI of further research on this variable is estimated for different possible sample sizes. Because the assumptions about the uncertainty in the parameters are simple (uniform distributions) the EVSI can be calculated analytically.

Although over 20 years old, this paper represents a simple case study illustration of almost all aspects of modelling for prioritising research, defining the process of prioritising research as a step-by-step process:

1. Model the disease and the possible intervention options using uncertainty analysis and expected value of perfect information.
2. This helps to define both the overall value of uncertainty in this area to society and the key uncertain parameters (given their value of information).
3. Consider designs for addressing the uncertainty in terms of primary research.
4. In order to do this, make assumptions about the delta results, the probabilities of those delta results, the uptake rate and the length of application of the study's results to practice.
5. Model the key aspects of study design as variables in a decision-analytic model, calculating out the EVI for different study designs.

Focusing technology assessment using medical decision theory¹⁰⁵

Methods

Phelps and Mushlin¹⁰⁵ compare the provision of a diagnostic test with two fallback positions:

- diagnostic test
- treat without testing
- do not treat without testing.

A decision-analytic framework incorporating data describing the sensitivity and specificity of the diagnostic technology, and costs and utility values associated with treatment and no treatment in both the sick and healthy population, demonstrates that patients with either a very low or very high prior probability of disease may not need the diagnostic test.

The estimation of the expected value of clinical information (EVCI), the EVPI and the expected value of imperfect information (EVImpI) are all

illustrated using a theoretical example. The EVCI is the net benefits associated with using the diagnostic test compared with allocating treatment on the basis of a patient's prior probability of the disease.

The expected value of a perfect diagnostic test, assuming perfect sensitivity and specificity for the test, is estimated for the full range of the prior probability of disease. Phelps and Mushlin¹⁰⁵ refer to the EVPI representing the first hurdle for the use of a diagnostic test, which is overcome if the EVPI is greater than the costs of the test (the costs of research are not mentioned).

The EVImpI estimates the EVCI, over the full range of prior probabilities of disease, for a variety of combinations of sensitivity and specificity of the test. The possible diagnostic capabilities of the test can be specified on the basis of a receiver operating characteristics (ROC) curve. If the anticipated combination of sensitivity and specificity results in an EVCI that is lower than the test costs, then the proposed test fails the second hurdle.

Phelps and Mushlin¹⁰⁵ then discuss the evaluation of a new diagnostic technology compared with an existing test. Assuming that a sufficiently precise ROC curve can be derived for the existing technology, and that the costs of the old and new technology are established, then a challenge region for the new technology can be created. The challenge region accounts for the differential cost of the two tests and describes the necessary performance of the new test for it to be considered cost-effective.

Results

A diagrammatic representation shows that the value of the EVPI peaks at the point at which uncertainty regarding the appropriate fallback positioning the absence of the test is greatest. The representation of the EVCI for a particular combination of sensitivity and specificity shows the EVCI peaking at the same prior probability of disease as the EVPI. It is stated that the EVPI can never be negative, although it is possible for the EVCI to be negative.

The theoretical derivation of the challenge region is described, and a diagrammatic representation is presented.

Discussion

In discussing potential forms of primary research, Phelps and Mushlin¹⁰⁵ recognise that RCTs

eliminate biased estimates of a medical technology's effects, but state that with most diagnostic interventions contamination is not a problem as each patient can receive all diagnostic interventions. While some invasive diagnostic interventions could contaminate findings, most technology comparisons involve only non-invasive methods.

To incorporate more meaningful measures of benefit than diagnostic accuracy, an RCT that follows patients beyond treatment (as indicated by a test) may be implemented. However, the emergence of new therapies will invalidate the results of such trials, while the decision-analytic approach can adjust to therapeutic changes, as well as alterations in service provision or even therapeutic misadventures.

Phelps and Mushlin¹⁰⁵ compare cost-effectiveness analysis with other possible methods of assessing diagnostic technologies, such as ROC curves, RCTs, group judgement or Delphi techniques, meta-analysis and case studies, and conclude that a decision-analytic framework extends the scope of analysis considerably because it incorporates patient outcomes and their values, and should systematically incorporate costs of therapy and testing. They further point out that undertaking the wider modelling of health gain, utility and cost will enable an understanding of crucial information that influences the allocation decision.

This paper acts as a forerunner of the later work (see following papers), providing a definition of the EVI in a healthcare context, and describing the use of decision analysis to support decisions regarding the use of a technology and whether further research on uncertain variables is potentially worthwhile.

However, the EVPI is not analogous to the EVSI (e.g. Claxton and Posnett⁹⁸), which estimates the value of further research on the basis of the predicted outcomes of that research. The EVImpI describes the cost-effectiveness of alternative combinations of sensitivity and specificity, as a function of the prior probability of disease. The challenge region describes combinations of sensitivity and specificity for which a new technology would be sufficiently cost-effective to improve upon the old technology, at least for some groups of eligible patients.

The value of identifying this challenge region is in the elimination of unnecessary research; for example, if the challenge region lies beyond the

projected capabilities of a new technology further research may be deemed unnecessary. However, Phelps and Mushlin do not incorporate estimates of the likely outcomes of potential research and, hence, do not estimate the expected value of further research.

In March 2000 a citation search on Phelps and Mushlin¹⁰⁵ identified 42 references including both methodological and case study papers. In particular, the identified papers focused on magnetic resonance imaging and diagnostic technologies for heart disease.

Meta-analysis by the confidence profile method⁵⁵

Methods

The confidence profile method (CPM) is introduced primarily as a "new set of meta-analytic methods" for multiparameter evidence synthesis, which specifies a chain of evidence structure (the general model) to estimate the relevant outcome parameter. The general application of the CPM can be used to estimate the effect of an intervention on health outcomes, given data estimates.

Eddy⁵⁵ also discusses the use of the CPM to incorporate the possible results of further research to inform the value of such research. Two alternative approaches to estimating the effects of new research on the posterior distribution of the parameter of interest are presented. First, the predicted results of further research can be added to the general model, which is re-estimated to calculate the new posterior distribution. Secondly, the posterior distribution for the parameter of interest estimated from the existing evidence is treated as a prior distribution, which is revised on the basis of the expected research results.

The chapter then discusses the expected value of research, which is described as the additional utility derived from the treatment decision informed by the additional research. The estimation of the value of infinite research (perfect information) is defined as the difference between making a treatment decision based on the true value of the parameter of interest over the existing posterior distribution for the parameter, and making the treatment decision based on the current mean estimate of the parameter of interest.

The value of finite research is described in the context of the postresearch valuation of the

additional data; that is, how to value the research once the results of the research are known. No discussion regarding the prediction of possible research outcomes is provided.

Results

Mathematical expositions of the estimation of the expected value of infinite (perfect) and finite (sample) information are provided. Use of the confidence profile method to assess the probability of a statistically significant result is presented, which covers the probability of a particular result and the effect on the posterior distribution. Eddy⁵⁵ also discusses the expected value of research and gives a simple mathematical and diagrammatic description of opportunity loss and probability of success.

Discussion

The potential use of the estimated value of infinite research as an upper bound on the value of research is noted. The main contribution of the CPM to the EVI approach, however, is in providing a method for combining data to describe the values of model input parameters.

An economic approach to clinical trial design and research priority setting⁹⁸

Methods

Claxton and Posnett⁹⁸ outline the basic principle of the EVI approach to clinical trial design, and in particular sample size definition, using a hypothetical example of a simple treatment decision between a conventional treatment (T_0) and an experimental treatment (T_1). The approach assumes that the INBs are normally distributed, and the methods used to estimate the EVPI, the EVSI and the ENBS have been described in Chapter 7.

Results

The results of the hypothetical EVI analysis show that if the MAICER is less than around £6500 then the baseline decision will be T_0 , otherwise T_1 would be provided on the basis of current information.

The analysis shows that for a MAICER under £3000, further research would be unlikely to be cost-effective because even if T_1 were better than T_0 , the value of the utility gained would be very small. The cost of a trial (of treatments T_1 versus T_0) is made up of a fixed cost plus a marginal cost

per patient included within the trial. A simple numerical example shows a curve of expected net benefit of trial, that is, the expected value of sample information minus the costs of the trial, which peaks at a sample size of 33.

Discussion

Claxton and Posnett⁹⁸ describe the arbitrary nature both in philosophical terms and in practice, of the power and significance parameters of the traditional sample size calculation. The marginal costs of obtaining sample information are rarely included, which implies an infinite value on the benefit of sample information. They also point out that this traditional approach to trial design does not directly relate to the problem facing clinicians, that is, it does not directly refer to the net benefit of taking one decision choice over another.

In terms of the application of the EVI process, they recognise that the EVPI is determined by three factors: the slope of the loss function ($2/g$), the distance of the prior mean of the incremental net benefit from zero ($\delta_0 - \delta_b$), and the spread of the prior distribution σ_0 . The relationship between the MAICER and the EVPI is not simple or linear because the MAICER itself affects which decision is taken given current information.

The paper suggests that the EVPI and EVSI approaches can be used to construct two hurdles that proposed research must pass before being considered cost-effective:

- EVPI is greater than the fixed cost of the research, as the EVPI represents the opportunity cost of rejecting cost-effective proposals.
- EVSI is greater than the costs of the research, which ensures that the research design is technically efficient and conducted at the optimal scale.

In general, Claxton and Posnett⁹⁸ conclude that research resources should be concentrated in areas of clinical research where the marginal net benefit of research is high. They also state that the transfer of funding from service provision to research should continue while the net benefit of the marginal research proposal is positive, as such funding will increase the overall health benefits provided by the healthcare budget.

Harper and co-workers⁸⁷ question the practicability of the EVI approach described by Claxton and Posnett,⁹⁸ expressing doubt as to

whether a health technology assessment funding organisation could implement this approach effectively. In addition to the estimation of the MAICER, they question whether a sufficient degree of confidence and precision in the estimation of the EVPI and EVSI could be achieved that would enable their use as methods for assessing research proposals.

This appears to be the major issue for the EVI approach. The case studies have now begun to accrue over time and the practicality of the approach is becoming clearer. However, the position of the EVI approach within health technology assessment prioritisation mechanisms still requires consideration and discussion.

Designing a cost-effective clinical trial¹¹¹

Methods

This paper presents three methods for estimating the optimal sample size for a trial, and the accompanying expected cost of the trial. The estimation of the expected costs for all three methods is based on the expected loss of adopting the wrong treatment (i.e. the treatment with negative net benefits) plus the cost of the trial itself.

The Bayesian method is similar to that described by Claxton and Posnett,⁹⁸ whereby a Bayes risk function is specified, although Hornberger and co-workers¹¹¹ assume that only the clinical effectiveness parameters are uncertain. The risk function is based on prior beliefs about the distribution of the effectiveness parameters for the two interventions, P_1 and P_2 , expressed as a joint density function, $g(P_1, P_2)$. The Bayes decision rule is to use the treatment that minimises the expected loss, and the optimal sample size is that which minimises the expected cost of the trial.

The other two methods, the Neyman–Pearson I and II approaches, calculate the sample size required to identify a significant increase in the 2-year survival rate in the experimental treatment (T_2) relative to the informed survival rate of the traditional treatment (T_1). The specified minimal clinical difference is determined by a cost function that subtracts the monetary value of the difference in treatment effects ($Q_2 - Q_1$) from the difference in the costs of treatment ($C_2 - C_1$) to estimate the net benefits of treatment.

The Neyman–Pearson I approach assumes that in the case of equal net benefits between the two

treatment options, clinicians will continue to use T_1 , so a threshold minimum clinical difference is estimated on the basis of T_2 achieving a value of positive net benefits (NB*) sufficiently large to compel clinicians to switch to T_2 . The interval between equal net benefits and the threshold value of positive net benefits for T_2 represents the clinician's indifference to the choice of alternative treatments.

On the basis of the defined critical effect size, and a specified one-tailed type I risk (α) and power ($1 - \beta$), the necessary sample size is estimated using standard sample size calculation methods. The expected costs associated with the assumed decision rule (i.e. that T_2 will only be used if a certain level of positive net benefits is identified) is estimated by solving the Bayes risk function using the assumed decision rule and the estimated sample size.

Alternatively, the Neyman–Pearson II approach assumes that if net benefits are similar between the two treatment options, clinicians are indifferent between the alternative interventions, and require either a sufficiently large value of positive net benefits for T_1 (NB**) to compel them to keep using T_1 , or a value of positive net benefits for T_2 (NB*) sufficiently large to compel them to switch to T_2 . The Neyman–Pearson II approach assumes that the interval of indifference extends from some value of positive net benefits for T_1 (NB**) to a value of positive net benefits for T_2 (NB*). Thus, the critical effect size for the Neyman–Pearson II will always be at least as big as the estimated effect size for the Neyman–Pearson I.

The necessary sample size is estimated as in the Neyman–Pearson I approach, and the associated expected costs are estimated by the Bayes risk function. However, for the Neyman–Pearson II approach, a further assumption about the point of indifference between the upper and lower estimates of clinical effectiveness that relate to positive net benefits for T_2 and T_1 , respectively, is required. The point of indifference is the point at which clinicians are indifferent between T_1 and T_2 , which informs the decision rule within the Bayes risk function.

Hornberger and co-workers¹¹¹ apply the three methods to a case study to estimate the optimal sample size for a trial comparing different levels of fractional urea clearance in haemodialysis. It is assumed that two million people could benefit from the trial, that the average cost per trial subject is \$8000, and that the MAICER for a QALY is \$50,000. The prior density function for the two clinical effectiveness parameters assumes

equal effectiveness, but the respective standard errors are specified such that the difference in the 95th percentiles is equal to the difference in clinical effects that produces equal net benefits between the treatment options.

For the Neyman–Pearson II approach, the point of indifference was assumed to be halfway along the interval of indifference.

Results

The results from the Neyman–Pearson I approach showed that as the value of positive net benefits required to compel clinicians to switch to T_2 increases, the minimum clinical effect increases and the required sample size decreases.

In the Neyman–Pearson II approach, the interval of indifference will always be at least as large as for the Neyman–Pearson I approach, so the critical effect size will be larger and a smaller sample size will be required.

The baseline results (if NB* and NB** equal US\$5000) from the three methods indicate optimal sample sizes of 16,046 (expected cost US\$204 million), 3890 (US\$237 million), and 6200 (US\$111 million) for the Neyman–Pearson I and II and the Bayesian approach, respectively.

Discussion

The Neyman–Pearson approaches are hybrids of the EVI approach and the use of decision models to inform traditional sample size calculations. A novel aspect of their application is the incorporation of assumptions regarding the level of cost-effectiveness required to compel clinicians to provide the more efficient treatment, although such assumptions could also be applied to the standard EVI approach. The drawbacks of the Neyman–Pearson approaches include the fact that the sample sizes calculations only accommodate uncertainty around the clinical effectiveness of the new intervention, and that the estimated sample size for each interval of indifference (i.e. each specified decision rule) does not necessarily minimise expected costs.

Unfortunately, Hornberger and co-workers¹¹¹ do not discuss the merits of the three methods presented, preferring to concentrate on the general advantages of the explicit calculation of the costs and benefits of potential trials, and consider a series of important issues in this regard:

- The choice of any proposed prior distribution and loss function needs to be justified, for

example, summarising results of a literature review using standardised techniques, such as meta-analysis or decision analysis.

- The parameters of the prior distribution can be altered to test the sensitivity of the results to the choice of distribution.
- External users may define different prior probability distributions to learn how their beliefs affect the calculation of the sample size.
- The need to specify values for variables for which there may be little information does not discredit the approach, as many clinical decisions require the analyst to include variables for which there is little information.
- Experts may be surveyed about certain parameters to assess the effect on the sample size of changes in this variable.

In relation to the costs of undertaking a modelling approach to the design of clinical trials, Hornberger and colleagues¹¹¹ cite the example of the Diabetes Control and Complications Trial (DCCT) in patients with type 1 diabetes, which cost more than \$100 million. They state that less than 1% of the trial budget would be required to perform the steps proposed, which could have avoided collection of excessive data (and hence reduced research costs).

In general, they recommend that cost–benefit assessment of potential research will enable important design and priority setting tasks to be undertaken, and will also facilitate a revised analysis of cost-effectiveness once the primary research is completed. In addition, working in this way will enable the research community to understand much better how research has reduced the uncertainty in the decision-making process.

It is recognised that substantial analytical efforts are required, but Hornberger and co-workers¹¹¹ suggest that these may well be worthwhile in the design of large-scale trials. One could go further and suggest that the early attempts at economic evaluation are very likely to provide useful input to both the priority setting and design of trials. They are not necessarily an additional cost but simply a rescheduling of the cost of economic evaluation to a time before the trial takes place.

The cost–benefit of a randomised trial to a healthcare organisation¹¹²

Methods

A simple decision tree analysis is presented comparing standard therapy with an experimental

treatment, about which there is uncertainty concerning its probability of success or failure. A Bayesian approach is suggested using a pretrial joint distribution for the probability of success on both the standard and experimental treatment. On the basis of this prior distribution, the difference in expected loss between the two treatments, and the probability of success for the experimental treatment leading to indifference between the experimental and standard treatments, can be estimated.

Hornberger and colleagues¹¹² then provide equations for the expected loss after the trial, which incorporate the probability of success of the standard and experimental treatments, given new evidence about the mean expected values of success. Given n patients in each arm of the trial and a total of N patients available for treatment after the conclusion of the trial, the total expected net benefit of the trial can be estimated.

A case study is presented that evaluates dapsons, a potential therapy for people suffering from multiorgan system failure in intensive care settings, the mortality for which is as high as 80–90%. The analysis only includes one clinical parameter, specifying prior probability distributions for the success of standard treatment (50% success with standard deviation of 4%) and for the experimental treatment (50% success with standard deviation $\pm 10\%$). A correlation factor of 0.4 between the likelihood of success between the two trial treatment arms is also specified.

Resource costs include the costs of treatment and expected lengths of inpatient stay given success (30 days) or failure (20 days) of treatment. They also estimate the expected QALYs lost if a patient dies at 17.3 years and the quality of life per day of hospitalisation. The estimation of the relevant patient population is based on an assumed impact of dapsons over 10 years. The value of an additional QALY is assumed to be US\$50,000.

Results

The expected loss is minimised with a trial where the number of subjects in each arm is 61. The sample size requirements for different levels of policy threshold and different parameters for the prior distribution are also estimated, and it is shown that negative net benefits would be expected from any trial if the prior expected rate of success with dapsons is less than 35% or higher than 65%.

It is also shown that if the uncertainty regarding the effectiveness of dapsons increases, then the

optimum sample size decreases; for example, specifying a probability distribution where the 95% intervals are between 5% and 55% the sample size per group declines to 31.

Discussion

Hornberger and Eghtesady¹¹² outline the benefits of classical hypothesis testing; namely, that it is widely used, based on valid statistical theory for drawing cause or influences from trial data, and has well-established criteria to limit the risks of making the wrong inference. However, limitations of the classical method include the facts that it does not enable consideration of the costs of the trial or the alternative treatments, or the effect of the choice of the wrong intervention on the patient's length and quality of life in the case of either a type I or a type II error.

They suggest the following advantages of using their method.

- It directly includes the costs and long-term outcomes associated with the trial and the alternative interventions in the estimation of sample sizes.
- The assumptions underlying the analyses are explicit and available for critical appraisal.
- A single measure of the maximum acceptable costs for prolonging life by 1 QALY could be applied across all projects contemplated, which enables a comparative assessment of different proposals.
- Sensitivity analyses highlight the areas of uncertainty that may improve trial design.

Responding to anticipated criticism of the use of subjectively defined prior probability distributions, Hornberger and Eghtesady¹¹² cite the subjectivity involved in defining the critical effect size in traditional sample size calculations as well as the use of the conventional values of type I and II errors.

