

How much Detail is needed in Cost Estimation in an Economic Evaluation alongside a Clinical Trial to Optimise Evidence for Decisions?

Edward Charles Frederick Wilson

MSc Health Economics, BSc (Honours) Economics & Finance

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Health Economics Group
Norwich Medical School
Faculty of Medicine and Health Science
University of East Anglia

April 2014

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Acquiring evidence to support decision making is expensive. Collecting resource use data alongside a randomised controlled clinical trial is particularly so due to the multi-dimensional nature of costs: different costs are incurred by different agencies with varying methods and systems to account for these.

Trialists are faced with decisions over how to collect such data, in particular different 'levels' of detail are possible. For example, hospitalisations can be costed (1) on a top-down, per admission basis multiplied by a representative unit cost, (2) a bottom-up basis measuring every component of care such as nursing and medic time, investigations and other procedures and drugs used which are each multiplied by relevant unit costs, or (3) some intermediate level of aggregation. The top-down data will be less expensive to obtain but may be less accurate (biased and/or over- or under-estimation of uncertainty) compared with the bottom-up. I refer to these alternative methods as 'data collection processes'.

Currently such decisions are based on the judgement of the trialist(s). However, formal quantification of the added value of one data collection process versus another compared with the added cost would inform the efficient allocation of research resources.

In this thesis I extend the use of value of information analysis to compare the incremental cost and benefit of one data process with another, further extending this to estimate the optimal mix of observations between two processes.

Using an example dataset I find that the method is workable, requiring prior information on the relationship between the two processes which can be obtained from either a pilot or feasibility study or expert opinion.

When incorporated with other concurrent developments in value of information analysis, the method has the potential to provide a decision analytic approach to the complete design of clinical trials.

Table of Contents

Contents

Abstract	i
Table of Contents	ii
List of Tables	vi
List of Figures	viii
Acknowledgements	x
Author's Declaration	xii
Mathematical Notation	xiii
Abbreviations	xiv
1. Introduction and Background.....	1
1.1. Overview.....	1
1.2. Background to Economic evaluation.....	6
1.2.1. Four approaches to economic evaluation.....	6
1.2.2. The need for economic evaluation.....	7
1.2.3. Welfarism and the "extra-" or "non-welfarist" foundations of economic evaluation.....	9
1.3. Conducting an economic evaluation.....	14
1.3.1. Define the study question	14
1.3.2. Describe each comparator	16
1.3.3. Establish the effectiveness of each comparator	16
1.3.4. Identify all important costs and consequences.....	17
1.3.5. Measure all costs and consequences	19
1.3.6. Value all costs and consequences	21
1.3.7. Adjust costs and consequences for differential timing.....	23
1.3.8. Undertake an incremental analysis.....	24
1.3.9. Analysis of uncertainty	25
1.3.10. Write-up and discussion.....	28
1.4. Value of Information Analysis	30
1.4.1. Analytic solution	31
1.4.2. Numeric solution	40
1.5. Decision models vs Clinical Trials and the Cycle of Evidence Based Medicine	50
1.5.1. Trials vs Models for economic evaluations	50

1.5.2.	Economic evaluation, Value of Information Analysis and the Iterative Approach to Decision Making	52
1.6.	Incorporation of EE & Vol / EBM into health policy in the UK and internationally.	57
1.6.1.	Economic evaluation in the UK, Europe and Globally	57
1.6.2.	Value of information analysis in health policy	60
1.7.	Statement of study problem	63
1.7.1.	Level of detail in costing in an economic evaluation.....	63
2.	Literature review	66
2.1.	Introduction.....	66
2.2.	Method	67
2.2.1.	Use of Vol or similar techniques to inform study design.	67
2.2.2.	The Cost of collecting resource use data alongside clinical trials and/or the cost of clinical trials themselves.....	69
2.2.3.	Comparisons of alternative data collection techniques.....	70
2.3.	Results	71
2.3.1.	Use of Vol or similar techniques to inform study design.	71
2.3.2.	The Cost of collecting resource use data alongside clinical trials and/or the cost of clinical trials themselves.....	75
2.3.3.	Comparisons of alternative data collection techniques.....	80
2.4.	Discussion	90
2.4.1.	Use of Vol or similar techniques to inform study design.	90
2.4.2.	The Cost of collecting resource use data alongside clinical trials and/or the cost of clinical trials themselves.....	92
2.4.3.	Comparisons of alternative data collection techniques.....	93
2.5.	Conclusion	94
3.	A critique of Value of Information Analysis.....	96
3.1.	Introduction.....	96
3.2.	Criticism of methods	97
3.2.1.	Issues relating to Vol in general	97
3.2.2.	Issues specific to the analytic solution	100
3.2.3.	Issues specific to the numeric solution	102
3.2.4.	Summary.....	104
3.3.	Ignoring correlation between model input parameters	106
3.3.1.	Introduction.....	106
3.3.2.	Method & Data.....	117
3.3.3.	Results	120
3.3.4.	Discussion	138

3.4.	Summary & Conclusion	152
4.	Using Vol to assess the level of detail required in resource data collection in an economic evaluation alongside a clinical trial.....	153
4.1.	Introduction.....	153
4.2.	Methods	157
4.2.1.	Data	164
4.3.	Results	167
4.3.1.	Value of information analysis for a repeated ELEVATE trial	168
4.3.2.	Value of information analysis of four separate studies reporting incremental QALYs, incremental cost, incremental non-drug cost and incremental drug cost alone	168
4.3.3.	Comparison of value of alternative data collection processes on drug costs	169
4.3.4.	Overall efficient trial design	169
4.4.	Discussion	173
4.4.1.	Implications of results	173
4.4.2.	Practical applications.....	174
4.4.3.	Determinants of results of 4.3.3.....	175
4.4.4.	Comparison with other studies	177
4.4.5.	Strengths and Weaknesses.....	177
4.5.	Conclusion	181
5.	Discussion & Further Work.....	183
5.1.	Summary of Findings	183
5.2.	Discussion Points	188
5.2.1.	Strengths and weaknesses of the method	188
5.2.2.	Broader issues	190
5.3.	Implications for practice.....	196
5.4.	Further work.....	199
5.4.1.	Methodological issues.....	199
5.4.2.	Applied issues	203
5.5.	Conclusion	206
	References.....	207
	APPENDIX A: Calculation of Beneficial Populations and Research Costs.....	226
A1.1	Potential beneficial population and research costs: BECCA.	226
A1.2	Potential beneficial population and research costs: ELEVATE.....	229
A1.3	Potential beneficial population and research costs: CESAR.	232
	APPENDIX B: Correlation Coefficients amongst Data components in the BECCA study	235

APPENDIX C: Full details of calculations for Chapter 4 237
APPENDIX D: Exploration of impact of reducing uncertainty in the correlation coefficient. 254
 Results 256
APPENDIX E: Power calculation..... 259

List of Tables

Table 1-1: EVPI Illustration	42
Table 1-2: EVSI & ENBS illustration	48
Table 3-1: 2x2 contingency table for conversion to open procedure	109
Table 3-2: Conversion rates to an open procedure.....	120
Table 3-3: Risk of inter-operative complications.....	120
Table 3-4: BECCA Summary Statistics.....	124
Table 3-5: BECCA Vol Statistics.....	125
Table 3-6: BECCA Vol statistics, $\rho=0$	125
Table 3-7: ELEVATE Summary Statistics	128
Table 3-8: ELEVATE Vol Statistics	129
Table 3-9: ELEVATE Vol Statistics, $\rho=0$	129
Table 3-10: CESAR Summary Statistics.....	132
Table 3-11: CESAR Vol Statistics	132
Table 3-12: CESAR Vol Statistics, $\rho=0$	133
Table 3-13: Cholecystectomy Summary Statistics: uncorrelated.....	136
Table 3-14: Cholecystectomy Summary Statistics : correlated	136
Table 3-15: Cholecystectomy Vol Statistics : uncorrelated.....	137
Table 3-16: Cholecystectomy Vol Statistics : correlated	137
Table 4-1: Summary statistics - means.....	166
Table 4-2: Summary statistics - variance and covariance	166
Table 4-3: Summary results.....	170
Table A-1: BECCA Beneficial Population.....	226
Table A-2: BECCA trial original budget	227
Table A-3: BECCA Budget, 2010 £.....	227
Table A-4: ELEVATE Beneficial Population	229
Table A-5: ELEVATE trial original budget.....	230
Table A-6: ELEVATE Budget, 2010 £	230
Table A-7: CESAR Beneficial Population	232
Table A-8: CESAR trial original budget.....	233

Table A-9: CESAR Budget, 2010 £ 234
Table B-1: Correlation Coefficients between cost components of BECCA data..... 235

List of Figures

Figure 1-1: Graphical Illustration of EVPI	33
Figure 1-2: Prior (a) and (expected) posterior (b) distribution of incremental net benefit....	36
Figure 1-3: Population EVSI (pEVSI), total cost of sampling (TC) and ENBS.	39
Figure 1-4: Structure of Decision Model	41
Figure 1-5: EVSI example	46
Figure 1-6: EVSI Example (continued)	47
Figure 1-7: EVSI, Cost of sampling and ENBS.....	49
Figure 1-8: The Cycle of ‘Economics’ Based Medicine	54
Figure 2-1: Flowchart, review 1.....	68
Figure 2-2: Flowchart, review 2.....	69
Figure 2-3: Flowchart, review 3.....	70
Figure 3-1: Example decision tree (part).....	108
Figure 3-2: Scatterplot of sampled costs and QALYs with (a) fixed costs and QALYs, probabilities allowed to vary, and (b) with uncertainty in cost, QALYs and probabilities	109
Figure 3-3: Scatterplot of 10,000 uncorrelated draws.....	110
Figure 3-4: Scatterplot of 10,000 correlated draws (a) natural units and (b) logarithmic....	110
Figure 3-5: BECCA cost-effectiveness plane and 95% confidence ellipse	123
Figure 3-6: BECCA cost-effectiveness acceptability curve.....	123
Figure 3-7: BECCA Plot of Incremental Net Benefit.....	123
Figure 3-8: BECCA per patient EVSI and EVPI	123
Figure 3-9: BECCA EVSI and ENBS.....	123
Figure 3-10: BECCA variance of INB and EVPI as a function of correlation coefficient.....	123
Figure 3-11: BECCA Optimal sample size (n^*) of a new study measuring INB, and ENBS at n^* as a function of the correlation coefficient.....	124
Figure 3-12: ELEVATE cost-effectiveness plane and 95% confidence ellipse.....	127
Figure 3-13: ELEVATE cost-effectiveness acceptability curve	127
Figure 3-14: ELEVATE Plot of Incremental Net Benefit	127
Figure 3-15: ELEVATE per patient EVPI and EVSI.....	127
Figure 3-16: ELEVATE Optimal sample size for a new trial.....	127

Figure 3-17: ELEVATE Optimal sample size (n^*) of a study measuring INB, and ENBS at n^* as a function of ρ	127
Figure 3-18: CESAR cost-effectiveness plane and 95% confidence ellipse	131
Figure 3-19: CESAR cost-effectiveness acceptability curve	131
Figure 3-20: CESAR plot of incremental net benefit.....	131
Figure 3-21: CESAR per patient EVSI and EVPI	131
Figure 3-22: CESAR Optimal sample size for a new trial	131
Figure 3-23: CESAR Optimal Sample size (n^*) of a study measuring INB, and ENBS at n^* as a function of ρ	131
Figure 3-24: LC Cost-effectiveness place: scatterplot and 95% confidence ellipse	135
Figure 3-25: LC Cost-effectiveness acceptability curve	135
Figure 3-26: LC Incremental Net Benefit	135
Figure 3-27: LC EVPI & EVSI(b) under (i) independent parameters and (ii) structurally induced correlation	135
Figure 3-28: Vol statistics for BECCA data as ρ is increased from -1 to +1	142
Figure 3-29: Vol statistics for ELEVATE data as ρ is increased from -1 to +1.....	144
Figure 3-30: Vol statistics for CESAR data as ρ is increased from -1 to +1	147
Figure 4-1: Calculation of ENGS of a trial using a combination of processes A & B.....	156
Figure 4-2: Distribution of Incremental Net Benefit	170
Figure 4-3: EVSI, total cost and opportunity loss, and ENGS.....	170
Figure 4-4: ENGS QALYs.....	170
Figure 4-5: ENGS Cost.....	170
Figure 4-6: ENGS Drug Cost	171
Figure 4-7: ENGS Non-drug Cost	171
Figure 4-8: EVPI and EVPPi @ $\lambda=\text{£}5,000$	171
Figure 4-9: Optimal mix of observations from each data process	172
Figure 4-10: Optimal mix of observations on each process as a function of ρ	177
Figure 5-1: Schematic of bivariate distribution of net benefit with two comparators	201

Acknowledgements

Firstly I would like to thank my supervisors, Prof Miranda Mugford, Prof Lee Shepstone and Dr Garry Barton for their help, encouragement and support over the last seven years. I would also like to thank Prof Andy Briggs, Dr Pelham Barton and others who commented on various earlier incarnations of parts of this work presented at Health Economists' Study Group meetings in Norwich (January 2008), Manchester (January 2009) and LSE (January 2010). Thank you also to Dr Marcello Morciano and Darren Burns for comments on presentation of this work at a Health Economics Group seminar, UEA (February 2013), and to Dr Katie Saunders, Cambridge Centre for Health Services Research for assistance with coding confidence ellipses.