The methodology described by Hornberger and Eghtesady¹¹² is essentially an assessment of the expected value of perfect information for a single parameter. Choosing a particular value of the sample size allows an assessment of the expected value of sample information under exactly the same methodology and algebraic approach as that expressed by Claxton and Posnett.⁹⁸

A unique aspect of the methodology adopted in this study is the specification of a correlation between the pretrial expectations of the success of the alternative interventions.

The irrelevance of inference: a decision-making approach to the stochastic evaluation of healthcare technologies¹¹³

This paper is split into two broad sections. The central argument of the first part of this paper is that if the MAICER can be defined, the acceptance of the “null hypothesis when a new treatment has a positive but statistically insignificant mean incremental net benefit imposes unnecessary cost which can be valued in either monetary or effectiveness terms.”¹¹³ The stated correct alternative is to base decisions solely on mean cost-effectiveness. A hypothetical example illustrates the increased expected net benefits derived over a prolonged period (10 years in the example) from the provision of an intervention with a statistically insignificant effectiveness advantage over the existing treatment.

Claxton¹¹³ recognises that establishing the MAICER may be problematic and may vary according to the time and location of the decision problem, the perspective of the decision-maker and the measure of health outcome adopted. He describes similar options to those suggested by Fenwick and colleagues⁹⁹ as either the estimation of the marginal societal willingness to pay for a unit of health benefit, or the use of the ICER associated with the least cost-effective intervention currently provided. To calculate the marginal ICER, the use of a linear programming approach with the objective of maximising health gain subject to the budget constraint is cited.¹¹⁴ Claxton¹¹³ argues that through the specification of the MAICER, it is not possible to allocate resources on the basis of cost-effectiveness without determining the price per effectiveness unit.

The first part of the paper then refutes some of the arguments that have been made against the net benefit decision criteria. In relation to the need to incorporate concerns about equity in the provision of healthcare (however defined), Claxton¹¹³ argues that explicit adjustments to the measure of outcome can be made to represent equity issues. Adjustments to the measure of outcome can also be implemented to address other similar concerns, such as a preference for the prevention or cure of rare but catastrophic events (and attitudes to risk generally), or a particular concern not to do harm.

The second part of the paper extends the application of the EVI approach described by

Claxton and Posnett⁹⁸ to more specific areas of trial design; in particular, noting that where expected costs differ between interventions a strategy of uniform sample distribution between the treatment arms will be suboptimal. The approach proposed by Claxton¹¹³ requires the estimation of the ENBS for every feasible allocation of each sample size between the relevant interventions. The optimal sample size, and allocation of that sample, is where the ENBS reaches a maximum.

Previous attempts to inform trial design using a decision-theoretical framework, which led to very large or unbounded predicted sample sizes, are discussed. Claxton¹¹³ argues that this is not surprising because these approaches excluded resource costs, which effectively assumes the marginal cost of sample information is zero. In relation to the impact of the results of the research on the decisions of clinical practitioners, Claxton¹¹³ argues philosophically that it is better first to identify the right treatments and then to devise methods to persuade clinicians to implement cost-effective interventions. On this last point, one could argue that it is possible to incorporate the uptake of research findings, including uncertainty around uptake, into a modelling approach.

Case study: EVI of research on diagnosis of urinary tract infections⁹⁹

Methods

Fenwick and co-workers⁹⁹ use a decision tree framework to assess the EVI of alternative approaches to managing symptoms presenting as possible urinary tract infections (UTIs). Following discussion with clinicians and a review of the literature, seven simple management strategies are identified:

- no treatment
- empirical treatment
- empirical treatment plus laboratory tests
- dipstick test followed by treatment
- dipstick test followed by treatment plus laboratory test
- laboratory test and wait for preliminary results
- laboratory test and wait for full results.

The model is intended to determine the cost and QALYs associated with each patient management strategy. Data to populate the model are obtained

from a review of the clinical literature and subjective expert opinion.

A separate analysis to estimate the EVPI for each individual strategy is undertaken on the basis of a defined maximum acceptable value of QALY (MAICER). As no explicit value of a QALY (to NHS decision-makers) is identified, the EVPI is estimated on the basis of three potential values: £5000, £10,000 and £20,000.

Probability distributions are assigned to every parameter within the model: log-normal distributions for unit costs; utility values and event time parameters (as they are all bounded by zero and tend to have skewed distributions), and triangular distributions to probability parameters.

Each simulation involved 1000 iterations, from which the probability of the optimality of each strategy is estimated. Those strategies that are never optimal can be confidently excluded from further data acquisition requirements. Choosing the expected optimal strategy on the basis of the mean INBs over the 1000 iterations, the EVPI for the full set of model parameters was estimated using the procedure described in Chapter 7.

To estimate the partial EVPI the values of the specified parameters were fixed at their mean values and the estimated EVPI was subtracted from the full EVPI. As noted in Chapter 7, such an approach may not fully account for the current (prior) level of information about the specified parameters. This issue constitutes one of the methodological differences between this study and other papers in the area.

Finally, the societal (as opposed to per patient) value of perfect information is estimated by applying the per person opportunity loss to the expected number of patients presenting to UK general practice in a year, multiplied by the 5 years for which the research is assumed to remain relevant (no specific reason is given for the assumed length of research relevance).

Results

The base case analysis suggests that empirical treatment is the optimal strategy with a cost of £264 per additional QALY. One-way sensitivity analysis suggests that the decision-maker can be reasonably confident that the cost-effectiveness for empirical treatment lies below £500 per QALY. However, there is less confidence associated with the dipstick strategy, which fluctuates from £220 per QALY up to £33,900 per QALY on the

variation of one parameter. Using a MAICER of £5000 per QALY, stochastic analysis suggests a 37% chance that dipstick treatment is optimal, while at £10,000 and £20,000 the dipstick strategy is optimal, with a 38% and 41% chance, respectively. The EVPI suggests that there is considerable value associated with further data acquisition concerning model parameters as a whole (£4.2 million; £8.6 million and £17.6 million for the three alternative MAICERs, respectively).

The analysis for groups of parameters suggests that further research on the relevant utility values provides the most value (£2.7 million, £5.7 million, £12.4 million), followed by event times (£1 million, £2 million, £4.1 million, respectively). The parameters of effectiveness of treatment are less important. The value of information for antibiotic effectiveness is £0.7 million, £1.3 million and £2.7 million. Dipstick accuracy (£0.7 million, £1.3 million and £2.8 million) and laboratory test accuracy (£0.6 million, £1.1 million, £2.2 million) are all much less uncertain and much less valuable in terms of further research.

The results indicate that further primary research would be justified. Moreover, the parameters for which the EVPI is greatest (i.e. utility values and probabilities of UTI given symptoms) would not require measurements within expensive RCTs. Further data could probably be generated using observational data collection methods. Non-experimental designs could also be appropriate to provide further data on dipstick and laboratory test accuracy and the probability of resolution of UTI without intervention.

The conclusion drawn from this early stage modelling is that a trial is not the most urgent research design in this area.

Discussion

Fenwick and co-workers⁹⁹ argue that the EVI approach is useful in unifying decision-making about the most efficient services to provide, the explicit value of additional information from primary and secondary research, and the most efficient means of acquiring that information. They further argue that modelling provides an iterative framework for evaluation, which should be instigated when a potential healthcare technology first emerges, and should continue to be used throughout a technology's life cycle to assist decision-making regarding routine service use and the value of additional information gathering.

The main benefits of the EVPI methodology that are cited include:

- the estimation of an explicit upper limit on the value of information
- the identification of parameters for which the collection of further data is unlikely to be cost-effective
- as an aid to the selection of appropriate research designs, such as when a key parameter is identified that is not vulnerable to selection bias (e.g. the incidence of a particular disease), then an experimental study design is unlikely to be required.

Fenwick and co-workers⁹⁹ also suggest that the EVI approach may usefully inform the design of secondary data acquisition exercises, in the form of systematic reviews. They argue that full systematic reviews should not be undertaken before the development of the initial model, as the efficiency of the search strategy could be improved, for example, by applying the following approach.

1. Describe a probability distribution for each parameter on the basis of the first five papers identified by a preliminary literature search.
2. Incorporate the system into the decision model.
3. Estimate the likely volume of published studies as yet unidentified.
4. Estimate the value of information offered by extending the literature search exercise.
5. Compare the value of information with the expected cost of further literature searching.

This is an exact analogy to the value of information collection from primary research. It does, however, presuppose the existence of reliable methods of meta-analysis and, perhaps more importantly, runs the risk of ignoring study selection bias within small samples of papers. Some of the formal requirements of a meta-analysis may therefore preclude such an approach.

Sensitivity analysis and the EVPI⁶⁹

Methods

This paper describes and compares four methodologies for assessing uncertainty in decision-making (as described in Chapter 4), including the EVPI approach. Felli and Hazen⁶⁹ describe the EVPI process and its general application to decision-making. Their description of the process for estimating the partial EVPI follows method 2 (see Chapter 7), where only the

parameters of interest are allowed to vary (and the opportunity loss is estimated assuming perfect information about the remaining parameters). As suggested in Chapter 7, a more systematic approach to estimating the EVPI involves a two-level simulation, where a series of sampled values for the parameter(s) of interest are held constant for a set of iterations and all other uncertain parameters are varied. However, the mathematical description of the EVPI process described by Felli and Hazen⁶⁹ describes the two-level simulation approach. The only difference is the order in which it is suggested that the simulation (or integration over the prior conditional probability distributions) is done, but the order should not affect the final answer since this depends only on the average opportunity loss over all the iterations.

Three case studies of decision analysis in the healthcare sector are used to compare the alternative approaches to sensitivity analysis: herpes simplex, deep vein thrombosis and a symptomatic bacteriuria.

Results

The EVPI approach is shown to describe the scale of importance of gaining further information and the parameters about which information is most important to collect.

Discussion

Felli and Hazen⁶⁹ believe that the EVPI approach is a natural extension to probabilistic sensitivity analysis, which provides additional benefits owing to the simultaneous assessment of the probability of making the optimal decision, and the change in payoff allied to an alternative decision. That is, the EVPI approach is consistent with the maximisation of expected value and the economic concept of marginal reasoning.

This paper is useful in clearly setting out the advantages of the EVPI approach to understanding decision uncertainty compared with other forms of uncertainty analysis, and contributes to the consensus view that EVPI is the most methodologically and philosophically consistent approach to measuring the priority of obtaining further information.

Bayesian approach to sensitivity analysis¹¹⁵

Methods

This paper compares two approaches to handling uncertainty – EVI and threshold sensitivity analysis

– in a case study evaluation of alternative treatments for pyriform sinus cancer.¹¹⁶ A decision tree describes the probabilities of survival and morbid events, leading to five alternative outcomes (disease-free survival, early relapse, delayed relapse, late relapse and death) for four treatment options (radiotherapy, surgery, radiotherapy then surgery, and surgery then radiotherapy).

The original analysis undertaken by Plante and colleagues¹¹⁶ included one-way threshold analyses of the probability of survival following surgery, quality-of-life parameters and the increased probability of survival due to irradiation following surgery. Seven combinations of parameters were also analysed using two-way sensitivity analyses.

Felli and Hazen¹¹⁵ undertook a full EVPI analysis, as well as a series of partial EVPI analyses for sets of two parameters (using method 2, see Chapter 7).

Results

The results of the EVPI analysis showed that the total parameter EVPI equalled 5.62% of the expected health benefits derived from the baseline decision (108 quality-adjusted weeks of survival). The partial EVPI results show that the probability of survival following radiation, combined with quality-of-life adjustment for radiation or mortality for surgery, provides the highest EVPI for parameter sets (3.29% and 2.95% of the baseline health benefits). All remaining parameter sets have EVPIs less than 2.7%.

Most importantly, the parameters identified as sensitive by the threshold analysis were clearly inconsistent with the EVPI results. The single parameter with the highest EVPI (disease-free survival following radiation) was not deemed sensitive by the threshold analysis. Furthermore, five of the parameters that were deemed sensitive had information values less than 1.2% of the optimal 108 quality-adjusted weeks.

Discussion

Felli and Hazen¹¹⁵ note that the differences between the conventional form of sensitivity analysis and EVPI is consistent with their general findings that conventional sensitivity analyses tend to overstate the sensitivity of model outputs to variation in input parameters. They conclude that traditional forms of sensitivity analysis address the question “how much do parameters affect results?”, rather than the more revealing question of “how likely and with what effect do parameters affect the results?”, which is better answered by the EVPI approach.

Bayesian value-of-information analysis: an application to a policy model of Alzheimer’s disease⁹⁷

Methods

This paper uses a pre-existing cost-effectiveness model of donepezil in the treatment of Alzheimer’s disease establishing probability distributions to describe uncertainty in the input parameters and undertaking an EVI analysis in the context of the US Alzheimer’s disease population. The model builds upon a placebo-controlled, double-blind clinical trial of donepezil with a follow-up period of just 24 weeks. The benefits of the drug were extrapolated using a state transition model (Markov process) to examine the progression through different disease states (mild, moderate and severe) and care settings (community and nursing home). The model was populated using trial data, longitudinal databases on disease progression and health state utilities for the seven states from observational data. Direct medical, non-medical and indirect costs were based on previous published analyses.

In this study, 10,000 Monte Carlo simulations are undertaken to establish the probability distribution for the INBs of treatment, which is done for different time horizons ranging from 24 to 210 weeks. The EVPI for partial sets of parameters is estimated by reanalysing the model holding the parameter of interests constant at their expected value (method 1, see Chapter 7). The EVI approach takes a societal perspective using a central estimate for the willingness to pay US\$50,000 per QALY and conducting analysis on values ranging from US\$1000 to US\$100,000.

One of the most significant aspects of this paper is the detailed discussion of the methodologies for selecting prior distributions for the uncertain parameters, which were described in Chapter 7.

Results

The analysis shows that the new treatment becomes cost-effective when cost and outcomes are considered beyond 54 weeks, although the uncertainty around the results increases as the time horizon of the analysis increases. For the modelled period of 210 weeks (4 years) the results are presented as a cost-effectiveness acceptability curve. For a MAICER of US\$50,000, the probability that donepezil is cost-effective is 0.6796, meaning that the error probability that the new treatment is not cost-effective is 0.3204,

which is greater than the conventional benchmarks of 0.05 or 0.025 used in both Bayesian inference and traditional frequentist statistics.

Claxton and co-workers⁹⁷ then estimate the expected costs of not adopting donepezil on the basis of statistically insignificant estimates of cost-effectiveness, which are US\$1220 or 0.0244 QALYs foregone for each individual patient. When calculated in terms of the US population for the next 4 years, the expected loss would be US\$164 million or 21,279 QALYs foregone.

The results of the EVPI per patient at different time periods show that the value of information over a short time horizon (24 weeks) is very small, but rises as the uncertainty about the extrapolated effect of donepezil increases across the time horizon considered. Assuming a MAICER of US\$50,000 and a time horizon of 210 weeks, the EVPI is US\$339 million.

The partial EVPI results show that the highest benefits from further research are most likely to come from data describing efficacy duration (US\$270 million), RR ratio beyond 24 weeks (US\$93 million), efficacy within the existing trial period (US\$84 million), and the dropout rate (US\$39 million). The authors stress that the partial EVPI is not additive across the individual EVPIs for each parameter because there are joint effects within the model.

Discussion

Claxton and colleagues⁹⁷ claim that the EVI approach answers nine different questions surrounding research priority and design:

- Is additional research in Alzheimer's disease potentially cost-effective?
- Are the estimates of the model inputs adequate?
- For which model inputs would more precise estimates be most valuable?
- Is experimental design required for subsequent research?
- If so, which end-point should be included in any future clinical trial?
- What is the optimal follow-up period?
- What is the optimal sample size?
- How should trial entrance be allocated between the arms of the trial?
- What is the value of this proposed research?

Discussing the results of the EVI analysis, the authors note that all of the important variables identified by the partial EVPI analysis are liable to

selection bias, so an RCT would be required to collect such data. Additional research on other parameters, such as baseline transition probabilities, direct costs or utility values, may be more efficient because an experimental trial design is not required and the associated data collection costs may be lower.

More specific to the design of an RCT, it is noted the exclusion of any end-points will always reduce the benefits derived from further research, but those losses may be balanced by an accompanying reduction in the fixed and marginal reporting costs of sampling. Claxton and co-workers⁹⁷ also point out that EVI methodology provides an empirical solution to the debate regarding large clinical trials versus trials with economic content, identifying circumstances where large and simple clinical trials may well be efficient, as well as cases where trials with economic end-points will be required. It is also suggested that the EVI approach is useful in identifying relevant alternatives in patient management, as the optimal sample allocation is based on the marginal benefit and the marginal costs of allocating patients to potential interventions. If a particular intervention is shown to be inefficient, then that alternative can be excluded from the research process.

Claxton and co-workers⁹⁷ conclude that, given the requirements of the US Food and Drug Administration for health economic claims to be "sufficiently substantiated" through the use of "competent and reliable scientific evidence", the EVI approach aids the design and conduct of research, as well as the setting of research priorities and the regulation of new technologies.

EVPI in liquid cytology for cervical screening¹⁰⁰

Methods

This paper was undertaken to help NICE to establish the likely cost-effectiveness of the new liquid cytology techniques for cervical screening. The report describes a central estimated value for every parameter within the model, together with maximum and minimum values for 32 key parameters. A mix of published literature evidence and personal communications was used to identify the central estimates and plausible range for each parameter.

Both full and partial EVPIs are presented. The relevant population is assumed to include 100,000

new women joining the screening programme at the age of 18 years each year for the next 5 years.

The two-level sampling approach (method 3, see Chapter 7) to the estimation of partial EVPI is used, whereby the value of the parameter of interest is sampled from its plausible range. A simulation is then run allowing all other parameters to vary, and the EVPI is estimated. A second value for the parameter of interest is then sampled and again the simulation is run allowing all the other parameters to vary. This process is repeated for 1000 samples of the parameter of interest, and the mean partial EVPI is obtained from the distribution of EVPIs.

Three different scenarios were examined for discounting costs and benefits, respectively; these were (6%, 1.5%) (3%, 3%) and (6%, 6%).

Results

The results show that the EVPI is sensitive to the discount rate, to the extent that, at a value of a life year of £20,000, the overall EVPI is around £200,000 at discount rates of 6% for costs and 1.5% for benefits, while the EVPI would approach £2.8 million if a discount rate of 6% were adopted for costs and benefits.

The EVPI for each parameter using (6%, 6%) discount rates and a threshold cost-effectiveness of £20,000 per life year gained show that the most important parameter is the marginal cost for a liquid cytology sample (this was assumed to be £3.50 with a plausible range of £0 to £7). The estimated partial EVPI was 2.4 million. The second most important parameter is the proportion of inadequate samples, which has an EVPI of around £900,000. All of the parameters had an EVI below around £600,000.

Discussion

The EVI analysis is used primarily as a technique for sensitivity analysis. The methodology clearly shows the prime causes of uncertainty in the model.

The clinical-effectiveness and cost-effectiveness of machine perfusion versus cold storage of kidneys for heart-beating and non-heart-beating donors¹⁰¹

Methods

This review was commissioned by the HTA R&D Programme to explore the clinical and cost-

effectiveness of machine perfusion (MP) versus cold storage (CS) in kidney preservation. This work was commissioned as a systematic and rapid review at the same time as the longer process for commissioning primary research in the same field was initiated. The study reports an analysis of the expected value of further information, undertaken with the intention of informing the design of a prospective clinical trial.

Multivariate sensitivity analysis was conducted for random variables within the model. Prior distributions of model parameters were estimated using a combination of published information and expert clinical guidance. Where a log-normal or normal distribution was assumed, the standard errors were chosen to allow for wide uncertainty in the model. Uniform distributions were assumed for those model parameters where little prior knowledge was available.

A systematic review was undertaken to identify the impact of MP and CS on delayed graft function (DGF) and graft survival in patients receiving kidneys from heart-beating and non-heart-beating donors. In addition, a Cox proportional hazards model, based on the literature, was used to estimate the impact of DGF on long-term graft survival. The assumed population of interest used within the model comprises both heart-beating and non-heart-beating kidney transplant recipients in England and Wales. Cost-effectiveness planes were constructed to explore the range of uncertainty surrounding the model parameters. The overall EVPI was reported alongside the partial EVPIs for each parameter within the model, assuming a maximum cost per QALY threshold of £20,000.

A two-level Monte Carlo approach was used to calculate the EVPI for parameters. The value of each parameter of interest was sampled from its plausible range, and a simulation was then run allowing all other parameters to vary, thus estimating the EVPI. This process was repeated for 1000 samples of the parameter of interest and the mean partial EVPI was obtained. The EVPI for heart-beating donors and non-heart-beating donors was analysed separately.

Results

The baseline results from the analysis suggest that MP dominates over CS, which is cheaper and more effective than CS, for both heart-beating donor and non-heart-beating donor recipients; however, key uncertainties in the assessment mean that MP may also be dominated by CS. Wight and

co-workers¹⁰¹ present cost-effectiveness acceptability curves (CEACs). The probability that this is the case is estimated at around 80% for non-heart-beating donor recipients and 50–60% for heart-beating donor recipients. The population EVPI was estimated to be £28,750 for heart-beating donors and £1,368,000 for non-heart-beating donors. Partial EVI analysis suggested that the percentage of patients with DGF receiving CS kidneys, the risk factor for graft loss, and in particular, the relative risk of DGF for MP were the most important parameters for which further information would yield the most value.

Discussion

This EVI study suggests substantial value in further research on the clinical benefit of MP over CS of kidneys for transplantation. Specifically, studies should focus on establishing the effect on long-term graft survival. Owing to the small predicted effect size on graft survival, long-term trials on this end-point may be infeasibly large. Thus, a more practical approach may be to undertake a short RCT focusing on the impact of MP on delayed graft function; this, however, should be supported by further work on validating the link between DGF and graft loss.

EVPI in screening for inborn errors of metabolism using tandem mass spectrometry¹⁰²

Methods

This review of the clinical and cost-effectiveness of tandem mass spectrometry (MS/MS)-based neonatal screening for inborn errors of metabolism was commissioned by the NHS HTA R&D Programme. Two previous systematic reviews had been undertaken, but were not entirely consistent in their conclusions. In particular, controversy existed about the requirement for and value of further research in this field. Economic modelling was conducted to indicate the likely effects of introducing MS/MS in the current UK screening programme for medium-chain acyl-coenzyme A dehydrogenase deficiency (MCADD) and for phenylketonuria and the ability of this new

technology to screen simultaneously for other conditions.

Probabilistic analysis was undertaken by assigning prior distributions to represent the uncertainty surrounding model inputs. A Bayesian approach to analysing uncertainty was adopted. Full and partial EVPIS were calculated, using two-level Monte Carlo simulation, in order to estimate the value of future research.

Results

CEACs were constructed across a range of thresholds. This analysis suggested that, given the available evidence used in the model, there is a high probability that using MS/MS for phenylketonuria and MCADD screening combined would be cost-effective even at relatively low threshold values. Simulations were also run to produce simple 'cost-per life years gained' and 'incremental net benefit' estimates for each of the other main conditions. These ranged from £703 per life year gained for long-chain fatty acid defects to £10,902 for tyrosinaemia.

The population EVPI was calculated for all model parameters across a range of thresholds. Even at a threshold of £30,000 per life year gained, the population EVPI was estimated to be only £2,119. These results showed that the most significant parameters to affect overall cost-effectiveness are the frequency or incidence of the condition and the future health and social care costs related to disabilities caused by a failure to recognise the disease early. The EVPI analysis reported by Pandor and co-workers¹⁰² also shows that another major driver of the cost-effectiveness of screening for metabolic conditions is the extent to which specific disorders lead to significant avoidable disabilities and the future resource costs that these problems impose on the health and social care sectors.

Discussion

The EVI analysis indicates that on the current evidence base there is little economic justification for further research into the potential effectiveness of MS/MS in identifying MCAD deficiency and phenylketonuria.

Chapter 9

Other related literature

Overview

This chapter briefly summarises literature that is of some importance to the topic but does not fit neatly into the other chapter categories. It includes three classes of literature:

- papers on R&D priority setting in commercial (particularly pharmaceutical company) settings
- the mathematical modelling of biological processes and its iterative use in trial design
- the role of direct simulation of clinical trials.

Commercial investment appraisal approaches

There is a very small published literature on modelling to inform commercial R&D, although a high volume of unpublished modelling work focusing on decisions about the R&D of pharmaceuticals has been undertaken. Three published papers are reviewed in this section.^{109,110,117} The review describes the methodologies and compares them with the methods reviewed in earlier chapters, as well as discussing the feasibility of translating these commercial approaches to a governmental or societal perspective.

An investment appraisal approach to clinical trial design¹¹⁷

Methods

From the commercial perspective, Backhouse¹¹⁷ identifies the role of clinical trials as providing access to a new drug's intended market or producing evidence of sufficient strength or relevance to secure or enhance the use of a product. Three main decision rules for the adoption of a proposed trial are suggested.

- A trial of given design is worth conducting if it yields a positive expected net present value (NPV).
- The optimal choice of trial design, in terms of factors such as sample size and primary end-points is that which maximises the (positive) expected NPV.
- When allocating funds to different studies competing for a limited trial budget, the funds should be allocated across potential trial designs

so as to maximise the expected NPV of the overall investment.

The NPV of a proposed trial is defined as the expected revenue minus the expected costs, which is estimated on the basis of the following equations:

- the discounted cost of the trial (which is a function of n = sample size and Q , the drug volume used in the trial).
- three different demand functions.
 - the desired demand function (a function of drug price and other factors including, in particular, the strength of available evidence)
 - actual demand (a function of desired demand and a diffusion or uptake factor over time)
 - expected demand (a function of the expected effect sizes for the product attributes, which is in turn a function of sample size, significance and power of the trial)
- revenue to the pharmaceutical company, which is a function of the mean expected demand (the aggregation of the expected demand of different effect sizes of possible trials and their probabilities of occurring), the price and the period of sales.

The estimation of the expected demand, as a function of the probability of alternative trial results, requires a decision-analytic approach. The company needs to estimate the expected outcomes with respect to potential trial end-points and their probability of occurrence. This method also requires an estimate of the useful life of the information provided by a trial.

The approach is illustrated with a simple example of a proposed Phase IV trial comparing new drug A versus existing drug B. There are two potential trial end-points: X_1 and X_2 . The performance of the existing drug B has been well studied and it is 25% successful on end-point X_1 and 75% successful on end-point X_2 . Classical sample size calculations are undertaken on the assumption that the trial should only investigate endpoint X_1 (to detect an absolute difference of 30%, with 90% power and a 5% level of significance).

The investment appraisal approach considers the NPV of conducting different size clinical trials

comparing drug A with drug B with alternative end-points (X_1 , X_2 , or X_1 and X_2). Prior distributions are postulated for the improvement of drug A on end-point X_1 (mean 15%, range +11% to 20%) and on end-point X_2 (mean 5%, range +1% to 10%). The desired demand function assumes that proven success in meeting endpoint X_2 has a greater impact on demand than in meeting endpoint X_1 , while success in meeting both end-points is assumed to yield a 100% market share for drug A.

Results

A larger sample size enables a more accurate assessment of the true performance of drug A, which creates a higher power for the detecting of a significant difference correctly. Following this logic and some algebra, the paper shows that as N (the sample size) increases then the expected value of the detected superiority of drug A over drug B on dimension X_1 rises to a threshold level (15.59%). This, in turn, affects the desired demand function and, hence, given parameters about discounting and market adoption, the discounted revenue. As N rises the discounted revenue rises to approach a ceiling value. The trial costs, however, are linear with sample size. The NPV of the trial therefore reaches a peak and then begins to drop as no extra-expected revenue is created from an even larger sample size.

Given the parameters outlined, the maximum NPV for a trial considering only endpoint X_1 is achieved at sample size $N = 225$ (maximum NPV around £1 million). The traditional sample size calculations required a trial of 125 patients per treatment arm, which shows that although the classically powered trial is worth undertaking it is suboptimal in terms of the strength of evidence produced and therefore the expected revenue.

If X_2 alone is the clinical end-point of the trial then the maximum NPV is £1.25 million and the sample size required would be 400. However, if both X_1 and X_2 clinical end-points are measured within the trial then, because of the additional strength of evidence, the expected revenue is maximised at £4.1 million (four times higher than the revenue with a trial on one of the single clinical end-point dimensions) and the maximising sample size is 425 per arm.

Discussion

Three main areas for further research are identified. First, practical issues around the description of the proposed RCT need to be addressed, such as choice of time horizon, the

determination of the trial's opportunity cost of capital, and how to assign capital, production, distribution, marketing and sales costs to individual trials.

Secondly, difficulties in accurately estimating the demand functions are raised. Specifically, Backhouse¹¹⁷ comments on defining the role of clinicians versus other decision-makers in the product adoption process, the relative importance of different types of product differentiation data to different decision-makers, the nature of the relationship between differences in product characteristics and prescriber take-up, and factors influencing product diffusion that are within the control of the company. The likelihood that such factors may be context specific forms a further problem.

Backhouse¹¹⁷ suggests that there is considerable benefit in an early analysis of the desired demand function or the key parameters that affect adoption and diffusion. This may well influence the design and also the priority of trials with different clinical end-point measures.

Finally, the ascertainment of prior distributions for the expected end-point outcomes is cited as a problem. One research-intensive solution is to develop a 'community of priors', based on alternative approaches to specifying prior distributions, and to test the sensitivity of the appraisal results to the choice of distribution.

Backhouse¹¹⁷ describes alternative applications of the approach, such as the choice of trial design parameters (e.g. the choice of comparators or the duration of patient follow-up), to optimise the design of late phase development strategies, or to maximise the NPV of a company's Phase III and IV trial portfolio. Backhouse¹¹⁷ is, however, pessimistic about the prospects for pharmaceutical companies adopting the investment appraisal approach, as the existing industry objective (for clinical trials) is perceived to be securing marketing approval, rather than boosting market adoption.

Possible links to a societal perspective

This investment appraisal approach is similar to the payback literature in that it calculates the expected profit from undertaking the trial on the basis of the probability of alternative trial outcomes. It also uses some of the probabilistic architecture associated with EVI approaches; for example, prior probability distributions for the effectiveness of the interventions.

This approach does not directly consider the potential cost-effectiveness of the proposed new intervention. Indeed, the discussion of the analysis suggests that the demand is a function of the clinical difference discovered between the new and existing therapy, rather than the cost of the new drug. However, Backhouse¹¹⁷ suggests that the approach might be used to evaluate the incremental costs and benefits of including economic data capture within proposed trials.

The comparison of private investment appraisal approaches with those that adopt a societal perspective (citing Detsky⁸⁹ and Claxton and Posnett⁹⁸) is suggested, in order to identify conditions under which societal and industry objectives reach similar conclusions. Backhouse¹¹⁷ recognises that the degree of concurrence will depend on the relative importance of cost-effectiveness considerations in drug-prescribing decision-making. In the UK, the increasing role of NICE and cost-effectiveness in informing resource allocation decisions suggest that societal and industry objectives will draw ever closer together.

Some statistical issues in project prioritisation in the pharmaceutical industry¹⁰⁹

Method

This paper focuses on commercial issues of R&D prioritisation and considers that an NPV approach, which accounts for the probability of (R&D) success, the expected rewards if successful, and R&D costs, should inform portfolio management. The Pearson index (expected profit divided by cost of development) is used to optimise an R&D process that has a sequence of stages.

Senn¹⁰⁹ notes the scale of research investment in the pharmaceutical industry (annual expenditure on R&D in 13 leading countries rose from US\$5 billion to US\$22 billion between 1981 and 1991) and the importance of investing wisely (the number of chemical entities declined from 66 to 40 between 1981 and 1993). He also notes the increasing importance of the evaluation of the cost-effectiveness of new drugs.

The central idea is that there may be projects with the same overall probability of success and the same overall reward if successful, but if projects have different probabilities of failure at stage 1, 2 or 3 then the expected net value of the projects will differ. A worked example concerning four projects with the same expected reward but different cost architecture is provided.

TABLE 10 Illustration of the Pearson index using four hypothetical projects¹⁰⁹

| | Project | | | |
|--|---------|-------|-------|-------|
| | A | B | C | D |
| Costs | | | | |
| Stage 1 | 3 | 3 | 1 | 1 |
| Stage 2 | 2 | 2 | 2 | 2 |
| Stage 3 | 1 | 1 | 3 | 3 |
| Total cost | 6 | 6 | 6 | 6 |
| Probability success | | | | |
| Stage 1 | 0.8 | 0.4 | 0.8 | 0.4 |
| Stage 2 | 0.6 | 0.6 | 0.6 | 0.6 |
| Stage 3 | 0.4 | 0.8 | 0.4 | 0.8 |
| Overall probability of success | 0.192 | 0.192 | 0.192 | 0.192 |
| Reward if successful | 28 | 28 | 28 | 28 |
| Expected development costs | 5.08 | 4.04 | 4.04 | 2.52 |
| Expected reward | 5.376 | 5.376 | 5.376 | 5.376 |
| Expected value (expected reward – costs) | 0.296 | 1.336 | 1.336 | 2.856 |
| Pearson index (expected reward/costs) | 0.058 | 0.331 | 0.331 | 1.133 |

Results

The results of the hypothetical example are presented in *Table 10*, which shows that, *ceteris paribus*, projects with costs loaded further down the R&D process, and with higher thresholds for success earlier in the process, have a higher expected value.

Senn¹⁰⁹ develops a simple mathematical description of projects that have N stages of equal cost and equal probability of success, and proves that a one-stage project is always less valuable than the N -stage project. He extends this to assume a cost of delay due to checking at each of the N stages, and describes the trade-off between the advantages of the “cost–probability architecture” of the N -stage approach and the reduction in the costs of delay from an all-in-one design. The mathematics to find the optimum number of stages for the project to increase the expected value is shown.

Discussion

Faster drug development is not always deemed to be better from a pharmaceutical company's

perspective. Moreover, if regulations slow down the R&D process they may encourage more optimal use of information structures.

Senn's¹⁰⁹ discussion of the weaknesses of the Pearson index relate to the commercial environment, which may not be fully reflected in the proposed approach. External factors may lead to the abandonment of a project; for example, a medical commission may decide that the appropriateness of a class of products can never be justified for a particular patient subgroup (although the probability of such factors could be incorporated into the prior estimates of the probability of success).

In general, Senn¹⁰⁹ presents an adaptation of the payback approach from a commercial perspective, although he clearly demonstrates the value of multistage projects as opposed to a single-stage project. This latter issue certainly has an application in the field of governmental health technology assessment, which has not been well addressed in any of the existing literature.

The effect of pharmacoeconomics on company research and development decisions¹¹⁰

Methods

The purpose of this paper was to review the rationale for integrating pharmacoeconomics into R&D project selection and termination decisions in the pharmaceutical industry.

Grabowski¹¹⁰ reviews "the new competitive dynamics", which includes increased assessment of cost-effectiveness by funders of healthcare services, and price discounting that can be necessary when there are other cost-effective products already in use. He describes seven sequential stages in the R&D process at which decision-making under uncertain conditions is required:

1. discovery programme in a particular disease area
2. preclinical development of a promising compound
3. first human testing
4. first efficacy testing in patients
5. large-scale clinical testing
6. regulatory submission
7. marketing launch.

Grabowski¹¹⁰ suggests an analysis of the current treatment options to produce benchmark values for the cost-effectiveness of existing therapies, as well as to identify the key parameters that

influence cost-effectiveness. The cost-effectiveness of the new drug can then be modelled, incorporating different assumptions about the efficacy, tolerability, pricing and formulation of the new therapy. Such analyses inform the targeted collection of data on this issue, and may indicate that a product candidate is potentially highly cost-effective. They could also lead to a re-estimation of the projected market size and/or price for the product and cause the company to assign a higher priority status to the development project within its overall R&D programme.

Results

A brief overview of this modelling approach in the area of anti-neoplastic therapy for non-small-cell lung cancer and cardiac transplantation is provided, as well as two further short examples where pharmacoeconomics have helped to shape the R&D portfolio of the pharmaceutical company. In the first case, a product was judged to have equal efficacy to an existing market competitor, so the R&D process was terminated because the price of entry to the market was too low to make the drug profitable. In the second case, although a product was able to show a significant incremental gain in efficacy over existing products, the gain was not large enough to demonstrate cost-effectiveness at the price level that would be profitable on rate-of-return grounds.

Discussion

Grabowski¹¹⁰ recommends the use of decision-analytic modelling to inform the continuation of the R&D process at a much earlier stage than immediately before Phase III trials. He also suggests that cost-effectiveness modelling should be used in conjunction with more traditional economic modelling concerning returnable investment analyses that incorporate cost, demand and other relative competitive considerations.

To company scientists, R&D directors and executives, the use of pharmacoeconomic analysis as an integral part of the go/no-go decision may be controversial because there is too much uncertainty about the cost-effectiveness of a product to make a decision using available data. Grabowski's response is that, at a minimum, the early integration of economic analysis into the strategic decision-making process will improve the targeted collection of data in Phase III clinical trials. In addition, the relative cost of such modelling is likely to be small.

Grabowski's recommendation that, in the commercial sector, early cost-effectiveness

modelling should be done alongside traditional return on investment analysis¹¹⁰ is exactly analogous to the idea that early modelling should be undertaken on government-sponsored health technology assessment in conjunction with assessment of the costs and benefits of the potential outcomes from the research.

Summary

The identified commercial literature indicates that the private sector has addressed the issue of the efficiency of conducting further research, and it is likely that significantly more work in this area has been undertaken, but is not in the public domain.

The general approaches described within the identified studies are all variations on the cost–benefit approach to the evaluation of potential trials, which raise similar practical issues to those described in Chapter 5. The main contribution to this area is provided by Senn,¹⁰⁹ through his explicit recognition and assessment of the value of splitting the research process into separate stages (although he does not describe how the probability of success for the sequential trials could be estimated). Thompson¹⁰³ estimated the value of splitting a single trial into separate periods, while Senn¹⁰⁹ proposes the implementation of the cost–benefit approach at the start of the R&D process and defining separate probabilities of success at each distinct development stage.

Mathematical modelling of underlying biological processes

A systematic examination of the use of mathematical modelling in the context of basic science research is explicitly excluded from this systematic review. However, in the course of the review, an example of the potential power of modelling in this area is illustrated through a case study, and the links between this approach and decision analytic modelling are considered.

For many years, mathematical modelling has been used to aid understanding of many disease processes and biological responses. Such modelling is often undertaken in the development of new pharmacotherapies, with complex mathematical models of pharmacokinetics and pharmacodynamics becoming a more common feature of the drug development process.

There is a distinct body of literature dealing with applications and methodological developments in this field, and the interested reader is directed to journals such as *Mathematical Biosciences* and *Journal of Mathematics in Biology*.

Illustration of the role of biological processes modelling: case studies in wound healing

There have been several mathematical studies in the area of wound healing, an area that illustrates some key features of the value of the approach (although wound healing is not unique in this regard).