I am most grateful to the chief investigators of the three trials for allowing me to use the raw datasets as examples in this thesis. These are Prof Georgina Charlesworth (BECCA), Prof David Price (ELEVATE) and Prof Diana Elbourne (CESAR). Thank you also to Rachael Mold and Dr Stan Musgrave for assistance with providing research cost estimates used in the Value of Information analyses on the three datasets.

I would like to thank the Health Foundation, the National Institute for Health Research and the Norwich Medical School for funding my research training fellowship, academic post and PhD fees respectively, without whose financial assistance it would not have been possible to undertake this research.

Finally I would like to thank my partner, Joe Murphy for his patience and for picking me up when I was down, and family and friends for keeping me at it. In particular I wish to thank my mother, Anne Wilson, for her constant encouragement and nagging.

For Mum

23rd March 1944 – 2nd December 2012

Author's Declaration

Parts of Chapter 1 describing value of information analysis (Section 1.4.2), Sections 3.2.1 and 3.2.3 have previously been published as Wilson & Abrams 2010,¹ although some wording and phraseology have been modified.

The following conference presentations have been made based on sections of this thesis:

Wilson E, Mugford M, Barton G, Shepstone L. A comparison of the relative value of information from different resource use data collection processes in an economic evaluation alongside a clinical trial. International Health Economics Association, Sydney, July 2013.

Wilson E, Mugford M, Shepstone L. Expected value of sample information with correlated parameters. Health Economists' Study Group, London School of Economics, January 2010.

Wilson E, Mugford M, Shepstone L. The consequences of ignoring correlations between model input parameters in value of information analyses. Health Economists' Study Group, Manchester, January 2009.

Wilson E, Shepstone L, Mugford M. Estimating the returns from reducing bias in clinical trials. Health Economists' Study Group, Norwich, January 2008.

Mathematical Notation

n	Sample size (per arm unless otherwise stated) of a proposed trial
n^*	Optimal sample size (per arm unless otherwise stated) for a data collection process (e.g. clinical trial, database analysis)
x_0	Prior estimate of mean of x
x_1	Posterior or preposterior estimate of mean of x (depending on context)
b	Incremental net benefit
C	Cost
Θ	Parameter of interest
E	Effect (e.g. QALYs)
Δx	Increment of x , i.e. $x_2 - x_1$
$V(x)$	Variance of x
$\text{Cov}(x,y)$	Covariance between x and y
$\rho(x,y)$	Pearson correlation coefficient between x and y
N	Total population who can benefit from information yielded from a data collection process
λ	Threshold / maximum willingness to pay for a unit of outcome (e.g. QALYs)
$L_{N^*}(x)$	Unit Normal Linear Loss Integral evaluated at x .
$\phi(x)$	Standard normal probability density function evaluated at x .
$\Phi(x)$	Standard normal cumulative density function evaluated at x .
$ x $	Absolute value of x , e.g. $ -2 = 2$
$I\{.\}$	Indicator function returning the value 1 if expression inside the brackets is true, else 0.

Abbreviations

CRF	Case Record Form
EMA	European Medicines Agency
ENGS	Expected Net Gain of Sampling
EVPI	Expected Value of Perfect Information
EVSI	Expected Value of Sample Information
FDA	US Food and Drugs Administration
HC	Human Capital
ICER	Incremental Cost Effectiveness Ratio
INB	Incremental Net Benefit
OC	Opportunity Cost
QALY	Quality Adjusted Life Year
TC	Total Cost
UNLLI	Unit Normal Linear Loss Integral
VoI	Value of Information analysis
WoK	Web of Knowledge

1. Introduction and Background

1.1. Overview

In an economic evaluation conducted alongside a clinical trial, analysts are faced with many design issues. On the cost side of the equation, analysts must first determine the perspective of their analysis. From this, a list of resource use items must be identified. Decisions must then be made as to how to measure, value and finally analyse them.

This thesis focuses on informing decisions as to how best to measure, value and analyse resource use data. Collecting resource use data alongside a clinical trial can be a complex, time-consuming and research intensive exercise. Unlike many clinical outcome measures, costs are always a 'compound' outcome measure comprising many individual components. For example, the total cost of a particular treatment strategy to the NHS will include not only the cost of the treatment itself, but any associated primary, secondary and tertiary care activity and prescribed medication. Furthermore, depending on the perspective of the analysis, total cost may also include costs borne by other public sector bodies (such as social services), patient out of pocket costs and the value of any productivity foregone to society due to morbidity or premature mortality.

Data can be collected at different levels of detail/aggregation. For example the cost of a hospitalisation could be estimated by measuring and valuing each separate component of care, such as nursing and medic time, individual diagnostics and procedures, pharmacy costs, 'hotel' costs and some allocation of overheads. Alternatively it may simply be approximated with a unit cost for a particular admission. The former ('micro-costing') would require considerable research expense to accurately measure the time spent on each activity, recording every diagnostic and exact quantities of drugs administered, whilst the latter ('gross costing') simply requires some record of the admission in the patient's case record form (CRF).

The appropriate level of detail may depend on the research question being asked. However, a formal economics-based approach would be to consider whether the additional benefit of the more detailed research method is 'worth' the additional cost. Mugford,² in an analysis of the costs of perinatal care found that use of simple (top-down) approaches to costing did not lead to consistently biased estimates compared with more complex, micro-costing

approaches although estimates were sensitive to modelling techniques employed. As a result, whilst simple costing approaches “would not bias policy decisions, more complex methods can reduce uncertainty about the limits of possible variation in different settings and are necessary for validation of simpler approaches”.²

However, it is unclear at what point a more detailed approach to costing is required and where a simpler approach will suffice. The appropriate choice would depend upon the added value of the information generated from a more detailed cost analysis compared with the additional cost (versus the simpler costing approach).

Value of information analysis (Vol) is a decision analytic technique for comparing the added cost of a research project with its added value.^{3 4} The gain from additional information in assisting a decision can be valued in terms of the consequences of a ‘wrong’ decision multiplied by the reduction in probability of a ‘wrong’ decision. The expected return in terms of reduced uncertainty of research can then be valued.⁵ Vol is warranted when:

- The proposed research has the potential of changing current practice;
- there is likely to be a large advantage over the new treatment compared with current practice; and
- the cost of gathering new information is not too large compared with its value.⁶

Vol is based firmly within a Bayesian statistical framework where probability represents degrees of belief about plausible values for a parameter rather than the long run relative frequency with which an event occurs. The key concept in Bayesian analysis is the updating of a prior belief with data to form a posterior belief using Bayes theorem.⁷ For this reason Bayesian analysis is sometimes referred to as posterior analysis.⁴ Value of information analysis requires prediction of data conditional on the prior to generate an expected posterior distribution. It is thus sometimes referred to as preposterior analysis.⁴

The inclusion of value of information analyses as a part of economic evaluations is becoming increasingly common.⁸⁻¹⁹ This is useful to direct future research effort to where it can achieve the greatest return for finite research funding. Its main use is to determine the optimal sample size for a future study based on the marginal gain from an additional trial enrollee versus the marginal cost. The optimal point is where the marginal cost is equal to the (value of the) marginal gain, a concept directly analogous to the profit maximising condition in the theory of the firm. However, it has potential for use in the planning stages

of a clinical trial to inform other trial design issues such as whether to conduct an RCT or observational study.²⁰

In this thesis I explore the use of Vol with a view to answering the following question:

Is it worth conducting a micro-costing exercise as part of a proposed economic evaluation alongside a clinical trial, or will a less detailed approach offer better value for money for finite research funds?

This can be thought of as a special case of the more general question:

Is there an optimal mix of observations using both data collection methods within the same trial?

The former question is a case where only one or the other process is considered. The latter is a more useful question to address as the former can be thought of as a special case of the latter where the number of observations on one data process or the other is zero.

The remainder of this thesis is structured as follows.

The remainder of this chapter covers the background to economic evaluation, including its origins, purpose and theoretical basis in (and departure from) traditional welfarist economics. I explain how economic evaluations are conducted, distinguishing between two implementations: either alongside a (randomised controlled) clinical trial or as a decision model. I focus in particular on methods to identify, measure and value resource use alongside RCTs.

Following this I introduce the concepts of Vol, defining two approaches: analytic and numeric (simulation). The former is more commonly associated with trial-based analyses whilst the latter is commonly conducted alongside decision models (although it is technically possible to conduct either method on either implementation). I then discuss the relative merits of decision models and RCT-based studies, concluding that they should be seen as complements rather than substitutes. This leads into a view of evidence based medicine (or 'economics based medicine'), bringing together the concepts of economic evaluation and value of information centred on an iterative approach to decision making. This comprises systematic review and synthesis of all relevant information to inform two decisions: firstly whether or not to adopt an intervention, and secondly whether further information should be sought to reduce decision uncertainty. This cycle is fundamental to the analytic approach of this thesis. I then review the use of components of the cycle of EBM to inform health

policy in the UK and elsewhere before concluding Chapter 1 with a restatement of the study question of this thesis.

Chapter 2 reviews the literature to determine whether Vol or any similar economics-based prioritisation process has been used to inform the design of clinical trials other than for sample size calculation. I also seek any studies examining the cost of collecting resource use data alongside clinical trials and comparisons of alternative approaches to collecting and measuring the same resource use data.

Chapter 3 is a detailed investigation and critique of the concepts of Vol. In order to adapt an existing process it is important to understand its current strengths and limitations. In this chapter I firstly explore issues relating to value of information analysis as applied to the health care sector in general. Briefly, these are defining the relevant beneficial patient population, the independence (or lack thereof) between the adoption and research decisions, and the existence of multiple jurisdictions and subsequent risk of free riding. I then consider issues specific to each of the two major implementations of Vol (the analytic and numeric solutions). Following this, I discuss one particular issue in more detail, that is, the assumption frequently made in decision modelling that input parameters are independent of one another. Ignoring correlation between input parameters may lead to incorrect estimation of decision uncertainty. As the Vol statistics are a function of decision uncertainty, ignoring correlation could potentially provide misleading research recommendations.