Gaffney¹¹⁸ reviews the role of mathematical modelling in the development of quantitative understanding of the wound-healing response and the impact of the modelling work on trial priorities and design. Gaffney discusses earlier modelling studies covering wound healing in general, and corneal epithelial wound healing in particular.

Simplified models of the cell kinetics and dynamics of a wound and the wound-healing process are developed using observational data from experimental wound-healing studies in the sequence of papers above. The core of these models is a mathematical description of factors affecting the progress of wound healing, including:

- cell density, which varies across the wound and throughout the time of wound healing
- cell migration and attrition rates
- concentration of growth factors
- mitosis rates (the rate of cell division which depends on both the cell density and the concentration of chemical stimulants for mitosis or growth factors).

The progress of these factors over the wound-healing period is inter-related and, when described mathematically, gives a system of coupled, non-linear partial differential equations that describe the wound-healing process. The equations can be solved (either analytically or numerically) to give predictions concerning observable outcomes.

The model predictions can be validated against observed phenomena. The results of this validation process inform future research direction. Validation confirming the model generates confidence in its continued use. Where discrepancies are found, they can be used as the

starting point for refining the model, either incrementally or through reviewing the basic assumptions of the model.

The following provides an illustration of the power of this iterative validation approach to inform future research. Dale and co-workers¹¹⁹ produced a model of wound healing and were able to validate the predicted speed of healing under normal conditions. This model was then used to make predictions of healing rates under different concentrations of topically applied growth factor. This acted as a basis for designing further trials to test the impact of topically applied growth factor and gave a rationale for defining experimental dosage rates. Later, Gaffney and colleagues¹¹⁸ reviewed the Dale model,¹¹⁹ particularly the relationship between the mitotic rate and the distance from the centre of the cornea. They found the predicted increase in the mitotic rate towards the centre of the cornea to be counter to newly observed data. This triggered a review of basic assumptions in the model structure, producing a series of refinements to reflect this increased understanding.

Sensitivity analysis has also been used to identify those parameters to which the healing effectiveness is sensitive. For example, in the model of collagen fibre formation during dermal wound healing, a one-way sensitivity analysis identified two sets of variables as the key parameters.¹²⁰ (These were the natural decay rates of latent transforming growth factor- β , and the rates of activation of procollagen I and III to produce collagen I and III.) The paper found that there were few experimental data for these parameter values and their determination thus became an important experimental goal. The construction of the mathematical model thereby serves to enlighten the understanding of the healing processes, targets the continuing research process and identifies key parameters for further investigation.

Summary

For many years, mathematical modelling has been used in aiding the understanding of many disease processes and biological responses. A systematic examination of the use of mathematical modelling in the context of basic science research is explicitly excluded from this systematic review. However, one example of this approach has been considered to show the potential power of modelling in this area. In these examples, the modelling approach provides a clear mathematical description and understanding of the factors and processes involved in wound healing.

The models described above provide examples of the iterative use of models as new data become available, such that the validation process enables an assessment and refinement of earlier models. The applications also show the value of sensitivity analysis in providing a detailed understanding of the impact of uncertain variables. In these basic science examples, modelling is an integral part of the R&D process, helping to identify important variables and hypotheses for experimentation and providing the tool to synthesise new research results with existing knowledge.

Mathematical models should not be viewed as a direct alternative to economic decision modelling, as the former can generally be used only to describe the relatively short-term efficacy of interventions. Such models are unlikely to provide the scope for assessing the full economic impact of interventions. The use of data produced by models that incorporate knowledge about the biological mechanisms of a disease or condition can usefully inform model parameters for which limited direct data exist.

Pretrial simulation of clinical trials

Hale and co-workers¹²¹ describe the potential impact of pretrial simulations in improving the efficiency of the drug development process, specifically through enabling a better understanding of the way in which alternative study designs and assumptions affect study outcomes. The proposed use of simulation to inform trial design is analogous to the application of simulation in other technology-based industries (e.g. electronics and aerospace), whereby the robustness of a specified trial design can be evaluated under a series of 'what if?' scenarios. The models can incorporate data describing the relationships between pharmacokinetic population parameters, pharmacodynamic end-points, enrolment and dropout rates, compliance rates, target population characteristics, etc. Hale and colleagues¹²¹ also state that the output from the simulations can be analysed to provide an understanding of the statistical properties of the proposed study.

Krall and co-workers¹²² present work in progress relating to the development of a central nervous system (CNS) drug, with the aim of understanding the application of simulation to the drug development process. The first step requires an understanding of dose-concentration relationship

over time, and its link with an appropriate clinical outcome (which can in turn be extrapolated to an economic outcome). They analyse data from a Phase I and II trial to estimate a population pharmacokinetic and pharmacodynamic model of the drug. The 'aptness' of the resulting pharmacokinetic/pharmacodynamic model is tested against Bayesian estimates of drug concentrations and scores on the Brief Psychiatric Rating Scale (the measure of clinical effect).

The second step incorporates the pharmacokinetic/pharmacodynamic model into the structure of a proposed trial design. In the CNS drug example, a Phase III trial has already been implemented, and it is intended to simulate this trial and compare the results of the simulation with the actual trial. The final step in the simulation process involves the repeated analysis of the model, from which the simulated data can be analysed as if from a real trial. Such analyses can inform power calculations, as well as test the robustness of the baseline assumptions.

Various examples of the use of simulation to test different options for the design of clinical trials are cited, such as alternative dose-ranging designs, concentration-controlled designs and effect-controlled designs.¹²¹ At present, however, the value of clinical trial simulation is unproven. At a minimum, Krall and colleagues¹²² believe that the simulation process proves the value of being explicit about the assumptions made regarding trial design. A key area in the acceptance of simulation as a trial design tool is the development of a rigorous validation framework, which would build confidence in simulation predictions. As for any proposed tool for informing trial design, it may be possible to evaluate the prospective use of simulation in cases

where simulation models are developed before the start of a trial, but where the trial is not informed by the simulation. The simulation-informed trial design could then be evaluated in the light of any deficiencies noted in the applied trial design.

Summary

Simulation models described in the clinical trial simulation literature synthesise mathematical models of underlying biological processes with assumptions regarding the use of potential drugs in a clinical trial setting (e.g. the target population and compliance rates). The same models could be adapted to accommodate economic outcomes by extrapolating from the specified clinical endpoints, and hence to inform either the cost-benefit or the EVI approach to trial design and prioritisation.

The simulation models replicate patient-level data from the simulated trial, which enables the estimation of confidence intervals around every possible trial result. The usefulness of such detailed predictions in estimating the economic impact of a trial is dependent on the assumed decision rule regarding the uptake of interventions based on the trial results. If the assumed decision rule is to allocate the intervention with highest expected net benefits (as in the applied EVI case studies), then the additional information does not add to the analysis.

If the decision rule is dependent on the new intervention achieving a minimum clinical difference at a certain level of significance (as has been assumed in some applications of the cost-benefit framework), then the use of patient-level simulation models may provide a more accurate assessment of the impact of a proposed trial.

Chapter 10

Conclusions and recommendations

Objectives

The aim of this review has been to clarify the potential role of modelling in planning and prioritising trials and other forms of primary research studies to support the process of health technology assessment. Health technology assessments generally examine both the clinical and cost-effectiveness of alternative interventions and are aimed at answering not whether an intervention works at all, but rather whether it should be used in practice. Funds for the evaluation of healthcare technologies are limited and programmes of research should be developed on the basis of the relative value of alternative evaluations. The original call for proposals identified five questions:

- In what ways can modelling extend the validity of trials (e.g. through adding to their generalisability)?
- What characteristics of the trial/technology affect the success of modelling?
- What aspects of trial design can modelling feasibly inform?
- How feasible, costly and beneficial might modelling be as part of the prioritisation process?
- How far can modelling substitute for low-priority trials?

This chapter brings together the discussion of the issues raised throughout the report and each of the above questions is addressed.

What do we mean by ‘modelling’?

This review has focused on mathematical modelling in health technology assessment and specifically on decision-analytic modelling; that is, modelling the choice between alternative health technologies.

Most decision-analytic models in health technology assessment describe pathways through health states and events that the population of interest can experience. Thus, mathematical models are used to describe the natural history of a disease and how that natural history is affected by the technologies under assessment. The models

are used to estimate health outcomes, resource usage and costs. Thus, models can be used to compare the clinical and economic effectiveness of the competing technologies.

Decision analysis is used in analysing such disease models. Decision analysis can be used in addressing commissioning decisions, such as which of two competing technologies should be used in practice. Furthermore, decision analysis techniques can be used to investigate the uncertainty underlying such a decision.

Many mathematical techniques are available and used in modelling for health technology assessment. The three most common are: decision analysis, state transition or Markov models (Markov chains and Markov processes) and DES. These techniques are not mutually exclusive; indeed, within one technology assessment it is possible to use decision analysis to evaluate a Markov disease model implemented with a DES. The appropriate technique depends on the characteristics of the treatment under evaluation.

Obtaining the relevant literature

The search strategy for the review included 12 electronic databases, citation pearl growing, handsearching of relevant journals, and direct contact with experts in 258 international and UK health technology assessment agencies. Although 601 full papers were reviewed, only 71 of these were directly relevant (either methodological studies or case studies) to the topic of modelling to inform the planning and prioritisation of clinical research.

The review also provided a valuable opportunity to inform the conduct of systematic reviews of methodology topics. The principal conclusions were as follows.

- Methodological reviews should undertake an iterative, systematic approach to literature searching, in particular incorporating citation pearl growing, rather than relying on traditional indexing methods of medical literature searching.

- Structured abstracts in methodology papers are currently rare but should be encouraged. Indexing for modelling methods should be improved.
- Research into structured methods for methodological reviews should be explored, potentially drawing from meta-ethnographic methods developed in the social sciences.

For more detail the interested reader should refer to Chapter 2.

The role of modelling in health technology assessment

Chapter 3 examines literature reviewing the use of modelling within the health technology assessment process in general. The conclusions and recommendations drawn from this chapter as they relate to the specific research questions are discussed below.

General benefits of modelling in health technology assessment

The review has systematically examined the discussion of the pros and cons of modelling. The conclusions are summarised in that models provide:

- a formal structure for addressing a decision problem
- a method for synthesis of evidence from a wide range of sources
- a framework enabling assumptions and data sources to be made explicit, transparent and thus open to debate
- the ability to explore the sensitivity of the results and recommendations to underlying uncertainty and to variations in assumptions.

In what ways can modelling extend the validity of trials (e.g. through adding to their generalisability)?

The conclusions show that modelling enables generalisability in several ways and the following roles were found to be commonplace:

- generalisation from specific trial populations to the full target group for an intervention and vice versa
- generalisation to other settings and countries
- extrapolation of trial outcomes to the longer term
- extension of intermediate end-points, such as reductions in cholesterol levels, to final outcomes, such as coronary events and mortality

- extension of the analysis to the relevant comparators, by incorporating data from external sources so that the evaluation analyses the full lifetime effects of both the intervention of interest and its comparators in practice
- extrapolation to extend survival curves beyond the time horizon of experimental or observational research
- adjusting for prognostic factors in trials
- synthesis of primary research results; for example, by combining trials (including meta-analysis), observational studies and routine resource use data.

What characteristics of the trial/technology affect the success of modelling?

The conclusions from the review do not suggest that there are any particular characteristics of a trial/technology that affect the success of modelling. Modelling is applied successfully to the full range of health technologies. Given the necessary level of analytical expertise, there are no theoretical distinctions between alternative disease areas. However, there are three related questions.

What characteristics of the trial/technology generate a greater need for modelling?

The review clearly identifies certain forms of health technology for which modelling may offer greater benefits as an evaluative tool. These include screening and diagnostics and other areas where typically the impact of the technology occurs over a particularly long duration or where key disease/technology characteristics (e.g. precancerous growth rates) are not directly observable. Long lead times for assessments can mean that a technology may become obsolete before the completion of long-term research and modelling the long-term impact of the technology is a necessity. Similarly, modelling is also of particular benefit for rapidly changing technologies.

What characteristics of the trial/technology generate a need for more complex modelling?

More complex models may be required for technologies with complex patient pathways, large numbers of different associated health events, many comparators and a large but disparate evidence base for the effectiveness of existing technologies and their cost in practice.

What characteristics of the evidence base for a technology affect the success of modelling?

If a new technology has only an emerging evidence base, such as a small case series or even subjective judgements (as opposed to large RCTs),

then a model-based preliminary evaluation is less likely to be considered as a valid method to inform adoption of the technology in practice. This issue has a different impact when the modelling is used to inform the design of further primary data collection; that is, future research. Quantifying the uncertainty in the evidence base is then crucially important in deciding whether and what further research is needed.

Similarly, if for a class of technologies there is significant uncertainty in linking surrogate and final end-points, this may well lead to differential uptake rates of modelling analyses by decision-makers considering adoption of the technology in practice. Again, this example has a different kind of importance when the modelling is used to inform the design of further research. The modelling may well identify further research on the relationship between intermediate and final end-points as crucial, rather than further short-term comparative trials measuring only the intermediate end-point.

The principal conclusion is that a limited evidence base will reduce the success of modelling, if the criterion for success is the usefulness of the model in deciding on the adoption of the technology in practice. However, if, as is the case in the context of this review, the criterion for the model's success is its usefulness in helping to decide on further research, then a limited evidence base is a fact of life and provides the key source material to describe the current uncertainty.

What aspects of trial design can modelling feasibly inform?

The review concludes that standard cost-effectiveness modelling and sensitivity analysis can inform four aspects of research design in particular:

- **Identifying key parameters for further investigation:** once a cost-effectiveness model has been developed, one-way sensitivity analysis and scenario sensitivity analyses indicate the importance of input parameters by examining the stability of the model results when the value of the parameter is varied across its plausible range. These approaches, however, can overestimate the sensitivity of model results and give misleading rankings for key parameters.
- **Specifying the minimum clinical difference required for sample size calculations for a proposed trial:** threshold analysis can be used to determine the minimum clinical difference for a technology to be economically acceptable and

then traditional statistical methods can be used to calculate the implied study sample size.

- **Deciding on the required duration of a proposed trial:** the estimation of the threshold clinical difference can also influence the duration of a trial because it has implications for the period required to observe the necessary differences. Alternatively, if a validated model enables the confident extrapolation of surrogate end-points to final end-points then the duration of a trial may be reduced.
- **Defining the population characteristics for a proposed trial:** models can also be used to explore potential differential effects of a new technology across different population subgroups and thus to define target population groups for a trial.

Modelling in health technology assessment: indications of a role in the planning and prioritisation of trials

There exist many applications of health economic modelling within the technology assessment literature. A subset of these explicitly claims value in informing research design and prioritisation, and this role is supported within the methodological literature. The majority of these case studies use one-way, multiway or threshold analysis towards this aim. However, a comparative methodological study has identified crucial weaknesses in these methods when applied for these purposes. Further analytical methods specifically focusing on trial design and prioritisation are therefore required. These methods are the subject of Chapters 6 and 7, and conclusions and recommendations are detailed below.

Good practice guidelines and critical appraisal of modelling in health technology assessment

One of the major criticisms of modelling studies has been the lack of transparency in many peer-reviewed publications concerning models. This, together with the potential for bias, intended or accidental, in the model development process, has led to much debate over the value of modelling studies. The ability to be able to distinguish 'good' modelling from 'bad' is crucial to all of the research questions under the scope of this review. In particular, however, the importance of robust and valid modelling techniques is fundamental to the last two questions: "*how far can modelling substitute for low-priority trials?*" and "*in what ways can modelling extend the validity of trials?*" In terms of the question "how far can modelling substitute

for low-priority trials?”, from a decision modelling perspective one can define a low-priority trial as one for which the research results would have a low likelihood of affecting the specific decision problem. Similarly, the question “in what ways can modelling extend the validity of trials” explicitly asks how far modelling can be valid in extending and generalising trial results. The reliance that can be placed on the results from modelling studies is clearly a fundamental issue in its acceptability as a basis for decision-making in these areas.

- There is considerable consistency between all the identified studies that have published guidelines for critically appraising modelling studies in technology assessment. This is especially noteworthy considering the time lapse of 15 years between the early Eddy study¹ and the consensus statement formulated during the Consensus Conference on Guidelines on Economic Modelling in Health Technology Assessment.⁷⁴
- All of the guidelines recognise the problems of generalisability when considering critical appraisal of modelling studies. They refrain from setting down detailed prescriptive checklists such as may be found in checklists for reporting statistical meta-analyses or clinical trials. Their approach is to identify a set of principles of good practice in undertaking and reporting modelling studies.
- A wide range of modelling methodologies may be appropriate under different circumstances to address different problems. It is not feasible to construct a prescriptive all-purpose toolkit that strictly defines appropriate methodologies for use in any given circumstances, although Sonnenberg and co-workers³⁶ have defined general conditions where decision trees, Markov models and simulation models may be appropriate.
- In the area of sensitivity analysis, much methodological development has been undertaken and the use of stochastic (probabilistic) sensitivity analysis is gaining popularity. However, no consensus on a prescriptive approach to sensitivity analysis has yet been reached. The recent HTA review presents an excellent starting guide,⁸² although it is not comprehensive.
- Eddy’s four levels of validation provide the benchmark for validation of modelling studies.¹ The four levels are:
 - expert concurrence
 - internal validity
 - predictions agree with non-source data
 - predict–experiment–compare.
 Again, however, the precise implementation of

these validation levels is not defined because of differences in the context, policy question and data availability for alternative models.

- The focus of the critical appraisal guidelines is towards transparency and explicitness in reporting of modelling studies. There are specific domains where this transparency is essential, in particular:
 - the modelling methodology used
 - the structure of the model
 - the sources of data, including subjective judgement, used to populate the model
 - validation of the model
 - analysis of uncertainty or sensitivity analysis of key outcomes.
- The need for transparency and explicitness may be compromised by the space limitations of published articles. There are often too few words, tables and figures to present a sufficiently complete picture of a typical modelling study. It is recommended that sufficient information to support peer review should be made available to reviewers. If necessary, this should be included in a supplementary report or technical appendix.
- In reviewing a modelling study it is necessary to review both the technical application of the modelling methods used and how well the underlying structure of the model reflects or incorporates known disease- or technology-specific factors. For this reason it is necessary to have both clinical and modelling input into the peer-review process and advisable for this to be coordinated.

The current place of modelling in research prioritisation

Chapter 5 reviews the literature on approaches to research prioritisation, with a specific focus on the use of quantitative modelling. The review, reinforced by a survey of health technology assessment organisations, suggests that there is little use of formal modelling in current health technology assessment prioritisation processes. There has also been very little applied research examining the feasibility and value of implementing mathematical modelling within a health technology assessment prioritisation process.

There have been several attempts to produce quantified scoring systems for weighting sets of subjective and quantitative criteria to assess the priority of research studies. These scoring systems would not come under the definition of models

considered by this review and are essentially subjective or arbitrary in the weights attached to the criteria involved. In general, prioritisation is based on the informed views of expert panels of decision-makers.

There is, however, general agreement on the criteria for prioritising research projects. These criteria can be reduced to three broad questions:

- How big is the problem?
- How likely is the assessment to make a difference?
- Are there any ethical, legal and social issues?

Modelling approaches would generally lay claim to informing the first two questions; the third is more likely to be considered separately from clinical and economic components.

Pilot studies have been undertaken to investigate the feasibility and impact of using health economic models in the prioritisation process. The main reasons cited for not using modelling currently are: the resources required to analyse many research projects, difficulties in obtaining reliable data to populate a model, the potential for spurious results given that assumptions will need to be made and, finally, difficulties quantifying the likely evolution of a technology. Potential benefits for a modelling approach are cited in terms of explicitness of the prioritisation process and improved decision-making.