After consideration of the strengths and weaknesses of Vol, I address the study question in Chapter 4. I consider the case where there are two methods to measure a particular cost item such as drug costs: a micro-costing approach identifying, measuring and valuing every milligramme of every drug consumed by a patient in a trial (Process A), or a gross-costing approach where a prescription for a particular drug or drug class is noted on the records from which assumptions are made about dose and duration of treatment (Process B). By specifying a prior bivariate relationship between the two processes, predicting the results of data gathered using process B can be used to revise belief about plausible values using process A. As Process A is believed to be the superior process, the resulting preposterior distribution of process A can then be used to generate the preposterior distribution of incremental net benefit and thus predict the expected reduction in decision uncertainty from data gathered using process B, and hence the value of information using process B, or a mixture of observations using both processes.

Finally in Chapter 5 I discuss the findings and implications of this thesis in full detail and present a pathway for future research and new questions this research has raised.

1.2. Background to Economic evaluation

In this section I outline the theoretical basis for economic evaluation.

Firstly, I briefly define economic evaluation, introducing the four main types (Section 1.2.1). I then discuss the need for such analyses and explore the theoretical foundations of economic evaluation, specifically its 'extra-welfarist' (or 'non-welfarist') basis (Sections 1.2.2 & 1.2.3).

1.2.1. Four approaches to economic evaluation

The purpose of much health services research is to assist in decisions as to whether or not to adopt a given technology (drug, device, programme or technique). The formal definition of economic evaluation is a “comparative analysis of alternative courses of action in terms of both their costs and consequences”.²¹ Conventionally, four types of economic evaluation are defined:²¹

- Cost benefit analysis (CBA)
- Cost effectiveness analysis (CEA)
- Cost utility analysis (CUA)
- Cost minimisation analysis (CMA)

The difference between each is in the outcome measure employed. In CBA, outcomes are valued in the same metric as costs (i.e. money). CEA measures outcomes in some 'natural' or clinical unit, for example mmHg change in blood pressure, deaths prevented, or life years gained. CUA is a subset of CEA where outcomes are measured as utility, typically expressed in Quality Adjusted Life Years (QALYs) gained, where the 'quality adjustment' is based on some revealed or stated preference over different health states. Preferences for particular health states are elicited either directly, using tools such as the standard gamble, or indirectly via completion of generic or disease specific quality of life scales, with scores matched to a previously elicited utility (e.g. the Measuring the Value of Health (MVH) project which used the time trade-off approach to value the EQ5D health states²²).

CMA is a specific type of evaluation where the outcomes from the comparator interventions have been demonstrated to be equivalent, and therefore the cost-effective intervention is simply the least costly (as it dominates its comparator). However, due to uncertainty around the sample estimate of costs and effects from a trial, it has been argued that CMA is an inappropriate analytic technique except in analyses specifically powered to detect equivalence between comparators (the failure to detect a statistically significant difference

in effect does not mean there is no difference in effect).²³ A recent review of the use of CMA built on this further, finding that it biases estimates of uncertainty and hence value of information analyses.²⁴

1.2.2. The need for economic evaluation

In this section, I explain the principles of the market mechanism, why such a system may fail in the health care field, thus providing a necessary (but not sufficient) criterion for government intervention or planning to replace individual decision making. Economic evaluation is then a mechanism for assembling all the necessary information to inform those planning decisions.

Resources are finite. Therefore individuals and organisations are faced with decisions as to how to allocate their finite resources to best meet their personal or organisational objectives, whether to maximise revenue, profit or some less tangible concept of happiness or welfare. Making a decision to consume one good or service (or treatment) means the resources cannot be used to consume another good or service. The foregone benefit from that 'next best' good or service is termed the opportunity cost. Therefore making decisions in order to maximise some concept of welfare or happiness (or other maximand) is analogous to minimising the opportunity cost.

In a perfectly competitive market for any good or service, individuals make their choices over what to demand and supply at what prices and market equilibria will exist at which supply and demand are equal. Markets will therefore 'clear' and the outcome will be Pareto efficient (that is, it is not possible to make anyone else better off without another being worse off – see 1.2.3 below).²⁵ Such an outcome is only realised when the assumptions of the perfectly competitive market are met. These are:²⁶ (i) many (infinite) buyers and sellers such that no one individual is able to influence the price of the good or service; (ii) homogeneity of product, thus buyers have no preference for the good of one supplier over another; (iii) freedom of entry and exit to the market, such that there are no barriers preventing new sellers setting up or quitting the market; and (iv) perfect information about the market including perfect knowledge of competitor's prices. Where these assumptions are violated, market outcomes will not necessarily be Pareto efficient, therefore providing a necessary (but not sufficient) justification for government intervention to plan the allocation of those particular resources.

The healthcare sector is characterised by violations of almost all these conditions. For example, most health care systems have one or few major purchasers of health care

(insurance companies or state health care systems purchasing on behalf of their populations), thus operating in a monopsonistic market. The supply side is characterised by a relatively small number of operators: hospitals and pharmaceutical companies, thus the market is also monopolistic. Furthermore patents for pharmaceuticals deliberately grant monopoly power to a manufacturer for a period of time.ⁱ Examples of product differentiation exist, intentional or otherwise, for example in published differences in outcomes from various procedures between centres.²⁷ The barriers to entry to the health care market are substantial. For example, in most countries individual medics must be a licensed member of their respective professional body before being legally allowed to practice, a license only being granted following extensive training (often involving considerable start-up costs to fund education which may themselves act as a barrier to entry) and proof of competence. Potential entrants to the pharmaceutical industry are likewise faced with prohibitive start-up costs in proving to licensing bodies that their products are safe and effective and establishing manufacturing facilities to comply with relevant legislation. Finally, consumers of health care products and services (i.e. patients) frequently have limited information about the services they need. The particular nature of this informational issue is characterised in the Principal-Agent problem with asymmetry of information between the patient (principal) and their doctor (agent). Where the doctor is also the supplier of the service there is the incentive for supplier induced demand leading to over-consumption of health care.²⁸ Reimbursement mechanisms (such as fee-for-service) and the existence of third-party payers exacerbates this even further.

Two further particular characteristics of the health care sector also lead to market failure.²⁵ The first of these is uncertainty. Generally, people do not know when they will fall ill. The insurance market provides a solution to this, but insurance leads to problems of its own with consequent impacts on the efficient allocation of resources.ⁱⁱ The second characteristic is the existence of externalities (benefits or costs imposed on a third party for which no compensation is paid or received).²⁵

ⁱ The reason for this being due to the public good nature of information (non-rival, non-excludable and non-diminishable). Without government intervention and protection in the form of a patent, anyone could make use of the research required to develop a new drug and undercut the original manufacturer as they have not incurred the cost of obtaining that research information. The original developer would then never see a return on their investment. Realising this, the manufacturer would not invest in the research and thus the drug would not be developed in the first place.

ⁱⁱ Namely moral hazard (itself a consequence of informational asymmetry). This is where knowledge that an individual is insured against a particular event affects their behaviour such that they increase their risk of experiencing the event. For example, a driver may drive less carefully in the knowledge that he or she will be insured against a loss in the event of damaging their car. The same may be true for health insurance, for example partaking in extreme sporting activities or failing to adopt a healthy lifestyle.

A classic example in the health care field is that of immunisation: as an increasing proportion of the population is immunised, the risk of contracting the disease in the un-immunised decreases due to the lower prevalence of disease in the population ('herd immunity'). Thus ultimately they can free-ride off those who have been immunised. The rational action for each individual would therefore be to avoid the pain (and/or expense) of immunisation themselves and let the rest of the population be immunised. Logically this would result in no individuals being immunised, or under-immunisation of the population with a consequent burden of disease.ⁱⁱⁱ Another example of externalities associated with health care is the 'caring externality'.²⁹ This is explained in more detail in Section 1.2.3.

Thus allowing the free market to operate throughout the health care sector is unlikely to lead to an efficient outcome, with over-consumption of healthcare by some groups and under-consumption in others: it is highly likely that resources could be reallocated away from over-consumers to under-consumers without deterioration in the welfare of the over-consumers (i.e. Pareto improvements are possible). This in itself does not provide a sufficient justification for government intervention: planned systems are costly to implement with heavy information requirements. If this cost is greater than the efficiency loss observed in the market then society would be better off with the imperfect operation of markets.²⁵ Despite this there may be other justifications for government intervention, namely concerns for equity (Pareto efficiency makes no comment on the 'fairness' of an allocation of resources - see below), or consideration of health as sufficiently important for it to merit being a maximand in its own right.²⁵

In the absence of a well-functioning market, economic evaluation provides a means to inform policy makers and planners as to the costs and consequences of different courses of action, and, given some socially accepted values placed on those consequences elicited by some legitimate process, to make recommendations as to which course of action best improves the welfare of society.

1.2.3. Welfarism and the "extra-" or "non-welfarist" foundations of economic evaluation.

Economics can be broadly divided into positive and normative methods. Positive economics is concerned with measurement and prediction; the statement of facts and hypotheses which can be formally tested. Normative economics is concerned with the relative

ⁱⁱⁱ This may not actually be a stable equilibrium because as the prevalence of disease rises, individuals will assess their own risk and the expected consequence of not immunising against the individual expected consequence of immunising, resulting in some individuals desiring immunisation, thus lowering the risk in the population and hence altering individual decisions again.

desirability of alternative states of the world, and is thus intrinsically concerned with subjective valuations and relative preferences.³⁰ The welfarist and extra-welfarist / 'non-welfarist' approaches can be considered two branches of normative economics.³⁰

In this section I summarise the key principles of welfarism, followed by a criticism of the approach, focusing on the arguments of Sen and Culyer, introducing the ideas of 'extra' or 'non' –welfarism on which current principles of economic evaluation are based.

Welfarism

Welfarism has been defined as "the systematic analysis of the social desirability of any set of arrangements, for example a state of the world or allocation of resources, solely in terms of the utility obtained by individuals".³⁰ Utility is the satisfaction or pleasure an individual obtains from the consumption of goods and services.^{iv} It is entirely individualist and consequentialist: individuals (defined either as single consumers or households) are deemed the best judge of what is good for them, and will therefore make consumption decisions in order to maximise their own utility. Furthermore the welfare of one individual is assumed not to depend on another individual's welfare. It is consequentialist as utility is assumed derived only from the consumption of goods and services. The method or process by which those goods and services are obtained is assumed to yield no utility (or disutility) in itself.^v

A key principle to judging whether a particular allocation of resources between members of society is 'optimal' or 'more desirable' compared with another was suggested by Pareto (1906).³⁵ A Pareto improvement in the allocation of resources is one where either the utility of all individuals is increased (known as a 'weak Pareto improvement' due to the weaker value judgement implied), or increases the utility of at least one individual without decreasing the utility of anyone else (a strong Pareto improvement). An allocation such that it is not possible to improve the utility of one individual without reducing another's is known as a Pareto efficient allocation.

There are thus an infinite^{vi} number of allocations which may be considered Pareto efficient. It is critical to note that Pareto's rule provides no information as to which of those allocations may be the most socially desirable (or specifically, distributionally equitable). For example, a Pareto efficient allocation may be one with vast inequality. Given a grossly

^{iv} Utilitarianism may be considered to have its origins in the work of the Greek philosopher Epicurus, but Jeremy Bentham³¹ is most often credited with development of the classical utilitarian approach

^v This is termed process utility.³² Empirical evidence in the health sector would seem to indicate the assumption of zero process utility is false.³²⁻³⁴

^{vi} 'Infinite' if goods and services are perfectly divisible. A 'large number' if resources are indivisible at some level.

unequal starting point, a (strong) Pareto improvement would be one that makes the rich better off whilst leaving the poor unchanged. This may be considered socially undesirable due to conflict with notions of equity.

A strength of Pareto's rule is that it is unambiguous as it avoids interpersonal comparisons of utility: where one person experiences a gain in utility at the expense of another. Arguably it achieves this simply by 'ducking the question': inter-personal utility comparisons are somewhat controversial³⁶ but unavoidable in any meaningful decision making process where there is any regard for equity. Therefore, extensions to Pareto's rule have been developed, most notably by Kaldor³⁷ and Hicks³⁸, based around compensation tests.^{vii}

Kaldor³⁷ suggested that following a policy change, if the winners could compensate the losers financially such that the losers returned to their original utility level, and the winners were still better off than before, then the policy change represents a net gain to societal welfare. Conversely, Hicks³⁸ suggested that the potential losers from a policy change could 'bribe' the winners into rejecting the change by an amount equivalent to their loss. If the winners would be better off rejecting the bribe, then the policy change would represent a net gain to societal welfare. The Kaldor-Hicks compensation tests thus allow for interpersonal comparisons of utility by equating utility to monetary trade-offs.

It is important to note that the tests do not require compensation to be actually paid to the losers in any policy, and thus whilst the tests extend Pareto's work to allow interpersonal comparisons of utility, and thus extend the set of comparable resource allocations, they still have very little to offer regarding the social desirability (with respect to equity) of any particular allocation or impact of a policy change.

In order to rank all the efficient allocations in terms of social desirability, a measure of the overall societal welfare from each allocation is required. This is termed the social welfare function, the most general form of this (the Bergson or Bergson-Samuelson function) defines social welfare as some function of individual utilities. Different authors have proposed different forms for the function. For example a strict utilitarian function would define social welfare as the sum of individual utilities. Alternatively a weighted utilitarian function allows for diminishing marginal returns to utility (society places less weight on gains to those with high utility). Rawls³⁹ and Nash (cited in³⁰) also proposed specific functional forms, based on maximising the utility of the worst off and the product of individual utilities respectively.

^{vii} Both published their ideas simultaneously in the same issue of *The Economic Journal* in 1939.

Criticism of Welfarism

The utilitarian social welfare function and variants thereof have come under criticism most notably from Sen.⁴⁰ He is critical of utility as an appropriate measure of well-being,⁴¹ suggesting that the pleasure obtained from consumption of goods and services does not sufficiently capture all that is relevant in the social welfare function. He distinguished 'functionings' and 'capabilities' as relevant attributes on the basis that not all individuals have the same capacity to obtain pleasure from a given allocation of goods and services. Whilst on a superficial level this may be explained away by differences in taste and preferences, Sen was concerned with the equity implications of utilitarianism: he distinguishes 'physical condition neglect' and 'valuation neglect'.⁴¹ Physical condition neglect refers to a situation where one may be in a poor physical state yet be happy and content due to adjustment of expectations and "[taking] pleasure in small mercies".⁴¹ Valuation neglect refers to utilitarianism's taking 'at face value' what people 'manage to desire' and thus ignores (or overestimates) the welfare of those "who are too subdued or broken to have the courage to desire much" (Sen 1985⁴¹ cited in Mooney & Russell⁴²). The consequence therefore would be that as people adjust their expectations to their current situation they may describe themselves as 'happy' and 'content', without realising their situation could be better, thus entrenching established social hierarchies and inequities. On the other hand, this raises interesting questions for policy makers. For example if those in poor health do not 'manage to desire' a better state of health,^{viii} then is it 'right' or appropriate to 'adjust' for their 'inadequate expression of desire'?⁴²

So Sen's^{40 41} argument was that concepts other than utility (i.e. functionings and capabilities) should comprise the social welfare function, indeed going so far as rejecting the notion of utility completely. His approach may therefore be termed 'non-welfarist'.⁴² Culyer⁴³ was the first economist to challenge the appropriateness of the welfarist approach within healthcare, arguing too that welfare comprises more than a function of the utilities of individuals derived from consumption of goods and services. He broadened out Sen's ideas on capabilities^{ix} to general 'characteristics of people', including their baseline health status (e.g. genetic inheritance), 'moral worth and deservingness', whether they are in pain or are stigmatised by society.⁴⁴ In particular he introduced the notion of the 'caring externality' where a person's utility may be a function not only of their own consumption of goods and

^{viii} For example, developing type 2 diabetes and its complications being seen as an inevitable part of aging in an area of high obesity (which tends to be associated with poverty and poorer socioeconomic status and education).

^{ix} Mooney and Russell⁴² dispute Culyer's claim to be building on Sen's ideas on the basis that whilst Culyer advocates including additional factors as well as utility into the social welfare function ("extra-welfarism"), Sen advocates omitting utility entirely in favour of capabilities and functionings ("non-welfarism").

services, but the utility of another too (and therefore their consumption of goods and services).²⁹ Brouwer and colleagues⁴⁵ clarified the distinction between the welfarist and extra-welfarist schools of thought, suggesting key areas of difference were in the outcomes considered relevant in an evaluation, the source of valuation of the outcomes, the approach to weighting those outcomes and the necessity for interpersonal comparisons of utility.

From a pragmatic or applied standpoint, inclusion of the 'extra-welfarist' elements (e.g. health) makes analysis somewhat complicated: health and utility are very unlikely to be independent as individuals may get utility from enjoying a particular state of health *per se*. Or perhaps (more specifically), health can be seen as an enabler of the consumption of utility-bearing goods and services. Furthermore Brouwer & Coopmanschap⁴⁶ argue that embedding CUA within a welfarist framework is impossible, due to the limited scope of QALYs and assumptions that they are comparable between persons and can be summed meaningfully.

In practice, attempts at measuring utility in applied economic evaluation are completely abandoned in favour of attempts to measure and maximise health itself. The development of economic evaluation (which should therefore perhaps be termed non-welfarist rather than extra-welfarist) is thus a pragmatic solution to applying the principles of economics to health care decision making, rather than being firmly rooted in a particular theoretical framework.

1.3. Conducting an economic evaluation

In Section 1.2.2 I argued that market failure, concern for equity and an intrinsic valuation of health for its own sake provide justification for government intervention in the health care sector. Any planning mechanism has considerable informational requirements to substitute for the very large number of individual supply and demand decisions that collectively constitute Smith's 'invisible hand'.⁴⁷ Economic evaluation is thus a means to structure this information in a form which can assist policy makers as to whether a new technology represents a net benefit or harm to society.

In this section I introduce the general methods for economic evaluation. There are two major 'implementations' of economic evaluation: those conducted alongside a clinical trial (often referred to as 'piggy-backed' studies), and decision models combining 'all relevant evidence' into a single framework.

Drummond et al. define ten elements to a sound economic evaluation⁴⁸ and these provide a structured approach to designing an analysis.

The elements are:

1. Define the study question in an answerable form
2. Describe each comparator comprehensively
3. Establish effectiveness of each comparator
4. Identify all important costs and consequences from each comparator
5. Measure all costs and consequences accurately in appropriate physical units
6. Value all costs and consequences credibly
7. Adjust costs and consequences for differential timing
8. Undertake an incremental analysis
9. Analysis of uncertainty
10. Write-up and discussion, including all issues of concern to the target audience.

These are considered in turn in sections 1.3.1-1.3.10 below.

1.3.1. Define the study question

Defining a study question in an answerable form is the critical first step. This must specify what is being compared with what (i.e. statement of comparators – see Section 1.3.2 below), in what population and from whose perspective the analysis is being conducted, for example society, the health care payer or individual hospital.

Cost-effectiveness is a subjective concept in the sense that it is dependent upon the perspective adopted; an intervention may be very cost-effective for one individual or organisation, but not from another simply because that individual or organisation's budget is not faced with all the relevant costs.^x Because of this, the preferred perspective for an economic evaluation is that of all society. This is the broadest point of view and allows judgement as to whether there is a net improvement to the whole of society from a proposed change, rather than simply a reallocation of resources between budgets.⁴⁹

However, the broader the perspective, the more difficult it can be to obtain good quality estimates of resource use. For example, relatively good data may be available for an analysis from the perspective of a single centre or the health care payer. A broader perspective may require collection of data from multiple agencies with differing record systems, requiring additional research effort to collect and potentially straining research resources. Finally methodological disagreements exist as to the appropriate approaches to valuing morbidity-related lost productivity (required for a true societal perspective, see Section 1.3.6).

In the absence of a societal perspective, typical alternatives are the health sector or an individual hospital. A different viewpoint of an analysis has the capacity to alter the results. For example, a recent trial of a befriending intervention for carers of people with dementia⁵⁰ suggested it was very unlikely to be cost-effective from a societal point of view. However, when considering the voluntary sector perspective alone, there was a high probability of cost-effectiveness. This was simply because the benefits of the intervention were the same (improvement in quality of life of carer), but the voluntary sector were only faced with a small proportion of the total costs. In a further example, Weisbrod and colleagues⁵¹ (cited in Drummond et al. 2005⁴⁸) evaluated the cost-benefit of a community-orientated programme for mental illness patients. From the perspective of the agency providing the service, the community-based programme was more expensive than conventional hospital-based care. However, when including broader costs such as those falling on other health care organisations and law enforcement, the increment was reduced substantially, and when a

^x The extreme example is all NHS care from the perspective of the individual. Generally speaking, an individual will pay the same (via tax contributions) for whatever treatment he or she receives (therefore the incremental cost of a treatment decision is zero), but receives the (incremental) benefit (of treatment over no treatment). Therefore the most cost-effective treatment from the individual patient's perspective is the most effective as a third party (the state) bears the cost differential. Another example would be the case of a treatment appearing to be cost-effective from the perspective of the NHS, but ignoring additional costs placed on other public sector agencies (e.g. social services) or patient out of pocket costs.

societal perspective is adopted, specifically including provision of food and shelter and productivity losses, the community-based programme was less expensive than conventional hospital-based care.

1.3.2. Describe each comparator

A clear description of the comparators is essential in order for decision makers to be able to implement the results of the analysis should they so wish. This is especially the case with 'complex interventions' (that is, where interventions are programmes of care rather than simple drugs or procedures). There are several reasons for this. Firstly, should a policy maker wish to adopt a particular intervention an adequate description is required in order to replicate it appropriately. Secondly, in order to generalise the results to a particular setting, sufficient information must be provided about the control in order for decision makers to judge how different the described protocols are to standard practice in their setting, and thus whether the intervention is likely to yield the incremental benefits (and incur the incremental cost) in their own setting. Thirdly, a full description of the intervention allows the decision maker to subjectively judge whether or not an adapted version suitable for their own setting would be more or less effective than the one described.

1.3.3. Establish the effectiveness of each comparator

Effectiveness evidence comprises either disease-specific or generic outcomes such as mmHg reduction in blood pressure and life expectancy / life years gained, or generic and disease specific health related quality of life scores. By this definition, QALYs are not considered an effectiveness measure: they are generated from a summary valuation of the results of a health related quality of life tool.

The method by which the effectiveness of the comparators is established depends on the implementation of the economic evaluation. For example, if the analysis is conducted alongside a clinical trial, effectiveness evidence will be from that trial. If the analysis is a decision model, ideally the effectiveness data will be obtained from a good quality systematic review of all relevant evidence. In both cases, particular attention should be paid to the risk of bias in the estimate of the relative treatment effect: a well-designed double-blind randomised controlled trial is considered less prone to bias than observational studies. In the case of a systematic review informing a decision model, reasons for inclusion or omission of particular source of evidence must be clearly stated, and appropriate techniques used to meta-analyse the results.

1.3.4. Identify all important costs and consequences

This and the following Sections 1.3.5 – 1.3.7 are relevant to both costs and outcomes in an economic evaluation. However, my focus will be on the cost side as this is the focus of problem addressed in my thesis.