Modelling the likely consequences of proposed research: the cost–benefit or ‘payback’ approach

A summary of the payback methodology

Chapter 6 reviews methodological and case studies, which attempt to analyse the payback or costs and benefits of a specific trial. Some of these studies are prospective, in that the trial has not yet been done, whereas others are retrospective, where the actual impact of the real trial results is being evaluated to assess its payback.

The review concludes that proponents of this approach are in general agreement on the framework required.

This framework requires the comparison of the costs and benefits of undertaking a predesigned research project with the costs and benefit of not

undertaking further research. There is broad agreement on the general approach to estimating the likely consequences of research:

1. List the possible results of the proposed technology assessment.
2. Attach a prior estimate of the likelihood of each possible research result.
3. For each possible research result estimate the uptake of the intervention of interest and of the competing interventions; that is, the pattern of resource allocation between the interventions.
4. Quantify the health benefits and cost consequences of each of these resource allocation patterns.

The estimation of the consequences of no further research involves only steps 3 and 4; that is, predict the future resource allocation between the competing interventions in the event of no further research, and then describe the benefit and cost consequences of each such resource allocation pattern.

The approach is decision analytic in the sense that it considers the choice between: (a) do this specified research and (b) do no further research.

The costs of conducting the research itself are clearly included. Also important are the costs of the intervention allocation decisions made in the absence and presence of the proposed research. The benefits of research are defined as the improvements in the health of the relevant patient population owing to better informed resource allocation decisions. In the case studies, these health benefits are generally converted to a monetary value on the basis of the maximum valuation of the unit of health benefit (e.g. a QALY), which enables the estimation of the net benefits (benefits minus costs) of the research.

Advantages and limitations of the payback methodology

The review concludes that the advantages of the approach include the following.

- It is an intuitive approach; it answers the question one would expect to ask regarding the value of further research: do the expected benefits of the research outweigh the expected costs?
- It is also intuitive in its application, in that the results of the research are estimated, followed by analyses of the implications of the alternative research results.

- The approach also has the advantage that it has been piloted in a series of case studies and in particular in a real research prioritisation context, which found that the approach could be operationalised for approximately 70% of 25 topics considered.

However, the review also identifies both practical and theoretical problems with the application of the approach.

- First, there are debates concerning the specification of possible outcomes of a trial. The simplest case studies, including the pilot discussed above, specify just two possible results: the new intervention is either proved (cost-) effective or not. Other studies allow for the possibility of an inconclusive result, while it is also an option to consider alternative results in different subgroups. Different specifications can and do give different results for the value of the trial.
- Secondly, there are practical concerns with specifying the probability of each possible research result occurring. In the identified case studies, the sources for such estimates ranged from the use of arbitrary probabilities (e.g. 50% success, 50% failure), through published probabilities of success across a sample of clinical trials, to non-specified sources. None of the case studies identified based their estimation of the prospective trial results on data specific to the trial being assessed. This aspect of current practice in implementing the approach is the most likely to invalidate the outcome of the analysis. It should rarely be the case that analysts are completely uninformed about the likely outcome of a trial, as data are usually available from pretrial research phases, and a well-designed (frequentist or Bayesian) trial should be informed by sample size calculations that include estimates of the anticipated effectiveness of new technologies. Such data should always be used to inform the relevant parameter values within the cost–benefit approach.
- Thirdly, methods vary significantly for quantifying the uptake of the alternative technologies. In the case studies, methods range from over-simple (e.g. assuming 100% uptake of trial results), through describing uptake scenarios or undertaking surveys of practitioners on changes in practice as a result of different outcomes from the trial, to much more complex economic models (e.g. specifying demand functions for a drug as a function of

alternative treatment characteristics including trial end-points, costs and a diffusion rate towards a stable demand level).

- Fourthly, the expected lifetime of the technology is difficult to predict. Indeed, only a couple of the applied studies attempted to define this variable, and no details of the assumed time horizon were provided.
- Finally, there is a more fundamental problem, in that this approach estimates the net benefits of a pre-designed research project; a position that implicitly assumes that the proposed research has been optimally designed. The optimality of a research design can only be proven in this framework following the comparison of the costs and benefits associated with a comprehensive set of alternative research designs.

The payback methodology is discussed in relation to the specific research questions below.

What aspects of trial design can modelling feasibly inform?

As noted above, the payback method presupposes a specific trial design; this methodology therefore does not explicitly address this issue. Specific applications, however, have focused on its role in informing the sample size of trials.

How feasible, costly and beneficial might this payback modelling approach be as part of the prioritisation process?

The approach has been attempted on occasions and although it has not achieved a proven track record of success in implementation, it does have potential feasibility. There are no published results concerning the cost of implementing the approach. The benefits are unproven but are conceived by many commentators as increased explicitness of the prioritisation process and improved decision-making.

The issue of whether sensitivity analysis should be applied to this prospective evaluation of a proposed technology assessment is important. Incorporating sensitivity analysis begins to move towards an approach that analyses the uncertainty in the expected benefit and tends towards the EVI approach, which is discussed in the next section.

The primary requirement for further research into payback methods is the implementation of stochastic sensitivity analyses methods within exemplar case studies.

The EVI approach: informing both the design and prioritising of research studies

Summary of EVI methods

Chapter 7 describes the methods for the EVI approach, while Chapter 8 charts the evolution and development of the approach by reviewing relevant methodological papers and case studies.

The review concludes that the rationale for the EVI approach is coherent and that the framework for undertaking an EVI analysis is agreed upon by all authors.

The EVI approach requires the development of an economic model comparing the intervention of interest with its relevant comparators. The variables within the model must be assigned probability distributions to describe their uncertainty. This is followed by an analysis of the expected INB of the interventions [$\lambda * (\text{QALY difference}) - \text{cost difference}$] and selection of the best to make a baseline decision. However, because there is uncertainty in the variables, the method also analyses the probability that the intervention not selected (given current information) could actually be the best. The benefits lost by not selecting the true best intervention given current uncertainties are calculated and known as the expected opportunity loss. Given the expected numbers of people in the system and the likely life span of the interventions, the monetary value of perfect information is calculated. This is measured by the reduction in expected opportunity loss if one had absolute certainty about the value(s) of the parameter(s) concerned. The method can estimate the value of further research overall and on individual parameters.

EVI analyses usually proceed in a sequence as follows.

1. **Overall EVPI:** this answers the question “how valuable would we expect it to be if we could obtain perfectly accurate information on the true value of all of the current uncertain random variables in the decision model?” This is the overall expected value of perfect information.
2. **EVPI for particular parameters or partial EVPI:** this answers the question “how valuable would we expect it to be if we could obtain perfectly accurate information on the exact true value of a specific chosen parameter or set of parameters in the decision model?” This is the

expected value of perfect information for particular parameters.

3. **EVSI for particular parameters:** this answers the question “how valuable would we expect it to be if we could obtain further sample data of a particular size on a specific chosen parameter or set of parameters in the decision model?” This is the expected value of sample information for particular parameters.
4. **ENBS for particular parameters:** this answers the question “what would be the net value of further sample data on a specific chosen parameter?” It is the EVSI for the parameters minus the costs of the data collection. This is the expected net benefit of sample information for particular parameters.

There have been debates on the methods for calculating EVPI for parameters, but these have largely been resolved. Analytical solutions have been formulated for instances where the INB is normally distributed. In the general case, the overall EVPI can be calculated numerically using a one-level Monte Carlo simulation. This requires only a small step from the generation of the CEAC representation of uncertainty and constitutes a minimum recommendation for the presentation of uncertainty in future modelling assessments for health technology assessment prioritisation.

Numerical approaches for estimating the partial EVPI for parameters have been defined and can be implemented in relatively simple economic models. Further research is required in the development of approximation methods, for instance meta-modelling, or efficient sampling algorithms to enable the general application of EVI methods.

Methods for calculating EVSI for parameters are more complicated because they require the revision of prior probability distributions to reflect the likely impact of further data collection of a specified sample size. Existing case studies have made assumptions that the uncertainty in net benefit is normally distributed and that all of the parameters would be measured in the sample data collection. In this restricted case there is an analytical formula to calculate the total EVSI.

In the more general case a two-level Monte Carlo simulation similar to that for EVPI for parameters will be required. The key to revising the prior probability distributions is in estimating the expected posterior variance in the parameters of interest following additional sample information.

This is possible for certain types of distribution. Further methodological research in this area is required.

Case studies using more general methods to calculate EVSI should be a priority for the analytical community.

Advantages and limitations of the EVI methods

In reviewing the literature there is a consensus on the advantages of the EVI approach. It is concluded that the main advantages are as follows.

- It is logical and coherent. The value of undertaking further research is directly related to its impact on technology commissioning decisions and the consequential health and economic benefits.
- Quantifying the value of research overall: the EVI approach quantifies the maximum potential value of research on the parameters concerned. In particular, this allows comparison across disease areas.
- It helps one to decide which specific uncertainties are important and quantifies the uncertainties in absolute rather than relative terms.
- The estimation of the full and partial EVPI has advantages over traditional forms of sensitivity analysis. This is because it answers the question “how likely and with what impact do parameters affect the conclusions of an assessment?”, rather than the simpler question “how much do parameters affect results?” Empirical comparisons find that conventional sensitivity analyses tend to overstate the sensitivity of model outputs to variation in input parameters and can lead to misleading rankings of uncertainties.
- The EVI approach does not require explicit decisions relating to the definition of possible research outcomes, or the prediction of the long-term costs and benefits associated with the defined patterns of uptake of technologies. This is because the underlying tool – a stochastic decision-analytic model that describes the costs and benefits of competing technologies – already incorporates both the possible research results (within the description of probability distributions for uncertain parameters) and the long-term costs and benefits (because they are a part of the cost-effectiveness model on which the EVI analysis is undertaken).
- Selecting research design: the most important advantage of EVI over the payback approach is that it does not start from a prespecified research design, but rather uses an iterative approach examining the decision problem,

identifying key parameters, exploring possible research designs and producing a valuation for each proposed data collection. This enables the research prioritising body to start from the question “what research is most important in this topic area?” rather than “what is the value of this particular randomised clinical trial design?” For example, an RCT is necessary to establish efficacy but may not be necessary to establish the utility associated with different health states.

It is also concluded that various practical issues and problems surround the application of the EVI approach. They include the following.

- The problem of defining how many people will face the choice of strategies modelled is exactly equivalent in the EVI and payback frameworks. The issue of prevalence of the disease area and likely lifetime of the technologies concerned before other options supersede them is an important area. Applied EVI studies often make simple assumptions on these issues.
- Applied EVI approaches commonly use an important and debated assumption with respect to the uptake of the alternative technologies. That assumption is the ‘adoption rule’, which assumes that the intervention with the highest level of expected net benefits will be provided to the whole patient population. Critics say that this is unrealistic and that diffusion of technologies will be dependent on a variety of factors. EVI proponents suggest either that this criticism should be ignored (e.g. just because healthcare organisations may not instantly implement fully rational decisions does not mean that research funding bodies should also be irrational) or that the uptake issues are able to be modelled explicitly within the EVI framework, provided valid relationships or assumptions on uptake can be constructed.
- The value of the EVI methodology hinges on the acceptability of the objective function used for decision-making in health technology assessment. Existing methods and case studies have used a simple INB function incorporating direct healthcare costs and benefits measured with generic quality-of-life instruments. Further research is required on the definition of an objective function that captures adequately the issues of importance to decision-makers in health technology assessment planning and prioritisation, includes the aspects that can be adequately quantified, and can be incorporated into a process that supports the arbitration of subjective judgement.

What aspects of trial design can modelling feasibly inform?

EVI analysis of economics models can be used in the following areas.

Identifying key parameters for further investigation

Partial EVPI analyses can identify key parameters or subsets of parameters that impact on a commissioning decision and can determine an upper limit for research expenditure in these fields. The partial EVSI analyses determine the potential value of further finite sample data on individual or sets of parameters. The ENBS analysis takes into account the costs of research and potential value of future research to determine the technical efficiency of research in different parameter subsets.

It should be noted that the results of these different levels of EVI analyses may give differing results; thus, the ranking of uncertainties from the partial EVPI analysis is likely to be different from the ranking achieved from the ENBS analysis. For example, an EVPI analysis may indicate that perfect information on the comparative efficacy of two interventions had a greater potential value than further information on resource usage, but the high cost of undertaking an RCT to address the efficacy issue compared with a relatively low cost of undertaking further cost analysis may mean that the costing research was more technically efficient.

Establishing an optimal sample size for proposed research

ENBS analysis directly addresses the problem of determining an optimal research sample size based on economic criteria. It is intuitive that increases in sample size have diminishing returns in terms of additional information as the sample size increases; however, the cost of a trial will increase linearly with the size of the trial. The ENBS analysis formalises this intuitive trade-off to establish the economically optimal sample size.

Influencing the required duration of a proposed trial

The focus on undertaking research to identify economically significant differences rather than 'clinically significant' differences is likely to impact on the proposed durations of many clinical trials.

In case studies where surrogate end-points are available that are related to required final end-points, an EVI analysis can determine the relative

value of potentially expensive long-term research, focusing directly on final end-points, in contrast to a coordinated programme of research focusing on identifying the impact of a technology on surrogate end-points together with research on validating the link between surrogates and final end-points. This impacts on the duration of research.

How feasible, costly and beneficial might modelling be as part of the prioritisation process?

The EVI approach to the prioritisation of research is more complex and requires more time and resources than an implicit process. However, the feasibility of undertaking such analyses has been demonstrated within the NHS R&D Programme. The review of liquid cytology screening for cervical cancer undertaken on behalf of NICE,¹⁰⁰ the review of screening for inborn errors of metabolism¹⁰² and an evaluation of kidney preservation systems¹⁰¹ have all demonstrated the feasibility of undertaking this type of analysis within the typical time and resource constraints of an R&D programme. This last case study is of specific interest here, as the rapid review was commissioned at the start of the NCCHTA process for commissioning primary research. Thus, the rapid review with EVI analysis was delivered in time to be used to support the review of detailed proposals for primary research.

The potential benefits of this type of analysis have already been identified. Further research is required to assess their benefits in practice.

How far can modelling substitute for low-priority trials?

Modelling is by no means a substitute for data collection. However, by identifying the absolute and relative value of further research on specific parameters and sets of parameters, EVI analysis directly identifies trial designs of low priority in informing technology commissioning decisions.

The place and feasibility of modelling within health technology assessment prioritisation processes

The following discussion further examines the place and feasibility of modelling within the health technology assessment prioritisation processes.

Shortlisting from hundreds of submitted topics is a significant task and evidence is limited in relation to the value of quantitative prioritisation scoring mechanisms for screening large numbers of potential research projects. It is clear, by common

sense alone, that there is not the capacity to subject literally hundreds of topics to a detailed early economic evaluation incorporating an EVI analysis.

There is a role for the EVI approach within systematic reviews. One important criticism of existing systematic reviews is the lack of rigour and consistency in the sensitivity analysis. The evidence from the review leads to a strong recommendation that systematic reviews should, where possible, incorporate economic evaluation that includes an EVI analysis to assess the importance and sensitivity of different parameters. It is exactly this analysis that can enable the systematic review to form conclusions on future research priorities.

There may also be potential for health technology assessment organisations to utilise modelling at the point of calls for proposals. Topics with a potentially large primary research expenditure are most likely to gain benefit. By commissioning an initial economic evaluation of the intervention concerned, including sensitivity analysis based on the EVI methods, a health technology assessment organisation can establish which are the key uncertain variables and what is the value of information for specific subsets, before investing heavily in primary data collection that may leave some important uncertain variables uninvestigated.

There may also be a possibility to use EVI analyses before the call for proposals, to support the identification of priority research areas. The calculated value of the research can also be compared with its expected costs to show value for money, both internally, and externally to the wider group of NHS, governmental and other stakeholders in society.

Such work would also enable health technology assessment organisations to issue more substantial guidance to researchers on the minimum requirements of research before they submit outline bids.

Recommendations for further R&D in modelling methodology in planning and prioritising trials

Further R&D is required in a number of areas.

- The issues of good practice in the undertaking and reporting of economic modelling studies require further dissemination and support. Areas for development are the handling of stochastic sensitivity analyses, and specifically the CEAC presentation of uncertainty, model validation and the explicit reporting of assumption. The guidelines identified in this report should be recommended to journals that publish economic evaluations to provide a structure for peer review.
- While the potential feasibility of the payback approach to prioritisation of future research is apparent, exemplar case studies using stochastic sensitivity analyses should be developed.
- The calculation of the overall EVPI for a decision problem requires only a very small extension over the generation of a CEAC. This presentation of the overall EVPI should be actively encouraged as a minimum requirement in modelling studies seeking to inform prioritisation and planning of health technology assessment.
- The potential benefits of the EVI analysis have been identified. Further research is required to assess whether these benefits can be realised in the R&D prioritisation and planning process in practice.
- Further research is required on the definition of an objective function that captures adequately the issues of importance to decision-makers in health technology assessment planning and prioritisation, and includes the aspects that can be adequately quantified and can be incorporated into a process that supports the arbitration of subjective judgement.
- Further research is required into the development of approximation methods, for instance meta-modelling or efficient sampling methods, to allow the general application of EVI methods.
- In order to develop a general method for the estimation of the expected value and expected net benefit of sample information, further methodological research is required concerning the updating of prior probability distributions. This should be developed alongside the explicit demonstration of these methods within case study examples.



Acknowledgements

We would like to thank a number of people without whom this report could not have been written. Claire Brooker and Sue James worked a great deal on the literature searching and obtaining articles. Matt Stevenson and Fiona Sampson reviewed over 8000 abstracts to decide on their inclusion in the review. Emma Warren assisted in the editing of this report. Simon Dixon, Jon Nicholl, Mark Sculpher and Roger Beech were involved in the Project Advisory Group providing helpful advice, guidance and encouragement. Elizabeth Fenwick and Karl Claxton were helpful in discussing some issues of methodology in the value of information approach and also providing early copies of papers, which have subsequently

been published. We thank Vanessa Wright, whose patience and hard work and cheerfulness in typing the report and managing the Reference Manager database, kept us on the road to completion. Thanks also to the anonymous referees of the original draft who provided detailed and constructive criticism on the report and enabled us to rewrite it in a more accessible way.

Contributions of the authors

Jim Chilcott, Alan Brennan, Jon Karnon and Paul Tappenden undertook the economics reviewing and methodological development and Andrew Booth was responsible for the information resources and methods.