Given a broadly technically efficient allocation of resources, the cost of a decision (and ultimately the only relevant concept of cost) is the opportunity cost. As I stated in Section 1.2.2, this is defined as the value of the benefits foregone from the next best alternative use of the resources employed in a particular use. In other words, given two alternative courses of action, A and B, the opportunity cost of choosing A is the value of the benefit foregone by not choosing B. For example, the NHS may choose to fund a particular novel cancer therapy. The opportunity cost is the benefit foregone from the other services which as a consequence are either withdrawn or could have been introduced or expanded instead.

In a perfectly competitive market equilibrium, the opportunity cost of A will equal the value of the resources shifted from B in order to achieve it.^{xi} To quantify this it is necessary to identify, measure and value each type of resource to be shifted.

Measurement and valuation are considered in Sections 1.3.5 and 1.3.6 respectively.

The first step is to define the categories of cost to be included in an evaluation. This will be dependent on the perspective chosen (see Section 1.3.1), and may include:^{48 52}

- health care resources directly relevant to the intervention or comparator(s);
- non-health care resources directly relevant to the intervention or comparator(s);
- patient and family out of pocket costs;
- patient and informal caregivers time including productivity loss;^{xii}

After defining the categories, the individual items within them need to be specified. For example health care resources would comprise *inter alia* all drugs (including study drug if applicable), surgical procedures, hospital ‘hotel’ costs and overheads, doctor, nurse and other specialist time.

^{xi} As stated in section 1.2.2, the health care sector violates many of the assumptions required for the market solution to achieve an efficient outcome. Where price is not equal to the opportunity cost of a good or service and the valuation of goods and services for which there is no market are considered in Section 1.3.6

^{xii} Traditionally, health economists divided costs into direct, indirect and intangible costs. However, due to confusion with other professions, this was later re-classified into the four categories listed.

The process of identification of resource use items is essentially a question of specifying the production function for the intervention of interest. However, it may be possible to exclude certain items, dependent on the purpose of the costing exercise: if the analysis is to be used solely to compare the interventions of interest in the economic evaluation, it is acceptable to exclude cost items common to all treatment arms.

For example, a cost-utility modelling study of mycophenolate mofetil (MMF) versus intravenous cyclophosphamide (IVC) in the treatment of lupus nephritis specifically excluded a number of additional drugs and outpatient visits from the analysis as these were identical for patients in both arms of the model.⁵³ This was acceptable for the comparison of the two alternatives considered due to the limited time horizon of the model,^{xiii} but if the cost estimates for treating a patient with MMF were taken 'as is' and incorporated into a subsequent analysis of a third, substantially different treatment, the cost estimate will be biased. Similarly, a finance department should not base budgeting decisions on the mean cost per MMF patient presented in the study, or percentage difference in cost: it would be necessary to include the cost of the additional items not included. Essentially, the economic evaluation was designed to demonstrate the incremental costs and outcomes of transferring patients from existing cyclophosphamide therapy to the newer MMF therapy, and therefore should be used to answer that question alone. Use beyond that purpose should be exercised with caution.

A second reason for excluding cost items is on the grounds that they will have zero, or very little, impact on the results. If they account for only a small proportion of the total cost (or rather, the cost increment between two interventions), then they will not impact on the study results, and it may be possible to safely exclude them. The limitation of this is that it is often difficult to know *a priori* the impact a particular cost item will have on the results, and thus whether it is worth going to the effort (and expense) of collection.

Finally, within the context of a clinical trial, it may be desirable to exclude 'protocol driven costs'. These are costs of activity incurred solely for the purpose of data collection for a clinical trial which would not be observed in routine practice (e.g. attendance at clinic for additional diagnostics or completion of case record forms). However, caution should be exercised in this approach as where the trial mandates more intensive monitoring of

^{xiii} The model was restricted to the management of an initial disease 'flare', rather than modelling a life time horizon. If the model had considered a longer time horizon it may have been necessary to include these cost items as if different treatments were associated with different rates of 'flare', the incremental lifetime cost of these elements would not be zero.

patients than would otherwise be observed, additional clinical events and outcomes may be detected than would be seen in routine care. This may then affect their subsequent care, incurring costs (and outcomes) that would differ from that seen in routine care. It may therefore be impossible to truly disentangle protocol-driven costs and outcomes from treatment effects, and inclusion of all cost items may be a more conservative approach.⁵⁴

1.3.5. Measure all costs and consequences

The next step is to define the approach to measuring resource use and outcomes. In the case of a clinical trial, health care resource use measurement is frequently incorporated as a part of the case report form (CRF), either as a resource use questionnaire for the patient to fill in, or for completion by study nurses or researchers as appropriate. Alternative approaches include retrospective database analysis, for example using primary records to measure primary or secondary care use,^{xiv} or skipping the resource measurement step entirely and extracting cost estimates directly from hospital billing records to patients or their third party payers (more common in insurance based systems such as the USA and parts of Europe).

Specific issues in resource use measurement in clinical trials are the length of time for which costs (and indeed outcomes) should be tracked, the handling of overhead and capital costs and the measurement of productivity costs.

Follow-up length

The guiding principle governing follow-up length in a clinical trial is that it should be of sufficient time to observe outcomes in order to reach conclusions as to the incremental efficacy or effectiveness of the comparators (whether that be survival, quality adjusted life years or some other interim outcome). For an economic evaluation, this should be interpreted as sufficient time to observe outcomes in order to reach conclusions as to the incremental cost-effectiveness of the comparators, that is, consideration of both cost and outcomes.⁵⁴

Frequently, resources or logistics do not permit a long enough follow-up time to observe all impacts on final outcomes and cost, so an intermediate outcome is selected in place. For example, the ideal outcome for an anti-osteoporosis drug is prevention of fractures (or indeed, ultimately the maximisation of quality and length of life through the prevention of fractures). However, a very long follow-up period may be required in order to observe any

^{xiv} Although the reliability of using primary care records to measure secondary care use in the English NHS has been questioned.⁵⁵

difference in fracture rates. Therefore an intermediate outcome of bone mineral density may be used as a primary endpoint in a clinical trial.^{xv}

Allocation of overheads and capital goods

Overheads are costs such as buildings, administration, heating and power which are not readily allocated on a per-patient or per-procedure level. The preferred approach to handling these is to allocate them directly to final costs,⁵² but this is a pragmatic solution due to lack of detailed data within typical accounting systems rather than being theoretically desirable. Within the NHS, the national tariff (based on the national schedule of reference costs) incorporates overheads within an average cost set for a given procedure.⁵⁶

Measurement of productivity costs

Productivity costs are the value of lost economic output to society as a result of illness or premature mortality. There is some disagreement as to whether it is appropriate to include productivity costs in a cost-effectiveness or cost-utility analysis. On the grounds of efficiency, it is argued that in an evaluation with a societal perspective, it is proper to include all costs borne by all parties. This would imply productivity costs should be included. However Gerard & Mooney⁵⁷ argued that as the only outcome of interest in a cost-effectiveness analysis is health related, it is logical to restrict the costs similarly. Koopmanschap & Rutten⁵⁸ countered this by pointing out that budgets were somewhat arbitrary divisions in resource allocations, and therefore resource allocations in a non-health budget may have an impact on health (and vice versa).

On the grounds of equity, an argument for excluding productivity costs is that they result in a bias towards those who are most economically productive (or who earn the highest salaries), which may be considered inequitable and therefore socially undesirable. However, if those patients are brought back into the workforce as a priority, they will then be paying taxes which leads to increases in resources available for health (as well as other) budgets. Anecdotal evidence suggests society may be willing to prioritise some workers above others to receive health care (e.g. prioritising NHS employees over the elderly for bird 'flu vaccines), and research evidence indicates that society as a whole may place greatest value on those of productive and child-bearing age, and least on the elderly (although not all empirical findings support this conclusion).^{59,60}

^{xv} This provides a justification for decision modelling to connect such intermediate outcomes to final outcomes and cost. See Section 1.5.1

Sculpher⁶¹ identified a number of questionnaires designed for use in economic evaluations conducted alongside clinical trials, including the Work Productivity and Activity Impairment questionnaire,⁶² and the Health and Labour Questionnaire.⁶³ Both aimed to measure absenteeism and ‘presenteeism’ (reduced productivity whilst in the workplace) attributable to ill health, whilst the latter⁶³ also included unpaid areas such as ability to do housework, shopping, childcare and general household maintenance.

1.3.6. Value all costs and consequences

The most obvious means of valuing resource use is to use price weights (i.e market prices). Total cost is then simply quantity of a resource multiplied by its unit price weight.⁶⁴ However, this assumes a constant marginal cost, and there remain questions as to whether national or centre specific weights should be applied. Use of national price weights enhances the external validity of the study (that is, the results will reflect whether or not an intervention is cost-effective on average across all centres within a country), but at the expense of internal validity (the results may not reflect the true cost-effectiveness of an intervention in the centre(s) in which the study was carried out), whilst the reverse is true of centre specific weights. The appropriate method is contingent on the perspective of the economic evaluation.

Other issues around the valuation of resource use relate to the valuation of goods and services for which there is no market, the valuation of goods and services in the presence of imperfect markets / market failure and valuing productivity changes.

Valuation of goods and services for which there is no market

In some cases, there may not be an appropriate market price to use to value a resource item. This is particularly a problem in valuing informal care-giving and foregone leisure time. For economic evaluations in areas where informal care is of considerable importance in the overall care of the patient (for example in the care of Alzheimer’s patients), omitting informal care will result in underestimation of the total resource consequences. However, there are complications in measuring and allocating the *extra* time input required as a result of e.g. Alzheimer’s over conventional household chores. For example, general supervision and surveillance may be carried out alongside other household activities, or a partner may cook meals for his/her partner, irrespective of whether the other partner was able to perform these activities.

Brouwer and colleagues identify three major approaches to valuing informal care time.⁵² These are the market price method,⁶⁵ the reservation wage method⁶⁶ and a method incorporating individual choice and societal costs of time.⁶⁷

The market price method⁶⁵ attaches a shadow price to the time spent caring at the rate a private carer would charge for the same service. However, there is a danger of overestimating the cost of this as an ‘amateur’ carer may take more time to do a specific task than a professional. To correct for this, the time spent caring can be adjusted for the expected length of time it would take a professional to achieve a specific task. Measuring this, however, increases the analytic burden of the study and does not equate to the opportunity cost of the carer’s time.

The reservation wage method⁶⁶ differs from the market price method in that time is valued at the wage rate the carer could have earned had he/she been in employment. Arguably, this approach, focussing on the input rather than output more closely reflects the true opportunity cost of caring, and so is the more appropriate for a societal economic evaluation.

The Brouwer approach⁶⁷ uses valuation techniques which vary according to the type of time sacrificed (paid work, unpaid work and leisure time). Paid work foregone should be valued using the friction cost approach (an approach which takes account that workers off sick temporarily may be able to catch up following return to work or for spare capacity elsewhere in the firm to cover the temporary absence; see “valuing productivity costs” below),⁶⁸ unpaid work (e.g. housework) valued at the shadow price of a housekeeper, whereas leisure time should be valued in terms of reduced quality of life rather than increased costs.

Valuation of goods and services for which there is an imperfect market

A perfectly competitive market results in prices for resources that represent the opportunity cost of those resources. In health systems where hospital reimbursement is based on accurate records of activity, billing records are an attractive means of valuing resource use. However, depending upon the perspective of the analysis, it may be necessary to convert charges to costs using a price to charge ratio, removing any surplus added on top of cost to more accurately reflect the opportunity cost of the activity to society. Although it should be noted that if the perspective of the analysis is the health insurance company, it is more appropriate to use charges rather than provider costs as these represent the opportunity cost to the insurance company.

Valuing productivity costs

Two main approaches are taken to the valuation of productivity costs, namely the human capital⁶⁹ and friction cost⁶⁸ approaches.

The human capital (HC) approach values productivity loss at gross earnings rates. A shadow price is usually assigned for those not in paid employment at either a market price⁶⁵ or reservation wage rate⁶⁶ (described in “valuation of goods and services for which there is no market”, above). Koopmanschap and colleagues⁶⁸ argued this approach was too crude and could grossly over-value lost productivity as for shorter term absences, colleagues may be able to cover absent employees, or they may themselves be able to recover their productivity on return to work. The friction cost method takes these situations into account, defining four possible outcomes when an employee is absent from work:

1. No change in productivity: the lost work can be made up by the employee on his/her return to work or covered by colleagues with no extra cost.
2. Productivity is unaffected but costs increase: where the absence is covered by colleagues working overtime or hiring of a temporary worker.
3. Productivity falls as a result of the absence
4. Productivity falls and costs are higher due to replacement by less experienced temporary workers / workers who are less familiar with the job than the incumbent.

Each of these situations has different implications for valuation of productivity loss: whereas the HC approach simply assumes situation (3), situations (1) and (2) may result in lower cost estimates than the HC approach, whilst situation (4) may lead to higher cost estimates.

1.3.7. Adjust costs and consequences for differential timing

The concept of discounting allows for the differential timing of costs (and effects). It is important to distinguish between inflation, interest rates and time preference. Inflation is the rate at which nominal prices increase; interest is a return paid on capital; whilst time preference refers to the concept that generally speaking, even in a world with zero inflation and zero interest rate, people still prefer to receive a benefit today rather than the same benefit tomorrow.

There are a number of reasons why an individual may have an intrinsic preference for benefits today rather than in the future, i.e. a positive rate of time preference. These include a degree of ‘myopia’ where living for today is intrinsically preferred to planning for

an uncertain future. Secondly, where wealth generally rises over time, the marginal value of a pound today is higher than a pound tomorrow as it represents a larger proportion of the individual's wealth today.⁷⁰

1.3.8. Undertake an incremental analysis

The results of an economic evaluation, whether trial or model based, will be estimates of the mean (expected) cost of each strategy and mean outcome (e.g. QALYs gained). These are combined into the Incremental Cost Effectiveness Ratio, defined as the difference in cost divided by the difference in outcome (Equation [1-1]) and is an expression of how much it costs to 'buy' one additional unit of outcome by moving from the old treatment to the new. If this is less than some threshold, λ (representing the maximum value or willingness to pay of the decision maker for a unit of the outcome) then the new intervention is considered cost-effective and should be adopted. For example, if the outcome is QALYs, the National Institute for Health and Care Excellence (NICE) recommends a threshold (λ) of £20,000 - £30,000 as a maximum willingness to pay for a QALY.

$$\frac{C_2 - C_1}{E_2 - E_1} \leq \lambda \quad [1-1]$$

From ICER to Incremental Net Benefit

Rearranging [1-1], the ICER can be expressed as incremental net benefit (INB), which can be interpreted as "adopt the new technology if we value the extra health gain greater than the loss" (Equation [1-2]).^{xvi} Note that the decision threshold, λ , is now incorporated within the equation. As λ is subjective to the decision maker, and unknown to the analyst, INB is usually calculated and plotted for a range of values of λ .^{xvii} If the INB is greater than zero at the decision maker's preferred threshold, then the decision maker should adopt the technology.

^{xvi} This is the incremental net monetary benefit. The equation can be rearranged into the incremental net health benefit by simply dividing both sides by λ , which makes the trade-offs in the health gain between different groups of patients particularly transparent, but the equation can cause divide by zero errors and thus net monetary benefit is preferable to work with.

^{xvii} A number of studies have been undertaken to elicit an empirical estimate of λ . A retrospective analysis of previous NICE decisions in 2004⁷¹ found that the revealed threshold may be higher than the stated £20,000-£30,000 claimed in its guidance.⁷² More recently, the 'Societal Value of a QALY' (SVQ) project⁵⁹ explored the feasibility of eliciting a valuation of a QALY using a number of willingness to pay methods. They used several plausible methods which in some cases yielded values close to the value used by NICE. However other methods yielded values of up to £250,000. A report by the University of York focused on the opportunity cost of services displaced within the health sector by new interventions (rather than by any other good or service as is implied in the WTP approach), yielding a 'best' estimate for the threshold of just under £13,000 per QALY gained.⁷³ Thus there is considerable variation in what constitutes an 'appropriate' threshold.

$$\lambda(E_2 - E_1) - (C_2 - C_1) \geq 0 \quad [1-2]$$

Note that the incremental net benefit will be positive if the net benefit of the new technology is greater than the net benefit from the old. Furthermore where there are more than two options, the incremental net benefit of the option with the highest net benefit will always be positive compared with any of the other options. Therefore the decision rule can be generalised to comparisons of multiple options as simply adopting the technology with the highest net benefit (Equation [1-3]).

$$D = \operatorname{argmax}_i (\lambda E_i - C_i) \quad [1-3]$$

Where D is technology i with the highest net benefit.

1.3.9. Analysis of uncertainty

The analysis of uncertainty is critical in order to assess the degree of confidence in the results of an economic evaluation and to test the robustness of conclusions, although the implications of this for the adoption decision are discussed in Section 1.5.2. It is important to distinguish between individual variability, heterogeneity and uncertainty.⁷⁴ Individual variability (also known as first order uncertainty) is the uncertainty of outcome associated with an individual. The spread of individual observations around the central tendency (i.e. the mean) of a sample with the same characteristics (i.e. from the same population) is measured with the standard deviation. This is a constant and a feature of the population itself and cannot be reduced with more information. Heterogeneity relates to different characteristics between patients which can be explained such as age and gender.

Variability and heterogeneity are not the subject of the analysis of uncertainty. The uncertainty under consideration is that in estimates of means (second order uncertainty), measured by the standard error of the mean. Parameter uncertainty should furthermore be distinguished from decision uncertainty: parameter uncertainty relates to the standard error of the input parameters of an economic evaluation, such as treatment effect, utility estimates and resource quantities. Uncertainty in these parameters is propagated through a decision model or mathematical formula to estimate decision uncertainty, i.e. standard error of the output parameter, namely the incremental cost effectiveness ratio or incremental net benefit.

In Bayesian statistics, the standard error around a parameter (or rather, the probability density function) can be interpreted as relative degrees of belief about different values of a parameter: a low standard error representing little uncertainty and thus a high degree of

belief that the parameter value is the ‘true’ value and vice versa. Initial beliefs about likely values form the prior probability distribution of a parameter (that is, plausible values of the mean of some measurement such as health gain from an intervention). Data can be gathered, which themselves will generate a sample mean and associated uncertainty expressed in the standard error of that mean (the distribution being termed the likelihood function). Bayes’ rule then combines the prior and the likelihood together (that is, the prior is updated with the likelihood) to form the posterior probability distribution. There are a number of approaches to analysis of uncertainty in an economic evaluation, determined to a certain extent by the implementation (i.e. whether alongside a clinical trial or in a decision model context).

Economic evaluations alongside clinical trials can be analysed using traditional statistical hypothesis tests such as Student’s T-test, which quantifies the probability of observing two sample means differing by at least as much as that observed in an experiment given an assumption of no difference in the population means. If this is sufficiently small then the assumption (termed the null hypothesis) is rejected.⁷⁵ These are conducted on the output parameters of the experiment: typically cost or outcome per patient in each arm, or (preferably) cost-effectiveness (most easily measured and tested by the net benefit per patient). However, there are sound methodological and practical reasons why this may be inappropriate. The methodological reasons are discussed in Section 1.5.2. The practical reasons are that studies are very rarely powered to detect a difference in cost-effectiveness, therefore there may be an unacceptably high likelihood of Type I error (or indeed, an unknown likelihood). Given the skewed nature of cost data, the sample size required to adequately power on cost-effectiveness may be unfeasibly high (although this is an empirical question). A review of the statistical analysis and interpretation of cost (rather than cost-effectiveness) data reported in randomised controlled trials published up to 1998⁷⁶ found that 56% of 45 trials reported results of statistical tests or some measure of precision of the comparison between groups, and according to the authors, only 36% of papers supplied conclusions which were justified on the basis of the results presented. Furthermore, none of the trials reported sample size calculations based on cost.

Economic evaluations based on decision models can be analysed using one or multi-way sensitivity analysis. Conceptually, the difference between the analysis of trial data and decision models is that the analysis of the latter focuses on the impact of uncertainty in model inputs (such as probabilities, health state utilities and resource quantities) on the outputs (i.e. the cost, outcome or net benefit per patient). A one-way sensitivity analysis

varies one input parameter in a model at a time, recalculating the ICER or INB for each value of that parameter. Likewise a multi-way sensitivity analysis varies two or more parameters simultaneously to explore their combined effect on the ICER. Typically, the results can be presented as either a threshold analysis (a table stating the maximum and/or minimum tolerable values for the parameters for the conclusions to be valid), or as a tornado diagram⁷⁷ (ordering parameters from most to least sensitive in terms of their impact on the ICER/INB).

The advantage of one-way sensitivity analysis is that it is simple to implement and to interpret. The disadvantages are that parameters are varied individually within arbitrary ranges: no account is taken of the true nature of uncertainty (that is, the probability distribution of the parameter where some values are more likely than others) and any interactions between parameters are not observed. Whilst multi-way sensitivity analysis does accommodate the interactions to a certain extent, interpretation becomes increasingly difficult as the number of parameters increases. For example, a two-way sensitivity analysis will define an 'acceptable set' of values for both parameters at which the ICER is within the decision threshold. As this increases to three or more parameters, the acceptable set becomes more complex to define and interpret.

The alternative is to analyse uncertainty in economic evaluations with probabilistic sensitivity analysis. The technique applies equally to analysis of economic evaluations conducted alongside clinical trials and decision models, albeit with some minor differences. I first explain the technique with a decision model before outlining the differences with clinical trial data.

Probabilistic sensitivity analysis of a decision model requires elicitation of a probability distribution for each input parameter. A Monte Carlo simulation approach is usually adopted where a value for each parameter is drawn from its respective distribution and the set plugged into the model. The resulting costs, outcomes, increments and ICER and/or INB are then recorded. This process is repeated many times to build up an empirical distribution around the ICER or INB and propagates uncertainty in input parameters into uncertainty in the ICER / INB (decision uncertainty).

It is important to note that each simulation represents a 'possible state of the world', but which state is ultimately realised is unknown. The input parameters therefore represent beliefs about plausible values of the parameter of interest in the Bayesian statistical sense,

rather than the long-run relative frequency with which particular sampled value may be observed in the future (the frequentist statistical interpretation).

Bootstrapping of clinical trial data simply requires resampling of the original data with replacement, and new mean costs and consequences (e.g. QALYs) calculated. Repeating this many times builds up the empirical distribution of mean costs and QALYs, and hence incremental net benefit. As the actual data points are resampled, no assumptions are required as to the distributional form of the data. This approach is therefore known as a non-parametric bootstrap.

1.3.10. Write-up and discussion

When interpreting the results of an economic evaluation it is important to evaluate whether the analysis has incorporated all relevant considerations, rather than unquestioningly accepting the results. Decision making is a complex process requiring the weighing up of many competing objectives, thus whilst technical solutions can be very appealing it is important not to accept their results uncritically.^{78 79} Thus they must be considered inputs into the decision making process and not the decision in itself.

If an economic evaluation is considered to be sufficiently well conducted and incorporate all relevant considerations, there is still the issue of how a decision should be made given the results.

The interpretation of the results of clinical trials are usually judged on whether, given an assumption of no difference and the experiment were to be repeated many times, there would be a sufficiently low probability of observing a treatment difference at least as big as that observed in the trial, such that chance could be ruled out as an explanation. This is hypothesis testing and is based on a frequentist interpretation of probability, although this is commonly (and erroneously) given a Bayesian statistical interpretation as the probability that an hypothesis is correct.