References

1. Eddy DM. Technology assessment: the role of mathematical modeling. Assessing medical technology. Washington, DC: National Academy Press; 1985.
2. Department of Health. Research for health. London: Department of Health; 1993. pp. 1–36.
3. Henshall C, Oortwijn W, Stevens A, Granados A, Banta D. Priority setting for health technology assessment. Theoretical considerations and practical approaches. Priority setting subgroup of the EUR-ASSESS Project [review]. *Int J Technol Assess Health Care* 1997;**13**:144–85.
4. Eddy DM. Selecting technologies for assessment. *Int J Technol Assess Health Care* 1989;**5**:485–501.
5. Raiffa H. Decision analysis: introductory lectures on choice under uncertainty. Reading, MA: Addison-Wesley; 1968.
6. Black N. Health services research methods: a guide to best practice. London: BMJ Books; 1999.
7. Jadad AR, Moher D, Klassen T. Guides for reading and interpreting systematic reviews: II. How did the authors find the studies and assess their quality? *Arch Pediatr Adolesc Med* 1998;**152**:812–17.
8. Singh R, Singh S. Research and doctors. *Lancet* 1994;**344**(8921):546.
9. Zielinski C. New equities of information in an electronic age. *BMJ* 1995;**312**:1480–1.
10. Gregoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a tower of Babel bias? *J Clin Epidemiol* 1995;**48**:159–63.
11. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, LeLorier J, *et al.* Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;**347**:363–6.
12. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989;**81**:107–15.
13. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias in clinical trials. *Control Clin Trials* 1987;**8**:343–53.
14. Egger M, Davey Smith G. Bias in location and selection of studies. *BMJ* 1998;**316**:61–6.
15. Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple graphical test. *BMJ* 1997;**315**:629–34.
16. Gotzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;**195**(6599):654–6.
17. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. In Chalmers I, Altman DG, editors. Systematic reviews. London: BMJ Publishing Group; 1995. pp. 17–36.
18. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med* 1997;**127**:380–7.
19. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;**16**:62–73.
20. The Cochrane Collaboration. Cochrane methodology database. Oxford: Cochrane Library; 2002.
21. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. *Health Technology Assessment* 1998;**2**(19).
22. Townsend J, Buxton M. Cost effectiveness scenario analysis for a proposed trial of hormone replacement therapy. *Health Policy* 1997;**39**:181–94.
23. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess* 1999;**3**(9).
24. Clarke M, Oxman AD, eds. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software. Updated quarterly.
25. NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews. CRD Report 4. York: NHS Centre for Reviews and Dissemination, University of York; 1996.
26. Buntrock-Robert E. The effect of the searching environment on search performance. *Online* 1979;**3**(4):10–3.
27. Pao-M-L. Term and citation retrieval: a field study. *Information Processing and Management* 1993;**29**:95–112.
28. Yoon-L-L. The performance of cited references as an approach to information retrieval. *Journal of the American Society for Information Science* 1994;**45**:287–99.

29. Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of published controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989;**84**:373–80.
30. Roberts I, Schierhout G. The private life of systematic reviews. *BMJ* 1997;**315**:686–7.
31. McGrath J, Davies G, Soares K. Writing to authors of systematic reviews elicited further data in 17% of cases. *BMJ* 1998;**316**:631.
32. Weinstein MC, Fineberg HV. Clinical decision analysis. Philadelphia, PA: WB Saunders; 1980.
33. Kassirer JP, Moskowitz AJ, Lau J, Pauker SG. Decision analysis: a progress report. *Ann Intern Med* 1987;**106**:275–91.
34. Pauker SG, Kassirer JP. Decision analysis. *N Eng J Med* 1987;**316**:250–8.
35. Thornton JG, Lilford RJ, Johnson N. Decision analysis in medicine. *BMJ* 1992;**304**:1099–103.
36. Sonnenberg FA, Roberts MS, Tsevat J, Wong JB, Barry M, Kent DL. Toward a peer review process for medical decision analysis models. *Med Care* 1994;**32**:js52–64.
37. Beck JR, Pauker SG. The Markov process in medical prognosis. *Med Decis Making* 1983;**3**:419–58.
38. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics* 1998;**13**:397–409.
39. Jain S. Markov chain model and its application. *Computers and Biomedical Research* 1986;**19**:374–8.
40. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Making* 1993;**13**:322–38.
41. Pidd M. Computer simulation in management science. Chichester: John Wiley; 1998.
42. Towse A, Drummond M. The pros and cons of modelling in economic evaluation. *OHE Briefing* 1997;**33**:1–12.
43. Kassirer JP, Angell M. The journal's policy on cost-effectiveness analyses. *N Eng J Med* 1994;**331**:669–70.
44. Luce BR. Policy implications of modeling the cost-effectiveness of health care technologies. *Drug Information Journal* 1995;**29**:1469–75.
45. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, *et al.* Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997;**6**:217–27.
46. Sheldon TA. Problems of using modelling in the economic evaluation of health care [editorial]. *Health Econ* 1996;**5**:1–11.
47. Schulman KA, Yabroff KR, Glick H. A health services approach for the evaluation of innovative pharmaceutical and biotechnology products. *Drug Information Journal* 1995;**29**:1405–14.
48. Rittenhouse B. Uses of models in economic evaluations of medicines and other health technologies. London: Office of Health Economics; 1996.
49. Shahani A. Commentary: Models can be powerful tools for making decisions about effective and efficient health care. *BMJ* 1996;**313**:1057.
50. Rittenhouse BE, O'Brien BJ. Threats to the validity of pharmacoeconomic analyses based on clinical trial data. In Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. Philadelphia: Lippincott-Raven; 1996. pp. 1215–23.
51. Drummond MF, Bloom BS, Carrin G, Hillman AL, Hutchings HC, Knill-Jones R, *et al.* Issues in the cross-national assessment of health technology. *Int J Technol Assess Health Care* 1992;**8**:671–82.
52. Hillman AL, Bloom BS. Economic effects of prophylactic use of misoprostol to prevent gastric ulcer in patients taking nonsteroidal anti-inflammatory drugs. *Arch Intern Med* 1989;**149**:2061–5.
53. Rittenhouse BE. The relevance of searching for effects under a clinical-trial lamppost: a key issue. *Med Decis Making* 1995;**15**:348–57.
54. Eddy DM. The confidence profile method: a Bayesian method for assessing health technologies. *Operations Research* 1989;**37**:210–28.
55. Eddy D. Meta analysis by the confidence profile method. *The statistical synthesis of evidence*. Boston: Academic Press. 1992. pp. 313–22.
56. Eddy DM, Hasselblad V, Shachter R. A Bayesian method for synthesizing evidence: the confidence profile method. *Int J Technol Assess Health Care* 1990;**6**:31–55.
57. Eddy DM, Hasselblad V, Shachter R. An introduction to a Bayesian method for meta-analysis: the confidence profile method. *Med Decis Making* 1990;**10**:15–23.
58. Morris S. A comparison of economic modelling and clinical trials in the economic evaluation of cholesterol-modifying pharmacotherapy. *Health Econ* 1997;**6**:589–601.
59. Mark DB, Hlatky MA, Califf RM, Naylor CD, Lee KL, Armstrong PW, *et al.* Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction. *N Eng J Med* 1995;**332**:1418–24.
60. Schulman KA, Glick HA, Rubin H, Eisenberg JM. Cost-effectiveness of HA-IA monoclonal antibody for Gram-negative sepsis: economic assessment of a new therapeutic agent. *JAMA* 1991;**266**:3466–71.

61. Oster G, Epstein AM. Cost-effectiveness of antihyperlipemic therapy in the prevention of coronary heart disease: the case of cholestyramine. *JAMA* 1987;**258**:2381–7.
62. Geelhoed E, Harris A, Prince R. Cost-effectiveness analysis of hormone replacement therapy and lifestyle intervention for hip fracture. *Aust J Public Health* 1994;**18**:153–60.
63. Stein K. Donepezil in the treatment of mild and moderate senile dementia of the Alzheimer type (SDAT). Development and Evaluation Committee Report. Wessex Institute for Health Research and Development. Bristol: NHS Executive South and West. 1997.
64. Government Statistical Service. Review of the Registrar General on Deaths in England and Wales 1993–95: Mortality Statistics – General. Series DH1(No. 28), Table 48. London: Stationery Office, Office for National Statistics; 1993. p. 221.
65. Beck JR, Pauker SG, Gottlieb JE, Klein K, Kassirer JP, *et al.* A convenient approximation of life expectancy (the 'DEALE'). II. Use in medical decision making. *Am J Med* 1982;**73**:889–97.
66. Pickin DM, McCabe CJ, Ramsay LE, Payne N, Haq IU, Yeo WW, *et al.* Cost effectiveness of HMG-CoA reductase inhibitor (statin) treatment related to the risk of coronary heart disease and cost of drug treatment. *Heart* 1999;**82**:325–32.
67. Gelber RD, Goldhirsch A, Cole BF. Parametric extrapolation of survival estimates with applications to quality-of-life evaluation of treatments. *Control Clin Trials* 1993;**14**:485–99.
68. Messori A. Survival curve fitting using the gompertz function: a methodology for conducting cost-effectiveness analyses on mortality data. *Comput Methods Programs Biomed* 1997;**52**:157–64.
69. Felli C, Hazen GB. Sensitivity analysis and the expected value of perfect information. *Med Decis Making* 1998;**18**:95–109.
70. Halpern MT, Luce BR, Brown RE, Geneste B. Health and economic outcomes modelling practices: a suggested framework. *Value Health* 1998;**1**:131–47.
71. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* 1974;**185**:1124–31.
72. Kassirer JP. The principles of clinical decision making: an introduction to decision analysis. *Yale J Biol Med* 1976;**49**:149–64.
73. McNeil BJ, Pauker SG. Decision analysis for public health: principles and illustrations. *Ann Rev Public Health* 1984;**5**:135–61.
74. Akehurst R, Anderson P, Brazier J, Brennan A, Briggs A, Buxton M, *et al.* Decision analytic modelling in economic evaluation of health technologies. *Pharmacoeconomics* 2000;**17**:443–4.
75. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models: a suggested framework and example of application. *Pharmacoeconomics* 2000;**17**:461–77.
76. McCabe C-J, Dixon S. Testing the validity of cost-effectiveness models. *Pharmacoeconomics* 2000;**17**:501–13.
77. Karnon J, Brown J. Selecting a decision model for economic evaluation: a case study and review. *Health Care Manage Sci* 1998;**1**:133–40.
78. Karnon J. An empirical comparison of the use of Markov processes and discrete event simulation in the economic evaluation of health care technologies. *Health Econ* 2003; in press.
79. Nuijten MJC, Pronk MH, Brorens MJA, Hekster YA, Lockefer JHM, De Smet PAGM, *et al.* Reporting format for economic evaluation part II: focus on modelling studies. *Pharmacoeconomics* 1998;**14**:259–68.
80. Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR, Hadorn DC. Assessing the effectiveness of health interventions. In Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. *Cost-effectiveness in health and medicine*, Oxford: Oxford University Press; 1996. pp. 135–75.
81. Schwartz M. An analysis of the benefits of serial screening for breast cancer based upon a mathematical model of the disease. *Cancer* 1978;**51**:1550–64.
82. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. *BMJ* 1999;**319**:635–8.
83. Donaldson MS, Sox HCJ. Setting priorities for health technology assessment (TA) and reassessment [abstract]. *Abstr Int Soc Technol Assess Health Care* 1992;32.
84. Sculpher M, Drummond M, Buxton M. Economic evaluation in health care research and development: undertake it early and often. Research Report 12. Discussion paper. Uxbridge: Health Economics Research Group, Brunel University; 1995.
85. Davies L, Drummond M, Papanikolaou P. Prioritising investments in Health Technology Assessment: can we assess the potential value for money? Discussion Papers. York: University of York Centre for Health Economics; 1999. 170.
86. Lilford R, Royston G. Decision analysis in the selection, design and application of clinical and health services research. *Journal of Health Services Research and Policy* 1998;**3**:159–66.
87. Harper G, Townsend J, Buxton MJ. The preliminary economic evaluation of health technologies for the prioritization of health technology assessment. *Int J Technol Assess in Health Care* 1998;**14**:652–62.

88. Buxton M, Hanney S. Assessing payback from Department of Health research and development: preliminary report: Vol.1, The main report. Uxbridge: Health Economics Research Group, Brunel University; 1994; 19 Part 1.
89. Detsky AS. Using cost-effectiveness analysis to improve the efficiency of allocating funds to clinical trials. *Stat Med* 1990;**9**:173–84.
90. Drummond MF, Davies LM, Ferris FL III. Assessing the costs and benefits of medical research: the diabetic retinopathy study. *Soc Sci Med* 1992;**34**:973–81.
91. Buxton M, Elliott R, Hanney S, Henkel M, Keen J, Sculpher M, Youll P. Assessing payback from Department of Health research and development: preliminary report, Vol. 2, Eight case studies. Uxbridge: Health Economics Research Group, Brunel University; 1994. 19 Part 2.
92. DiMasi JA, Hansen RW, Grabowski HG, Lasagna L. The cost of innovation in the pharmaceutical industry. Tufts University, Boston, MA: Center for the Study of Drug Development; 1990.
93. Hampton JR. Evidence-based medicine, practice variations and clinical freedom. *J Eval Clin Pract* 1997;**3**:123–31.
94. Lamas GA, Pfeffer MA, Hamm P, Wertheimer J, Rouleau JL, Braunwald E. Do the results of randomised clinical trials of cardiovascular drugs influence medical practice? The SAVE investigators. *N Eng J Med* 1992;**327**:273–4.
95. Romeo AA, Wagner JL, Lee RH. Prospective reimbursement and the diffusion of new technologies in hospitals. *J Health Econ* 1983;**3**:1–24.
96. Warner KE. A 'desperation reaction' model of medical diffusion. *Health Serv Res* 1975;(Winter):369–83.
97. Claxton K, Neumann PJ, Araki S, Weinstein MC. Bayesian value-of-information analysis. An application to a policy model of Alzheimer's disease. *Int J Technol Assess Health Care* 2001; **17**:38–55.
98. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ* 1996;**5**:513–24.
99. Fenwick E, Sculpher M, Claxton K, Briggs A. Improving the efficiency and relevance of health technology assessment: the role of decision analytic modelling. CHE Discussion Paper 179, Centre for Health Economics, University of York; 1999.
100. Payne N, Chilcott JB, McGoogan E. Liquid-based cytology in cervical screening. *Health Technol Assess* 2000;**4**(18).
101. Wight JP, Chilcott JB, Holmes MW, Brewer N. The clinical and cost-effectiveness of pulsatile machine perfusion versus cold storage of kidneys for transplantation retrieved from heart-beating and non-heart-beating donors. *Health Technol Assess* 2003;**7**(25).
102. Pandor A, Eastham J, Beverley C. The clinical and cost-effectiveness of neonatal screening for inborn errors of metabolism using MS/MS (tandem mass) spectrometry. *Health Technol Assess* 2003; in press.
103. Thompson MS. Decision-analytic determination of study size: the case of electronic fetal monitoring. *Med Decis Making* 1981;**1**:165–79.
104. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves [letter]. *Health Econ* 2001;**10**:779–87.
105. Phelps CE, Mushlin AI. Focusing technology assessment using medical decision theory. *Med Decis Making* 1988;**8**:279–89.
106. Karnon J, Brown J. Tamoxifen plus chemotherapy versus tamoxifen alone as adjuvant therapies for node-positive postmenopausal women with early breast cancer: a stochastic economic evaluation. *Pharmacoeconomics* 2002;**20**:119–37.
107. Karnon J. Economic evaluation of health care technologies: a comparison of alternative decision modelling techniques. PhD Thesis; Brunel University, Uxbridge; 2001.
108. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technol Assess* 2000;**4**(38).
109. Senn S. Some statistical issues in project prioritization in the pharmaceutical industry. *Stat Med* 1996;**15**:2689–702.
110. Grabowski H. The effect of pharmacoeconomics on company research and development decisions. *Pharmacoeconomics* 1997;**11**:389–97.
111. Hornberger JC, Brown BW, Halpern J. Designing a cost effective clinical trial. *Stat Med* 1995;**14**:2249–59.
112. Hornberger J, Eghtesady P. The cost benefit of a randomised trial to a health care organisation. *Control Clin Trials* 1998;**19**:198–211.
113. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;**18**:341–64.
114. Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. *J Health Econ* 1996;**15**:641–53.
115. Felli JC, Hazen GB. A Bayesian approach to sensitivity analysis. *Health Econ* 1999;**8**:263–8.

116. Plante DA, Piccirillo JF, Sofferman RA. Decision analysis of treatment options in pyriform sinus cancer. *Med Decis Making* 1987;**7**:74–83.
117. Backhouse ME. An investment appraisal approach to clinical trial design. *Health Econ* 1998;**7**:605–19.
118. Gaffney EA, Maini PK, Sherratt JA, Tuft S. The mathematical modelling of cell kinetics in corneal epithelial wound healing. *J Theor Biol* 1999;**197**:15–40.
119. Dale PD, Maini PK, Sherratt JA. Mathematical modelling of corneal epithelial wound healing. *Math Biosci* 1994;**124**:127–47.
120. Dale PD, Sherratt JA, Maini PK. A mathematical model for collagen fibre formation during foetal and adult dermal wound healing. *Pro R Soc Lond* 1996;**263**:653–60.
121. Hale M, Gillespie WR, Gupta S, Tuk B, Holford N. Clinical trial simulation as a tool for increased drug development efficiency. *Appl Clin Trials* 1996;**5**:35–40.
122. Krall RL, Engleman KH, Ko HC, Peck CC. Clinical trial modelling and simulation – work in progress. *Drug Information Journal* 1998;**32**:971–6.
123. Parsonnet J, Harris RA, Hack HM, Owens DK. Modelling cost-effectiveness of *Helicobacter pylori* screening to prevent gastric cancer: a mandate for clinical trials. *Lancet* 1996;**348**:150–4.
124. Howard S, Mugford M, Normand C, Elbourne D, Grant A, Field D, *et al.* A cost-effectiveness analysis of neonatal ECMO using existing evidence. *Int J Technol Assess Health Care* 1996;**12**:80–92.
125. Cantor SB, Spann SJ, Volk RJ, Cardenas MP, Warren M. Prostate cancer screening: a decision analysis. *J Fam Pract* 1995;**41**:33–41.
126. Briggs A, Sculpher M, Britton A, Murray D, Fitzpatrick R. The cost and benefits of primary total hip replacement. *Int J Technol Assess Health Care* 1998;**14**:743–61.
127. Gillespie WJ, Pekarsky B, O'Connell DL. Evaluation of new technologies for total hip replacement: economic modelling and clinical trials. *J Bone Joint Surg* 1995;**77**:528–33.
128. Laheij RJ, Severens JL, Jansen JBMJ, van der Lisdonk EH, Verbeek ALM. Management in general practise of patients with persistent dyspepsia. *J Clin Gastroenterol* 1997;**25**:563–7.
129. Davey PG, Malek M. Defining criteria for the pharmacoeconomic evaluation of new oral cephalosporins. *Pharmacoeconomics* 1994;**5**:11–19.
130. Sculpher MJ. A preliminary economic evaluation of the diode laser in ophthalmology. *Lasers Med Sci* 1993;**8**:163–9.
131. Sculpher M, Michaels J, McKenna M, Minor J. A cost–utility analysis of laser-assisted angioplasty for peripheral arterial occlusions. *Int J Technol Assess Health Care* 1996;**12**:104–25.
132. Neil N, Sullivan SD, Lessler DS. The economics of treatment for infants with respiratory distress syndrome. *Med Decis Making* 1998;**18**:44–51.
133. Sculpher MJ. Final report: phase II medical laser technology assessment. Research Report 18. Uxbridge: Health Economics Research Group, Brunel University; 1994.

Appendix I

Review of case studies addressing the planning and design of future research

Introduction

Table 11 summarises the case studies referenced in Chapter 3 that explicitly address the role of modelling in trial planning and design. The main characteristics of the case studies that are described include:

- evaluation characteristics and modelling methodologies used
- examples of the roles of modelling applied
- trial design issues addressed.

The remainder of the appendix reviews the case studies in more detail.

Respiratory distress syndrome: the economics of treatment for infants with respiratory distress syndrome¹³²

This paper describes the development of a disease treatment pathway model for infants with respiratory distress syndrome in the USA. Clinical and resource use data were obtained from a panel of expert clinicians, using a modified Delphi technique. The primary role of the simulation model is the estimation of the costs associated with this disease; as such, this represents a burden of disease model. It is recognised, however, that the model supports preliminary evaluations of new technologies in this disease area, whereby the relevant parameters affected by any new technology may be altered to estimate their impact on the burden of the disease. The model allows the key components of care to be identified and thus allows effort to be concentrated on determining the cost associated with these elements rather than undertaking an exhaustive cost search. The methods for doing this are not expanded upon within the paper, although a series of scenario analyses is undertaken.