There are a number of criticisms of this approach, not least that the choice of cut-off (p-value) is arbitrary and does not take into account the costs and consequences of a wrong decision:³ it seems unlikely that one is willing to accept the same probability of being wrong about the effectiveness of an indigestion remedy as for an anti-cancer drug.

Statistical decision theory provides a more logically consistent approach to decision making.⁴ If an analysis represents the best estimate of the costs and consequences of different

courses of action, then the decision should be based solely on those expected costs and consequences, irrespective of uncertainty. I discuss this further in Section 1.5.2.

1.4. Value of Information Analysis

In this section I introduce the concepts and purpose of value of information analysis and methods for calculating the relevant statistics.

Information theory has its origins in the early 1960s in the work of Raiffa & Schlaifer,⁸⁰ but recently interest has grown in its application to healthcare decision making to inform future research. Value of information (VoI) analysis values the returns from investment in further research to reduce decision uncertainty and thus provides a justification for whether research should be conducted, and if so, on which uncertain parameters, and the appropriate sample size for such a study.⁸¹⁻⁸³ It can therefore be used in place of conventional power calculations to estimate the appropriate sample size for future trials of the technology under consideration, based on a comparison of the return from the marginal trial enrollee and the associated marginal cost of including her/him in the research.

Pilot studies have been undertaken to inform future research priorities in the NHS Health Technology Assessment programme⁸⁴ and for the National Institute for Health and Care Excellence (NICE),⁸⁵ and VoI analyses are increasingly appearing alongside published economic evaluations.⁸⁻¹⁹

The value of information statistics are the expected value of perfect information (EVPI), the expected value of sample information (EVSI), and the expected net benefit of sampling (ENBS). The sample size which maximises the ENBS is the optimal sample size for a new study which I denote n^* .

The expected value of 'perfect parameter' or 'partial perfect' information (EVPPI) is defined as the value of perfect knowledge about a particular input parameter, rather than the entire decision problem itself. In other words it is the expected value of perfect information for that parameter, rather than for incremental net benefit.

There are two approaches to estimating the value of additional information; an analytic solution requiring knowledge solely of the parameters of the distribution of incremental net benefit (and assumptions regarding the dispersion [standard deviation] of those parameters in each arm) and a numeric solution based on Monte Carlo simulation of a decision analytic model.

The former is the approach favoured by economists such as Willan^{82 86} whilst the latter is associated with the work of Claxton.^{3 81 83-85}

1.4.1. Analytic solution

The analytic solution illustrated below assumes incremental net benefit (denoted b where the subscript '0' indicates the prior) is a simple linear combination of incremental cost and outcomes (Equations [1-4] and [1-5]). I assume outcomes are measured in QALYs throughout and a threshold of £30,000 is assumed unless otherwise stated. I express the covariance between incremental cost and QALYs as the product of the standard errors and correlation coefficient.

$$b_0 = \lambda \Delta E_0 - \Delta C_0 \quad [1-4]$$

$$\begin{aligned} v_0 &= \lambda^2 v(\Delta E)_0 + v(\Delta C)_0 - \lambda 2 \text{Cov}(\Delta E, \Delta C)_0 \quad [1-5] \\ &= \lambda^2 v(\Delta E)_0 + v(\Delta C)_0 \\ &\quad - \lambda 2 \sqrt{v(\Delta E)_0} \sqrt{v(\Delta C)_0} \rho(\Delta E, \Delta C)_0 \end{aligned}$$

Where:

b_0 = (prior) mean incremental net benefit

ΔE = incremental outcome (e.g. QALYs)

ΔC = incremental cost

v_0 = (prior) variance of mean incremental net benefit

$v(x)_0$ = (prior) variance of mean of x

$\rho(x,y)_0$ = (prior) correlation coefficient between mean of x and y .

Expected Value of Perfect Information

Conceptually, the EVPI is the probability of making the 'wrong' decision multiplied by the associated loss (Equation [1-6]).^{xviii} Graphically, this probability is the area under the probability density function, denoted $f(b)$, on the opposite side of the y-axis to the mean (Figure 1-1, red shaded area). The loss associated with every value of b is plotted as $L(b)$.

$$EVPI_0 = N \left[I\{b_0 \geq 0\} \int_{-\infty}^0 -bf_0(b)db + I\{b_0 < 0\} \int_0^{\infty} bf_0(b)db \right] \quad [1-6]$$

$$N = \sum_{t=0}^T \frac{I_t}{(1+r)^t} \quad [1-7]$$

Where:

^{xviii} Note Willan & Briggs⁸⁶ set the conditions within the indicator function such that when b is exactly zero, New would be rejected. I have reversed this in Equation [1-6], representing a 'benefit of the doubt'. At this point the decision maker is indifferent between Old and New. However, this makes no difference to the solution as the probability of observing a particular point value (e.g. exactly zero) with a continuously distributed variable is zero.

$EVPI_0 =$ (prior) expected value of perfect information

$b_0 =$ (prior) mean incremental net benefit

$N =$ total population (present and future) who can benefit from one or the other treatment.

$I_t =$ Incident population at time t .

$r =$ discount rate.

$I\{\}$ is the indicator function which returns 1 if the condition $\{\}$ is satisfied, otherwise 0.

$f_0(b) =$ prior density function of b .

If mean incremental net benefit (b) is positive then the indicator function in Equation [1-6]

means the second term in the equation is zero, and the EVPI is $\int_{-\infty}^0 -bf_0(b)db$.

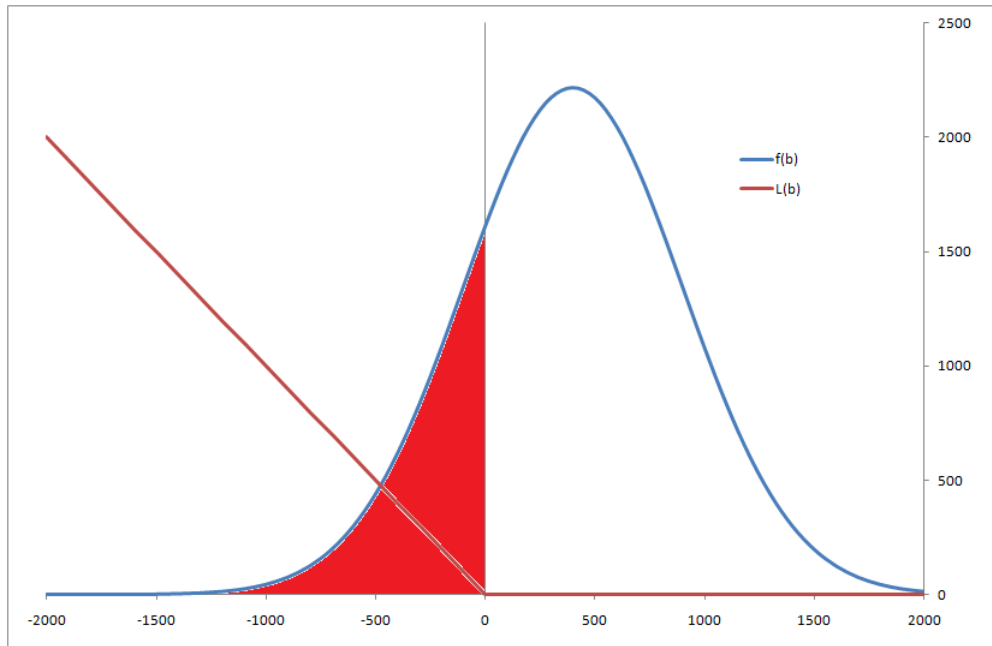
The integral is from $-\infty$ to zero because if the 'true' value of b is greater than zero then the correct decision has been made and there is thus no opportunity loss. However, if the 'true' value of b is actually negative, then the wrong decision has been made, and the loss is $-b$.

This is best explained by example:

Suppose current evidence yields an estimate of mean incremental net benefit (b_0) greater than zero. The decision on this evidence would be to adopt New. But due to uncertainty as to the true value, it could actually be $-\text{£}10$. If this is the case the wrong decision will have been made. The payoff would have been $\text{£}10$ higher had New been rejected, so the loss is $\text{£}10$, which is $-b$. The probability that $b=-10$ is $f_0(-10)$. Integrating across all values of b from $-\infty$ to zero yields the expected loss if b is negative. Similarly $\int_0^{\infty} bf_0(b)db$ is the expected opportunity loss per patient from retaining Old and rejecting New.

The per-patient EVPI is multiplied by N , the total present and (discounted) future population who could benefit from the information (Equation [1-7]). Depending on the disease, this may comprise the current prevalence, plus the net incidence over an 'appropriate' time horizon, discounted at an 'appropriate' rate. (The definition of an 'appropriate' time horizon is considered in Chapter 3).

Figure 1-1: Graphical Illustration of EVPI



If incremental net benefit is assumed normally distributed, the EVPI can be estimated using the unit normal linear loss integral (UNLLI denoted L_{N^*} ; Equation [1-8]).^{4,86} The unit normal loss is calculated as the absolute normalised prior incremental net benefit (the absolute prior mean divided by its standard error), and then rescaled by multiplying back by the standard error.

$$EVPI_0 = N\sqrt{v_0}L_{N^*}\left(\frac{|b_0|}{\sqrt{v_0}}\right) \quad [1-8]$$

$$= N\sqrt{v_0}\left(\phi\left(\frac{|b_0|}{\sqrt{v_0}}\right) - \frac{|b_0|}{\sqrt{v_0}}\left[\Phi\left(-\frac{|b_0|}{\sqrt{v_0}}\right) - I\{b_0 < 0\}\right]\right)$$

Where:

N = the present value of the current and future beneficial population, as defined in Equation [1-7]

v_0 = (prior) variance of mean incremental net benefit, b_0

$\phi(x)$ is the standard normal probability density function evaluated at x

$\Phi(x)$ is the standard normal cumulative density function and is thus the probability of observing a value less than or equal to x .

$\phi(x)$ and $\Phi(x)$ can be obtained from published tables, or most simply using the 'normdist' command in excel.^{xix} I present their formulae in Equations [1-9] and [1-10].

^{xix} The relevant excel command for the pdf at x is "=normdist(x,0,1,0)", and for the cdf at x "=normdist(x,0,1,1)"

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2}\right)} \quad [1-9]$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\left(\frac{x^2}{2}\right)} \quad [1-10]$$

Example

Suppose a trial based economic evaluation comparing 'Old' with 'New' yielded the following:

Mean incremental net benefit (INB) b_0 = £1,000

Standard Error of Mean INB $\sqrt{v_0}$ = £1,000

Further suppose that there are 10,000 patients who could benefit from the treatment. This represents both present and future patient population over an 'appropriate' time horizon.

Therefore the EVPI is:

$$\begin{aligned} EVPI_0 &= 10000 * 1000 (\phi(1) - 1 * [\Phi(-1) - 0]) \\ &= 10000 * 1000 * (0.2420 - 0.1587) \\ &= £833,000 \end{aligned}$$

Expected Value of Perfect Parameter information

The EVPPI is the expected loss resulting from uncertainty in a particular parameter or group of parameters that form a part of incremental net benefit (in the case of this example, simply incremental health gain and incremental cost, but in principle can be extended to any component of INB). It can be estimated by assessing the impact of reducing the standard error of the target parameter to zero on the reduction in standard error of overall incremental net benefit. In other words, the EVPPI is the (expected) reduction in expected loss from the reduction in decision uncertainty attributable to eliminating uncertainty in the target parameter.