New oral cephalosporins

Defining criteria for the pharmacoeconomic evaluation of new oral cephalosporins¹²⁹

This paper discusses a series of economic issues related to clinical choices in the use of third-generation cephalosporins, focusing on antibiotic therapy as a whole rather than on any specific disease area. The use of economic evaluations in informing trial design is discussed in relation to the choice between oral and intravenously administered therapy, where the additional cost of intravenous therapy can only be justified by increased efficacy.

The key parameter for the use of the cephalosporins is assumed to be their relative efficacy compared with intravenous delivery; this assumption is justified.

The cephalosporins model combines an estimate of the maximum willingness to pay to avoid a treatment failure (e.g. US\$1000) with the additional costs of the intravenous treatment option (e.g. \$100) to calculate the level of increased efficacy (of the intravenous option) that must be excluded for the oral therapy to be considered cost-effective, which is 10%:

$$\text{Implied marginal value (US\$1000)} = \frac{\text{Incremental cost (US\$100)}}{\text{Incremental efficacy (0.1 or 10%)}}$$

The value of avoiding a treatment failure and the incremental cost of treatment define the minimum clinically important difference in efficacy, and hence the classically defined trial sample size.

Total hip replacement

Evaluation of new technologies for total hip replacement: economic modelling and clinical trials¹²⁷

This study undertakes a preliminary evaluation of new hip prostheses in comparison to conventional hip prostheses. It is noted that many new hip prosthesis designs have been introduced with little

TABLE 11 Summary table of case studies

| Reference | Neil et al. ¹³² | Sculpher et al. ¹³¹ | Sculpher ¹³⁰ | Davey and Malek ¹²⁹ | Laheii et al. ¹²⁸ | Gillespie et al. ¹²⁷ | Briggs et al. ¹²⁶ | Cantor et al. ¹²⁵ | Howard et al. ¹²⁴ | Parsonnet et al. ¹²³ |
|--|-------------------------------|-----------------------------------|-------------------------|-----------------------------------|---------------------------------|------------------------------------|---------------------------------|---------------------------------|---------------------------------|------------------------------------|
| Intervention type | | | | | | | | | | |
| Pharmaceutical | | | | ✓ | ✓ | | | | | ✓ |
| Screening | | | | | | | | ✓ | | ✓ |
| Diagnostic | | | | | | | | | | |
| Surgical | | ✓ | | | | ✓ | ✓ | | | |
| Service delivery & organisation (SDO) | ✓ | | ✓ | | | | | | | |
| New technologies | | ✓ | | | | ✓ | ✓ | | ✓ | |
| Place in R&D cycle | | Post-trial | Pretrial | | Pretrial | Pretrial | Pretrial | Pretrial | Pretrial | Pretrial |
| Modelling technique | | | | | | | | | | |
| Decision analytical | | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ |
| Markov | | ✓ | | | | | ✓ | ✓ | | ✓ |
| Simulation | ✓ | | | | | | | | | |
| Analytical modelling | ✓ | | ✓ | ✓ | | ✓ | | | | |
| Outcomes | | | | | | | | | | |
| Clinical | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Cost | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |
| Health economic | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |
| Roles of modelling | | | | | | | | | | |
| Studying assessment problems that have not been previously addressed | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Synthesising head-to-head comparisons between competing treatments | ✓ | | | | | ✓ | ✓ | | | |
| Generalising between trial and actual treatment populations | | A | | | | | | | | |
| Generalising between different geographical/temporal populations | | | | | | | | | | |

Continued

TABLE 11 Summary table of case studies (cont'd)

| Reference | Neil <i>et al.</i> ¹³² | Sculpher <i>et al.</i> ¹³¹ | Sculpher ¹³⁰ | Davey and Malek ¹²⁹ | Laheii <i>et al.</i> ¹²⁸ | Gillespie <i>et al.</i> ¹²⁷ | Briggs <i>et al.</i> ¹²⁶ | Cantor <i>et al.</i> ¹²⁵ | Howard <i>et al.</i> ¹²⁴ | Parsonnet <i>et al.</i> ¹²³ |
|--|--------------------------------------|--|-------------------------|-----------------------------------|--|---|--|--|--|---|
| Extrapolating from surrogate to final outcome measures | | ✓ | ✓ | | | | | | | ✓ |
| Extrapolating the duration of treatment effects | | ✓ | | | | | | | | ✓ |
| What trial design issues identified | | | | | | | | | | |
| Key parameters for study | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Minimum clinical difference (delta) | | ✓ | | ✓ | | ✓ | ✓ | ✓ | | ✓ |
| Duration of study | | ✓ | | | | ✓ | ✓ | | | |
| Sample size | | | | ✓ | | ✓ | ✓ | | | ✓ |
| Study population age/gender | | | | | | ✓ | ✓ | ✓ | | ✓ |
| Incremental cost of new technology | | ✓ | | | | ✓ | ✓ | | | |
| Further research worthwhile | Yes | Yes | N/A | Yes | Yes | Yes | Yes | | | Yes |
| Type of research | Cohort study | RCT | | Clinical trials | Not specified | RCT | RCT | | | RCT |

A: explicit assumptions made; NA: not applicable.

or no evidence regarding clinical effectiveness in terms of long-term prosthesis survival.

An NPV model of lifetime costs for patients with hip replacement is developed.

$$PV_c = C_c + H + \sum_{i=n}^n \frac{l_{i,m} p_{i,c,m} (C_c + R_c + H)}{(1+r)^{(i-1)}}$$

where PV_c = present value, C_c = purchase cost of prosthesis type C , H = hospital costs associated with fitting the prosthesis, R_c = recuperation costs in addition to H (for prosthesis type C), $l_{i,m}$ = probability of survival to year i , $p_{i,c,m}$ = probability of prosthesis failure in year i .

Cost-effectiveness is dependent solely on the difference between the expected lifetime costs, as morbidity associated with failure of hip prosthesis is not included.

The model aids the design of clinical trials of new prostheses through the prediction of cost or survival limits within defined patient groups outside which a component system will not prove cost-effective. For example:

“A claim for a 20% reduction in prosthetic failure rate at 15 years for a new prosthesis priced at twice the conventional cost would not justify an RCT or licensing of the prosthesis. The same projected reduction in failure rate with a cost of 1.2 times the conventional cost, however, would lead to a trial being considered in men under 65 and women under 55 years of age.”

These limits in the potential improvements in effectiveness define the smallest relevant improvement in outcome that is worth detecting, which can then be used to determine the sample size of trials of new prostheses through the application of standard statistical techniques.

Appropriate age and gender entry criteria are also informed by the preliminary evaluation. The analysis also identifies a minimum duration of a proposed trial of 15 years, since at that point conventional survival is still in the order of 80%. The practical difficulties of running a trial over such a prolonged period led to an assessment of alternative study designs, for example:

- A study could include intermediate study analysis of prosthesis survival, and once again the necessary smallest relevant improvements at these interim points can be used in defining stopping criteria.

- A multicentre design would be indicated in order to achieve rapid recruitment of the large patient numbers implied by the analysis.
- The study identifies the potential usefulness of short-term outcome measures, for example, femoral loosening or acetabular wear, although the validity of these surrogate measures as predictors of long-term survival would need to be established through long-term monitoring.

Costs and benefits of primary total hip replacement: how likely are new prostheses to be cost-effective?¹²⁶

The scope of the study detailed in this paper is similar to that of the previous study,¹²⁷ although it uses a different methodological approach in two aspects.

- Morbidity and mortality associated with hip prosthesis failure is accounted for by the inclusion of health-related quality-of-life measures associated with prosthesis failure.
- A Markov state transition model is used to track lifetime costs and health-related quality-of-life of hip replacement patients.

The paper similarly investigates the potential cost-effectiveness of new hip replacement technologies in different age/gender populations over their full lifetime.

Elasticities, representing the percentage change in the cost-effectiveness of new prostheses in response to a percentage change in the value of a parameter, are estimated for the individual parameters within the model. These analyses show that the two most important parameters in the model are the cost and effectiveness of the prosthesis relative to the standard prosthesis. The majority of the other parameters have low elasticities, indicating that they are unlikely to be key parameters in terms of determining cost-effectiveness.

The required reduction in revision rates for alternative forms of cementless prostheses to be considered cost-effective was also estimated from the sensitivity analyses of the effectiveness parameter. These analyses showed that a trial comparing some of the newer prostheses with the established prostheses may be cost-effective, since reductions in revision risk of around 17% are required to show cost-effectiveness. Such estimates are analogous to the minimum clinically significant difference that can be used to inform sample size calculations.

Medical laser technology assessment

Final report: Phase II medical laser technology assessment¹³³

An iterative framework for technology assessment, equivalent to the four phases typically used in clinical drug development, is described (*Table 12*).

The report summarises three assessments of medical laser technologies that have been characterised as Phase II preliminary economic evaluations. These are:

- a preliminary economic evaluation of the diode laser in ophthalmology
- an economic evaluation of the Nd:YAG laser in gastroenterology
- the role of the laser in percutaneous artery revascularisation.

Two of these studies have given rise to peer-reviewed published articles. The role of modelling in these studies is discussed below.

A preliminary economic evaluation of the diode laser in ophthalmology¹³⁰

The newer diode laser is compared with the existing argon laser used in a range of conditions within ophthalmology. An assumption of equal effectiveness is supported by the substantive equivalence between the two technologies and their role in treatment. However, the authors recognise that small studies have reported greater levels of pain in patients treated with the diode laser than in patients treated with the argon laser.

A simple cost model (details of which are not provided) is used to calculate the expected cost per patient treated for the argon and the diode laser under five different purchasing scenarios.

- routine replacement of an argon laser
- early replacement of an existing argon laser
- establishment of a new ophthalmology facility
- purchasing of a diode laser for specific conditions
- reorganisation of the delivery of ophthalmic care.

The model is used to highlight the determinants of differential cost and investigate the potential scope for decreased costs with the introduction of the diode laser in each scenario. The analyses identify contexts where the diode laser is potentially cost-minimising, the implication being that further trials examining those contexts where there is no potential for the technology to be cost-minimising are of little value.

The cost modelling does not inform specific trial design issues between these technologies because the impact of the uncertainty regarding the relative clinical effectiveness of the two lasers is not explored within the model.

Key parameters are identified as the relative clinical effectiveness of novel and conventional technologies in terms of patient-based rather than technical outcomes, in order to validate the assumption of equivalent effectiveness between the technologies. The importance of the number of treatment sessions required when using the diode laser is also identified as a key parameter through one-way sensitivity analyses.

TABLE 12 Stages of economic evaluation of medical technologies

| Stage of study | Extent of diffusion | Level of clinical evaluation | Level of economic evaluation |
|----------------|-------------------------------|------------------------------|--|
| Phase I | One or two pioneering centres | Case studies | Indicative economic evaluation to identify scope for technology to be cost-effective |
| Phase II | Key clinical centres | Series | Using sample from series and literature data and modelling methods |
| Phase III | Main centres | Comparative trials | Alongside trials or using trial (and other) data in models |
| Phase IV | Poised to diffuse widely | Comparative trials | Synthesis of data within models to assess generalisability of Phase III results |

Source: Sculpher.¹³³

A cost–utility analysis of laser-assisted angioplasty for peripheral arterial occlusions¹³¹

This paper details a cost–utility analysis of the use of laser technology as a secondary adjunct to angioplasty to treat peripheral arterial occlusions. The cost–utility model combines a decision tree framework to assess the options available at the time of angioplasty with a Markov model to extrapolate the short-term outcomes of different angioplasty techniques. Three short-term outcomes from angioplasty are specified: asymptomatic, claudicant, and rest pain/ulceration. Long-term outcomes are defined in terms of QALYs gained.

The probabilities of the different clinical outcomes within the short-term model are obtained from an RCT (which is assumed to be representative of the general population), while some of the transition probabilities within the long-term model are based on expert judgements.

Analysis of the uncertainty within the model is undertaken through a series of one-way sensitivity analyses together with threshold analyses. Three categories of uncertainty are identified within the model:

- generalisability: proportion of patients who after successful recanalisation become asymptomatic
- analytical methods: methods for valuing health states, discount rates
- data variability: proportion of symptomatic patients whose symptom status moves to claudication.

These key determinants are identified through their impact on the key outcome measures when each parameter is varied within its defined range (informed by published data sources and expert judgements).

The minimum clinical difference, in terms of the secondary recanalisation rate necessary to achieve a variety of cost–utility thresholds, is estimated from the model using a series of threshold analyses. The sample size and study populations could also be informed by use of the minimum clinical difference as estimated by the model, although this is not explicitly expanded upon in the paper.

Threshold analysis is also used to determine the equipment utilisation rates, which impact on the incremental cost, under which the novel

technologies dominate, that is, are cheaper and more effective than the conventional techniques.

To investigate the long-term costs and consequences of the novel laser-based technologies through empirical trials, these would require large sample sizes and an extensive follow-up period. The model, by formalising the link between the short-term and long-term outcomes, provides an alternative strategy, whereby short-term empirical studies can be targeted at the key uncertainties in the model-based assessment. The modelling thereby determines the duration of the subsequent trials.

Dyspepsia

Management in general practice of patients with persistent dyspepsia: a decision analysis¹²⁸

The objective of this paper is to examine the potential for an empirical drug treatment strategy within general practice to increase the appropriate referral and use of endoscopy in treating patients with persistent dyspepsia. The outcomes from the model are the percentage of patients undergoing endoscopy, percentage of patients with symptom relief and mean cost per patient.

The structure and parameterisation of a decision tree are informed by reviews of published literature, which identified baseline and ranges of values for model parameters, although details of search methods are not given. The study claims that the empirical treatment strategy dominates the conventional treatment strategy in all scenarios considered, that is it is both cheaper and more clinically effective. The key parameter influencing the outcome measures is identified as the prevalence of the different dyspeptic disorders in the patient population, although it is not clear what type of sensitivity analysis, for example, one-way, multiway or parametric, is undertaken.

Although the authors recommended the need for further research, the relevant health technology assessment agency concluded that the preliminary modelling answered the main questions and refrained from funding further primary research in this area.

Prostate cancer

Prostate cancer screening: a decision analysis¹²⁵

A dearth of RCT evidence is identified for the screening of men for prostate cancer. This,

together with the lack of evidence regarding whether early treatment leads to improvements in either length of survival or quality of life, makes this intervention controversial.

A Markov model is used to assess the likely health impact of such a screening policy in terms of QALYs (costs are not included in this analysis). The structure and parameterisation of the model are derived from published literature (details of the search strategies are not given). Quality-of-life adjustments reflecting the utilities of potential health outcomes and adverse events of treatment were obtained from a time trade-off study of a small subgroup of individuals with no diagnosis of prostate cancer. This study was undertaken by the same study group and is reported fully elsewhere.

The baseline result shows that fewer QALYs are gained by any of the screening options than by the no-screening option. One-way sensitivity analysis is undertaken on all parameters, supported by further two-way analyses for key outcomes of treatment. The baseline result was found to be insensitive to most parameters within the model except for (1) changes in the underlying prevalence of asymptomatic clinically significant prostate cancers and (2) patients' preferences regarding adverse effects of treatment. Both factors are subjected to threshold analyses to determine the critical values that these parameters must attain for screening to be effective. This information would be of use in determining the sizes of any further experimental investigations, although this is not elaborated upon in this paper.

This study identifies the low potential of prostate cancer screening to be an effective treatment in terms of quality-of-life outcomes. It also notes that screening would be likely to increase overall healthcare costs, indicating that no screening would be a dominant policy if health economics were considered. Thus, the collection of further RCT-based evidence in this field should be treated as a low priority.

Neonatal care

A cost-effectiveness analysis of neonatal ECMO using existing evidence¹²⁴

This paper describes the development of a preliminary economic evaluation prepared in support of a planned RCT. This trial investigates the clinical and economic effectiveness of extracorporeal membrane oxygenation (ECMO) in

giving temporary support to patients with severe respiratory failure. The specific aim of this study is to assess the quality of existing evidence on cost and effectiveness, as an aid to identifying what information was required for the economic evaluation in the trial.

The structure of a decision tree describing the alternative paths of cure is derived from the protocol of the proposed trial. Parameterisation of the model used evidence from published sources and routine health-service data. Systematic literature searches were used to identify existing clinical and economic evidence.

The reported sensitivity analysis consists of simple worst and best case scenario analyses, with parameters set to extreme values, although the nature of these extremes is not defined. The results indicate that this new technology is likely to increase the costs of care in both the best and worst case scenarios, but could be either more or less effective. The study concludes that restricting the availability of ECMO pending completion of the prioritised trial is justified.

The sensitivity analysis identifies a wide range of uncertainty in the cost-effectiveness of this treatment, but fails to identify which parameters have the greatest impact, and hence does not allow targeting of further experimental investigation.

Gastric cancer

Modelling the cost-effectiveness of *Helicobacter pylori* screening to prevent gastric cancer: a mandate for clinical trials¹²³

This study reports a preliminary economic evaluation of screening, leading to the eradication of *Helicobacter pylori* infection with a focus on the prevention of gastric cancer. Early evidence suggesting a link between *H. pylori* infection and gastric cancer is identified, which supports this as a potentially effective treatment strategy. However, inadequate experimental data demonstrating that *H. pylori* eradication modifies cancer risk means that further experimental investigation would be required before widespread introduction of this intervention. This study aims to investigate clinical and economic effectiveness, with the objective of assessing the value of further research, as well as to suggest interim treatment strategies while awaiting the results of these trials.

A decision tree is used to compare two potential screening/treatment strategies to a baseline strategy of no screening or eradication. A Markov model is used to assess long-term costs and outcomes associated with the different strategies. The model structure and probabilities were obtained with reference to published literature, although it is unclear whether or not this was obtained through a systematic search. Healthcare costs were obtained from databases of US fees and charges.

The baseline results indicate that the screening and treatment policy would increase healthcare costs but would result in the prevention of cancer incidence and related deaths, with a cost per life year saved of US\$25,000. One-way sensitivity analysis is undertaken for each parameter within the model across its plausible range; this analysis is supported by a best and worst case scenario analysis.

The potential cost-effectiveness of this intervention supports further experimental investigation. This study assumes that the appropriate way forward is the undertaking of a large economic trial to estimate the cost-effectiveness of treatment directly, rather than undertaking smaller substudies aimed at increasing knowledge concerning some of the individual factors in the analysis. The model does not explicitly target key parameters for further investigation.

A role for the model in defining population characteristics for further investigation is claimed within the paper, based on its use in investigating the cost-effectiveness of screening in different risk groups. The study also indicates that the model could assist the sample size calculations of a prospective trial, citing the use of the methodology proposed by Hornberger and co-workers¹¹¹ (see Chapter 7) in this respect.

Appendix 2

Summary of identified modelling guidelines

This appendix presents three sets of published guidelines (*Tables 13–16*) describing good practice for the use, and critical appraisal of decision-analytic modelling as a tool for the assessment of health technologies.

TABLE 13 *Content analysis of modelling guidelines*¹

1. A statement of the problem
 2. A description of the relevant factors and outcomes
 3. A description of the model
 4. A description of data sources (including subjective estimates), with a description of the strengths and weaknesses of each source
 5. A list of assumptions pertaining to
 - (a) the structure of the model (e.g. factors included, relationships and distributions) and
 - (b) the data
 6. A list of parameter values that will be used for a base case analysis, and a list of the ranges in those values that represent appropriate confidence limits and that will be used in a sensitivity analysis
 7. The results derived from applying the model for the base case
 8. The results of the sensitivity analyses
 9. A discussion of how the modelling assumptions may affect the results, indicating both the direction of the bias and the approximate magnitude of the effect
 10. A description of the validation method and results
 11. A description of the settings to which the results of the analysis can be applied and a list of factors that could limit the applicability of the results
 12. A description of research in progress that could yield new data that could alter the results of the analysis
- If the analysis recommends a policy, the report should also contain:
13. A list of the outcomes that required value judgements
 14. A description of the values assessed for those outcomes
 15. A description of the sources of those values
 16. The policy recommendation
 17. A description of the sensitivity of the recommendation to variations in the values
 18. A description of the settings to which the recommendations apply

TABLE 14 *Content analysis of modelling guidelines*⁷⁰

Model approach

1. Study question specified
2. Need for modelling versus alternative methodologies discussed
3. Type of model identified
4. Reason for use of this model type discussed
5. Model scope specified:
 - time-frame
 - perspective
 - comparator(s)
 - setting/country/region
6. Basis of scope discussed

Model specifics

7. Source and strength of model data specified
8. Model assumptions discussed
9. Model parameters available in technical appendix
10. Values and sources for model parameters specified:
 - event probabilities
 - rates of resource utilisation
 - costs
 - health utilities
11. Criteria for evaluating quality of data specified
12. Relevant treatment strategies included
13. Relevant treatment outcomes included
14. Biases discussed and explored

Model analysis

15. Base case results presented and described
16. Sensitivity analyses performed:
 - unidimensional
 - multidimensional
 - best/worst case
 - threshold
17. Key cost drivers identified
18. Verification performed
19. Validation performed

TABLE 15 Content analysis of modelling guidelines³⁶**Principles**

The type of model should reflect the nature of the clinical problems, e.g. if the clinical problem persists over a long period, a Markov model is likely to be the most appropriate type of model

All reasonable treatment options, including extremes such as watchful waiting, should be included in the model

The key characteristics of the disease should be included in the model

The key clinical outcomes should be included in the model

The utility structure should incorporate all relevant attributes

Errors

Model syntax

Conditioning of action on unobservable states

Violations of symmetry in modelling prognosis

Failure to link variables that are inherently related

Inconsistent bias in assumptions

Modelling results of diagnostic tests

Modelling of treatment

Appendix 3

Data extraction sheet for identifying the key roles of modelling

**DATA
EXTRACTION
SHEET 1**

METHODOLOGY: DEVELOPMENT PHASE I

**PUBLICATION
DETAILS**

| | |
|-----------------------|---------------------------------------|
| Author | <i>Lead author</i> |
| Title | <i>Full title</i> |
| Publ. Date | <i>Year</i> |
| Type | <i>e.g. Monograph/Journal article</i> |
| Source | <i>Journal name</i> |
| Ref ID | <i>Reference manager index</i> |
| Classification | <i>M1/M2/M3</i> |

| ISSUE HEADING | DISCUSSION AND QUOTES | PAGE | EXTERNAL REF |
|----------------------|--|-----------------|--|
| NOTES | <i>Free text discussion and interpretation of specific issues regarding the context of the report and key features</i> | <i>Page no.</i> | <i>Referring information to other key articles</i> |
| | | | |
| | | | |
| | | | |

POTENTIAL ROLES OF MODELS IN PRIORITISING AND PLANNING TRIALS

| | | | |
|----------------------------------|---|-----------------|--|
| <i>Free text key issue title</i> | <i>Free text detailed discussion of key issue</i> | <i>Page no.</i> | <i>Referencing information to other key articles</i> |
| | | | |
| | | | |
| | | | |
| | | | |

KEY ISSUES IN GOOD AND BAD PRACTICE

| | | | |
|----------------------------------|---|-----------------|--|
| <i>Free text key issue title</i> | <i>Free text detailed discussion of key issue</i> | <i>Page no.</i> | <i>Referencing information to other key articles</i> |
| | | | |
| | | | |
| | | | |
| | | | |

Appendix 4

Data extraction sheet for discussion on the roles of modelling

| | <i>Reference ID (Author, Year)</i> |
|---|---|
| Notes | |
| | |
| | |
| Generalising between populations (across time and space) | <i>Direct quotations and commentary from the article giving page numbers and external reference</i> |
| Linking information from diverse sources Studying assessment problems that have never been the subject of comprehensive experimental studies (e.g. screening/clinical modelling early studies) | |
| Synthesising head-to-head comparisons | |
| Extrapolating results beyond the trial duration | |
| Extrapolating results from surrogate to final end-points | |
| Analysis of disease dynamics | |
| Description, testing/validation of hypotheses about the natural history of a disease | |
| Value of collecting additional information through research priority setting | |
| Communication/bookkeeping tool | |

Appendix 5

Data extraction sheet for case studies on the role of modelling

| | |
|---------------------------------------|--------------------------------|
| Ref ID | <i>Reference manager index</i> |
| Classification of case study | <i>C1/C2</i> |
| Disease area | <i>Free text description</i> |
| | |
| Intervention type | |
| Pharmaceutical | <i>Tick Box</i> |
| Screening | <i>Tick Box</i> |
| Diagnostic | <i>Tick Box</i> |
| Surgical | <i>Tick Box</i> |
| Service delivery & organisation (SDO) | <i>Tick Box</i> |
| New technologies | <i>Tick Box</i> |
| Place in R&D cycle | <i>Pretrial/Post trial</i> |
| Modelling undertaken | <i>Yes/No</i> |
| Decision analytic | <i>Tick Box</i> |
| Markov | <i>Tick Box</i> |
| Simulation | <i>Tick Box</i> |
| Cost modelling | <i>Tick Box</i> |
| Risk modelling | <i>Tick Box</i> |
| Bayesian analysis | <i>Tick Box</i> |
| Delphi | <i>Tick Box</i> |
| Other | <i>Free text description</i> |
| Outcomes | |
| Clinical | <i>Tick Box</i> |
| Cost | <i>Tick Box</i> |
| Health economic | <i>Tick Box</i> |
| | |
| Identification of key uncertainties | <i>Tick Box</i> |
| Further research worthwhile | <i>Yes/No</i> |
| Type of research | <i>Free text description</i> |



Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Kent Woods,
Director, NHS HTA Programme
& Professor of Therapeutics,
University of Leicester

Professor Bruce Campbell,
Consultant Vascular & General
Surgeon, Royal Devon & Exeter
Hospital

Dr John Reynolds, Clinical
Director, Acute General
Medicine SDU, Radcliffe
Hospital, Oxford

Professor Shah Ebrahim,
Professor in Epidemiology
of Ageing, University of
Bristol

Dr Ron Zimmern, Director,
Public Health Genetics Unit,
Strangeways Research
Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Kent Woods, Director,
NHS HTA Programme,
Department of Medicine and
Therapeutics, Leicester Royal
Infirmary, Robert Kilpatrick
Clinical Sciences Building,
Leicester

Chair,
Professor Shah Ebrahim,
Professor in Epidemiology of
Ageing, Department of Social
Medicine, University of Bristol,
Canyngge Hall, Whiteladies
Road, Bristol

Deputy Chair,
Professor Jenny Hewison,
Professor of Health Care
Psychology, Academic Unit of
Psychiatry and Behavioural
Sciences, University of Leeds
School of Medicine, Leeds

Professor Douglas Altman,
Professor of Statistics in
Medicine, Centre for Statistics
in Medicine, Oxford University,
Institute of Health Sciences,
Cancer Research UK Medical
Statistics Group, Headington,
Oxford

Professor John Bond, Professor
of Health Services Research,
Centre for Health Services
Research, University of
Newcastle, School of Health
Sciences, Newcastle upon Tyne

Professor John Brazier, Director
of Health Economics, Sheffield
Health Economics Group,
School of Health & Related
Research, University of
Sheffield, ScHARR Regent
Court, Sheffield

Dr Andrew Briggs, Public
Health Career Scientist, Health
Economics Research Centre,
University of Oxford, Institute
of Health Sciences, Oxford

Dr Christine Clark, Medical
Writer & Consultant Pharmacist,
Cloudside, Rossendale, Lancs
and
Principal Research Fellow,
Clinical Therapeutics in the
School of Pharmacy, Bradford
University, Bradford

Professor Nicky Cullum,
Director of Centre for Evidence
Based Nursing, Department of
Health Sciences, University of
York, Research Section,
Seeborn Rowntree Building,
Heslington, York

Dr Andrew Farmer, Senior
Lecturer in General Practice,
Department of Primary Health
Care, University of Oxford,
Institute of Health Sciences,
Headington, Oxford

Professor Fiona J Gilbert,
Professor of Radiology,
Department of Radiology,
University of Aberdeen, Lillian
Sutton Building, Foresterhill,
Aberdeen

Professor Adrian Grant,
Director, Health Services
Research Unit, University of
Aberdeen, Drew Kay Wing,
Polwarth Building, Foresterhill,
Aberdeen

Professor Alastair Gray, Director,
Health Economics Research
Centre, University of Oxford,
Institute of Health Sciences,
Headington, Oxford

Professor Mark Haggard,
Director, MRC ESS Team, CBU
Elsworth House, Addenbrooke's
Hospital, Cambridge

Professor F D Richard Hobbs,
Professor of Primary Care &
General Practice, Department of
Primary Care & General
Practice, University of
Birmingham, Primary Care and
Clinical Sciences Building,
Edgbaston, Birmingham

Professor Peter Jones, Head of
Department, University
Department of Psychiatry,
University of Cambridge,
Addenbrooke's Hospital,
Cambridge

Professor Sallie Lamb, Research
Professor in Physiotherapy/Co-
Director, Interdisciplinary
Research Centre in Health,
Coventry University, Coventry

Dr Donna Lamping, Senior
Lecturer, Health Services
Research Unit, Public Health
and Policy, London School of
Hygiene and Tropical Medicine,
London

Professor David Neal, Professor
of Surgical Oncology, Oncology
Centre, Addenbrooke's Hospital,
Cambridge

Professor Tim Peters, Professor
of Primary Care Health Services
Research, Division of Primary
Health Care, University of
Bristol, Cotham House, Cotham
Hill, Bristol

Professor Ian Roberts, Professor
of Epidemiology & Public
Health, Intervention Research
Unit, London School of
Hygiene and Tropical Medicine,
London

Professor Peter Sandercock,
Professor of Medical Neurology,
Department of Clinical
Neurosciences, University of
Edinburgh, Western General
Hospital NHS Trust, Bramwell
Dott Building, Edinburgh

Professor Martin Severs,
Professor in Elderly Health
Care, Portsmouth Institute of
Medicine, Health & Social Care,
St George's Building,
Portsmouth

Dr Jonathan Shapiro, Senior
Fellow, Health Services
Management Centre, Park
House, Birmingham

Diagnostic Technologies & Screening Panel

Members

| | | | |
|--|---|--|---|
| <p>Chair, Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge</p> | <p>Dr David Elliman, Consultant in Community Child Health, London</p> <p>Dr Andrew Farmer, Senior Lecturer in General Practice, Institute of Health Sciences, University of Oxford</p> <p>Dr Karen N Foster, Clinical Lecturer, Dept of General Practice & Primary Care, University of Aberdeen</p> <p>Professor Jane Franklyn, Professor of Medicine, University of Birmingham</p> <p>Professor Antony J Franks, Deputy Medical Director, The Leeds Teaching Hospitals NHS Trust</p> | <p>Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London</p> <p>Dr Susanne M Ludgate, Medical Director, Medical Devices Agency, London</p> <p>Dr William Rosenberg, Senior Lecturer and Consultant in Medicine, University of Southampton</p> <p>Dr Susan Schonfield, CPHM Specialised Services Commissioning, Croydon Primary Care Trust</p> <p>Dr Margaret Somerville, Director of Public Health, Teignbridge Primary Care Trust, Devon</p> | <p>Mr Tony Tester, Chief Officer, South Bedfordshire Community Health Council, Luton</p> <p>Dr Andrew Walker, Senior Lecturer in Health Economics, University of Glasgow</p> <p>Professor Martin J Whittle, Head of Division of Reproductive & Child Health, University of Birmingham</p> <p>Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow</p> |
| <p>Dr Paul Cockcroft, Consultant Medical Microbiologist/Laboratory Director, Public Health Laboratory, St Mary's Hospital, Portsmouth</p> <p>Professor Adrian K Dixon, Professor of Radiology, Addenbrooke's Hospital, Cambridge</p> | | | |

Pharmaceuticals Panel

Members

| | | | |
|--|--|---|---|
| <p>Chair, Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Oxford Radcliffe Hospital</p> <p>Professor Tony Avery, Professor of Primary Health Care, University of Nottingham</p> <p>Professor Iain T Cameron, Professor of Obstetrics & Gynaecology, University of Southampton</p> <p>Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London</p> | <p>Dr Christopher Cates, GP and Cochrane Editor, Bushey Health Centre, Bushey, Herts.</p> <p>Mr Charles Dobson, Special Projects Adviser, Department of Health</p> <p>Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham</p> <p>Dr Karen A Fitzgerald, Pharmaceutical Adviser, Bro Taf Health Authority, Cardiff</p> <p>Professor Alastair Gray, Professor of Health Economics, University of Oxford</p> | <p>Mrs Sharon Hart, Managing Editor, <i>Drug & Therapeutics Bulletin</i>, London</p> <p>Dr Christine Hine, Consultant in Public Health Medicine, Bristol South & West Primary Care Trust</p> <p>Professor Robert Peveler, Professor of Liaison Psychiatry, Royal South Hants Hospital, Southampton</p> <p>Dr Frances Rotblat, CPMP Delegate, Medicines Control Agency, London</p> <p>Mrs Katrina Simister, New Products Manager, National Prescribing Centre, Liverpool</p> | <p>Dr Ken Stein, Senior Lecturer in Public Health, University of Exeter</p> <p>Professor Terence Stephenson, Professor of Child Health, University of Nottingham</p> <p>Dr Richard Tiner, Medical Director, Association of the British Pharmaceutical Industry, London</p> <p>Professor Dame Jenifer Wilson-Barnett, Head of Florence Nightingale School of Nursing & Midwifery, King's College, London</p> |
|--|--|---|---|

Therapeutic Procedures Panel

Members

Chair,

Professor Bruce Campbell,
Consultant Vascular and
General Surgeon, Royal Devon
& Exeter Hospital

Dr Mahmood Adil, Head of
Clinical Support & Health
Protection, Directorate of
Health and Social Care (North),
Department of Health,
Manchester

Professor John Bond, Head of
Centre for Health Services
Research, University of
Newcastle upon Tyne

Mr Michael Clancy, Consultant
in A & E Medicine,
Southampton General Hospital

Dr Carl E Counsell, Senior
Lecturer in Neurology,
University of Aberdeen

Dr Keith Dodd, Consultant
Paediatrician, Derbyshire
Children's Hospital, Derby

Professor Gene Feder, Professor
of Primary Care R&D, Barts &
the London, Queen Mary's
School of Medicine and
Dentistry, University of London

Ms Bec Hanley, Freelance
Consumer Advocate,
Hurstpierpoint, West Sussex

Professor Alan Horwich,
Director of Clinical R&D, The
Institute of Cancer Research,
London

Dr Phillip Leech, Principal
Medical Officer for Primary
Care, Department of Health,
London

Mr George Levy, Chief
Executive, Motor Neurone
Disease Association,
Northampton

Professor James Lindsay,
Professor of Psychiatry for the
Elderly, University of Leicester

Dr Mike McGovern, Senior
Medical Officer, Heart Team,
Department of Health, London

Dr John C Pounsford,
Consultant Physician, North
Bristol NHS Trust

Professor Mark Sculpher,
Professor of Health Economics,
Institute for Research in the
Social Services, University of
York

Dr L David Smith, Consultant
Cardiologist, Royal Devon &
Exeter Hospital

Professor Norman Waugh,
Professor of Public Health,
University of Aberdeen

Expert Advisory Network

Members

Mr Gordon Aylward,
Chief Executive,
Association of British Health-
Care Industries, London

Ms Judith Brodie,
Head of Cancer Support
Service, Cancer BACUP, London

Mr Shaun Brogan,
Chief Executive, Ridgeway
Primary Care Group, Aylesbury,
Bucks

Ms Tracy Bury,
Project Manager, World
Confederation for Physical
Therapy, London

Mr John A Cairns,
Professor of Health Economics,
Health Economics Research
Unit, University of Aberdeen

Professor Howard Stephen Cuckle,
Professor of Reproductive
Epidemiology, Department of
Paediatrics, Obstetrics &
Gynaecology, University of
Leeds

Professor Nicky Cullum,
Director of Centre for Evidence
Based Nursing, University of York

Dr Katherine Darton,
Information Unit, MIND – The
Mental Health Charity, London

Professor Carol Dezateux,
Professor of Paediatric
Epidemiology, London

Professor Martin Eccles,
Professor of Clinical
Effectiveness, Centre for Health
Services Research, University of
Newcastle upon Tyne

Professor Pam Enderby,
Professor of Community
Rehabilitation, Institute of
General Practice and Primary
Care, University of Sheffield

Mr Leonard R Fenwick,
Chief Executive, Newcastle
upon Tyne Hospitals NHS Trust

Professor David Field,
Professor of Neonatal Medicine,
Child Health, The Leicester
Royal Infirmary NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and
President, National Childbirth
Trust, Henfield, West Sussex

Ms Grace Gibbs,
Deputy Chief Executive,
Director for Nursing, Midwifery
& Clinical Support Servs., West
Middlesex University Hospital,
Isleworth, Middlesex

Dr Neville Goodman,
Consultant Anaesthetist,
Southmead Hospital, Bristol

Professor Robert E Hawkins,
CRC Professor and Director of
Medical Oncology, Christie CRC
Research Centre, Christie
Hospital NHS Trust, Manchester

Professor F D Richard Hobbs,
Professor of Primary Care &
General Practice, Department of
Primary Care & General
Practice, University of
Birmingham

Professor Allen Hutchinson,
Director of Public Health &
Deputy Dean of SCHARR,
Department of Public Health,
University of Sheffield

Professor Rajan Madhok,
Medical Director & Director of
Public Health, Directorate of
Clinical Strategy & Public
Health, North & East Yorkshire
& Northern Lincolnshire Health
Authority, York

Professor David Mant,
Professor of General Practice,
Department of Primary Care,
University of Oxford

Professor Alexander Markham,
Director, Molecular Medicine
Unit, St James's University
Hospital, Leeds

Dr Chris McCall,
General Practitioner, The
Hadleigh Practice, Castle
Mullen, Dorset

Professor Alistair McGuire,
Professor of Health Economics,
London School of Economics

Dr Peter Moore,
Freelance Science Writer,
Ashtead, Surrey

Dr Andrew Mortimore,
Consultant in Public Health
Medicine, Southampton City
Primary Care Trust

Dr Sue Moss,
Associate Director, Cancer
Screening Evaluation Unit,
Institute of Cancer Research,
Sutton, Surrey

Professor Jon Nicholl,
Director of Medical Care
Research Unit, School of Health
and Related Research,
University of Sheffield

Mrs Julietta Patnick,
National Co-ordinator, NHS
Cancer Screening Programmes,
Sheffield

Professor Chris Price,
Visiting Chair – Oxford, Clinical
Research, Bayer Diagnostics
Europe, Cirencester

Ms Marianne Rigge,
Director, College of Health,
London

Professor Sarah Stewart-Brown,
Director HSRU/Honorary
Consultant in PH Medicine,
Department of Public Health,
University of Oxford

Professor Ala Szczepura,
Professor of Health Service
Research, Centre for Health
Services Studies, University of
Warwick

Dr Ross Taylor,
Senior Lecturer, Department of
General Practice and Primary
Care, University of Aberdeen

Mrs Joan Webster,
Consumer member, HTA –
Expert Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (<http://www.nchta.org>) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.