The variance of mean incremental net benefit, v is defined as the sum of the variances of ΔE and ΔC less twice the covariance (note λ^2 converts the $v(\Delta E)$ into monetary units; Equation [1-5], repeated here for convenience).

$$v_0 = \lambda^2 v(\Delta E)_0 + v(\Delta C)_0 - \lambda 2 \sqrt{v(\Delta E)_0} \sqrt{v(\Delta C)_0} \rho(\Delta E, \Delta C)_0 \quad [1-5]$$

If ΔC were to be known with certainty, then the predicted posterior variance of ΔC , $v(\Delta C)_1$ would equal 0. Noting that $v(\Delta E)_1 = v(\Delta E)_0$ and $\rho(\Delta E, \Delta C)_1 = \rho(\Delta E, \Delta C)_0$, the predicted posterior variance of b , v_1 reduces to the prior estimate of the variance of ΔE (expressed in

monetary terms by multiplying by λ^2 , Equation [1-11]). The (expected) reduction in variance of b is therefore as per Equation [1-12]. Thus the expected value of perfect partial information is as per Equation [1-13]. The equivalent is true for the value of eliminating uncertainty in ΔE , where the reduction in uncertainty is as per Equation [1-14].

Thus the equations for EVPI and EVPPI can be thought of as the UNLLI evaluated at the normalised mean, where the prior mean incremental net benefit is divided by the expected reduction in uncertainty, multiplied by the expected reduction in decision uncertainty (then multiplied by the beneficial population). In the case of EVPI, it is all decision uncertainty, i.e. the standard error of incremental net benefit. In the case of EVPPI it is the parameter uncertainty, i.e. standard error of the relevant parameter adjusted for the covariance.^{xx} EVPI is therefore a special case of the EVPPI.

$$\begin{aligned}
 v_1 &= \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - \lambda 2\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}\rho(\Delta E, \Delta C)_1 \\
 &= \lambda^2 v(\Delta E)_0 + 0 - 0 \\
 &= \lambda^2 v(\Delta E)_0
 \end{aligned}
 \tag{1-11}$$

$$\begin{aligned}
 v_0 - v_1 &= \lambda^2 v(\Delta E)_0 + v(\Delta C)_0 - \lambda 2\sqrt{v(\Delta E)_0}\sqrt{v(\Delta C)_0}\rho(\Delta E, \Delta C)_0 \\
 &\quad - \lambda^2 v(\Delta E)_0 \\
 &= v(\Delta C)_0 - \lambda 2\sqrt{v(\Delta E)_0}\sqrt{v(\Delta C)_0}\rho(\Delta E, \Delta C)_0
 \end{aligned}
 \tag{1-12}$$

$$EVPPPI_{\Delta C} = N\sqrt{v_0 - v_1}L_{N^*}\left(\frac{|b_0|}{\sqrt{v_0 - v_1}}\right)
 \tag{1-13}$$

$$\begin{aligned}
 v_0 - v_1 &= \lambda^2 v(\Delta E)_0 + v(\Delta C)_0 - \lambda 2\sqrt{v(\Delta E)_0}\sqrt{v(\Delta C)_0}\rho(\Delta E, \Delta C)_0 \\
 &\quad - v(\Delta C)_0 \\
 &= \lambda^2 v(\Delta E)_0 - \lambda 2\sqrt{v(\Delta E)_0}\sqrt{v(\Delta C)_0}\rho(\Delta E, \Delta C)_0
 \end{aligned}
 \tag{1-14}$$

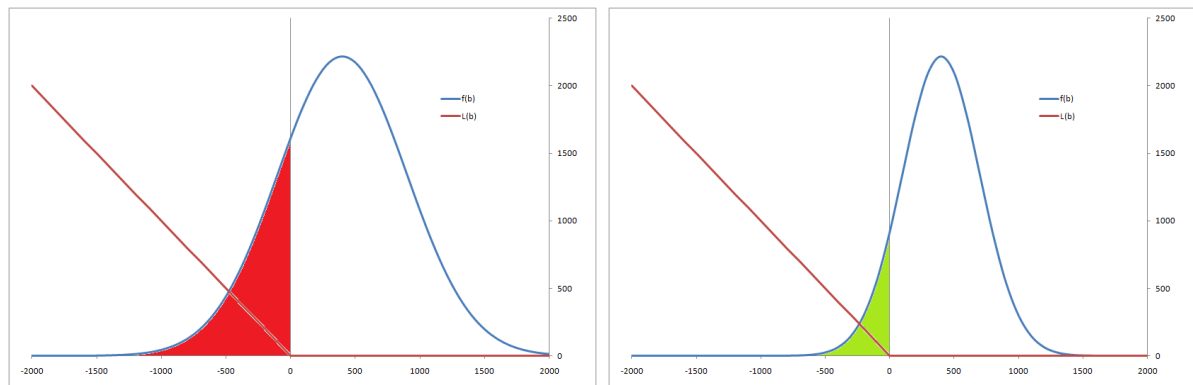
Expected Value of Sample Information

The expected value of sample information is the difference between the expected posterior EVPI (that is, with the additional sample information and denoted $EVPI_1$), and the prior EVPI

^{xx} Note this is only true under assumptions of normally distributed parameters and linear relationships. This is discussed in section 2.2.1

(denoted $EVPI_0$).⁸⁶ In other words, it is the expected reduction in the expected value of perfect information from a trial of a given size. As the posterior EVPI is predicted before data collection has actually taken place EVSI analysis can be termed 'preposterior analysis'.⁴ Graphically it is the EVPI represented in Figure 1-2a ($EVPI_0$) less the EVPI represented in Figure 1-2b ($EVPI_1$). Panel a represents current decision uncertainty (i.e. prior distribution of incremental net benefit). Panel b represents the 'preposterior' distribution, that is, current information combined with the expected results of a new trial of sample size n. The difference between the two is the expected reduction in EVPI from the trial, or expected value of sample information (EVSI). Note that both the prior and preposterior distributions share the same mean as the expected sample mean, given the prior mean, is the prior mean itself.

Figure 1-2: Prior (a) and (expected) posterior (b) distribution of incremental net benefit.



$EVPI_1$ is uncertain as it is conditional on the trial information, which is unknown. Therefore the expected $EVPI_1$ is the $EVPI_1$ associated with a particular sample result, \hat{b} , multiplied by the probability of observing that result (Equation [1-15]). The predictive distribution of \hat{b} , denoted $\hat{f}(\hat{b})$, is simply the prior distribution of incremental net benefit. The EVSI is thus the difference between prior EVPI and expected posterior EVPI, which is then multiplied by the patient population, N, as previously but less those enrolled in the study as they cannot benefit from the information (Equation [1-16]).

$$E(EVPI_1) = \int_{-\infty}^{\infty} EVPI_1 \hat{f}(\hat{b}) d\hat{b} \quad [1-15]$$

$$E_{\hat{b}} EVSI(n, \hat{b}) = (N - 2n) \left[EVPI_0 - \int_{-\infty}^{\infty} EVPI_1 \hat{f}(\hat{b}) d\hat{b} \right] \quad [1-16]$$

Willan and Pinto⁸² provide a comprehensive but rather complex approach to calculating the EVSI. Alternatively, a simpler notation can be derived based on Equation [1-13] above where the reduction in standard error of incremental net benefit from a trial of sample size n is denoted $\sqrt{v_s}$ and the potentially beneficial population is the total population less those enrolled in the study [1-17].⁴ Thus v_s is the difference between prior and (expected) posterior variance of mean incremental net benefit and is calculated as per Equation [1-18].

$$EVSI = (N - 2n) \sqrt{v_s} L_{N^*} \left(\frac{|b_0|}{\sqrt{v_s}} \right) \quad [1-17]$$

$$v_s = v_0 \left(\frac{n}{n' + n} \right) = v_0 \left(\frac{n}{\frac{\sigma^2}{v_0} + n} \right) = v_0 - \left(\frac{1}{v_0} + \frac{n}{\sigma^2} \right)^{-1} \quad [1-18]$$

n' is the notional prior sample size which may be known where there are actual prior data or inferred by dividing the sample variance of incremental net benefit (σ^2) by the variance of the mean (v_0).

Example

Continuing the example above, suppose $\sigma^2 = \text{£}100,000,000$. This would be estimated from prior evidence as the sum of the sample variance of net benefit in each arm. Such prior data may be obtained from existing studies into the same decision question or a pilot study. Alternatively where there are no data a literature search of studies in a similar field may yield plausible estimates. Failing this, expert opinion may be sought.^{xxi} Where patient level data on cost and outcomes in a two arm trial comparing treatment (T) and control (C) are available (such as from a pilot study), this is calculated as per Equation [1-19].

^{xxi} As is routinely undertaken for conventional sample size / power calculations

$$\begin{aligned}
\sigma^2 &= [\lambda^2 s_{e,T}^2 + s_{c,T}^2 - 2\lambda Cov(e_T, c_T)] + [\lambda^2 s_{e,C}^2 + s_{c,C}^2 - 2\lambda Cov(e_C, c_C)] \\
&= \lambda^2 \frac{\sum_i (e_{i,T} - \bar{e}_T)^2}{(n_T - 1)} + \frac{\sum_i (c_{i,T} - \bar{c}_T)^2}{(n_T - 1)} \\
&\quad - 2\lambda \frac{\sum_i (e_{i,T} - \bar{e}_T)(c_{i,T} - \bar{c}_T)}{(n_T - 1)} + \lambda^2 \frac{\sum_i (e_{i,C} - \bar{e}_C)^2}{(n_C - 1)} \\
&\quad + \frac{\sum_i (c_{i,C} - \bar{c}_C)^2}{(n_C - 1)} - 2\lambda \frac{\sum_i (e_{i,C} - \bar{e}_C)(c_{i,T} - \bar{c}_C)}{(n_C - 1)}
\end{aligned} \tag{1-19}$$

Where $e_{i,j}$ and $c_{i,j}$ are the QALYs and cost of patient i with treatment j ($j = T, C$ for Treatment and Control), n_j is the sample size in arm j , and λ is the value placed on a unit of outcome. For example, the outcome may be QALYs and the value per QALY, £30,000.

Suppose a study of sample size $n=100$ per arm is proposed. First calculate the (expected) reduction in variance of mean incremental net benefit (Equation [1-18]):

$$v_s = 1,000,000 - \left(\frac{1}{1,000,000} + \frac{100}{100,000,000} \right)^{-1} = £500,000$$

The EVSI is then the unit normal loss multiplied by the reduction in standard error and by the beneficial population as previously (Equation [1-17]):

$$\begin{aligned}
EVSI &= (10,000 - 2 * 100) \sqrt{500,000} L_N^* \left(\frac{|1,000|}{\sqrt{500,000}} \right) \\
&= (10,000 - 2 * 100) \sqrt{500,000} \left(\phi \left(\frac{1,000}{\sqrt{500,000}} \right) - \frac{1,000}{\sqrt{500,000}} \left[\Phi \left(-\frac{1,000}{\sqrt{500,000}} \right) - 0 \right] \right) \\
&= 8000 * \sqrt{500,000} * (0.1468 - 1.414 * 0.079) \\
&= 8000 * \sqrt{500,000} * 0.035094 \\
&= £246,247
\end{aligned}$$

Expected Net Benefit of Sampling

The expected net benefit of sampling is the expected gain from the trial (i.e. EVSI) less the cost of sampling (Equations [1-20] & [1-21]). The cost of sampling is usually simplified into a fixed and variable component. Added to this is the opportunity cost for the patients randomised to the arm currently believed to be inferior.⁸⁶

$$TC(n) = C_f + 2nC_v + n|b_0| \tag{1-20}$$

$$ENBS(n) = EVSI(n) - TC(n) \tag{1-21}$$

Note that both the EVSI and TC are functions of n . The optimal n (denoted n^*) is that which maximises the ENBS.

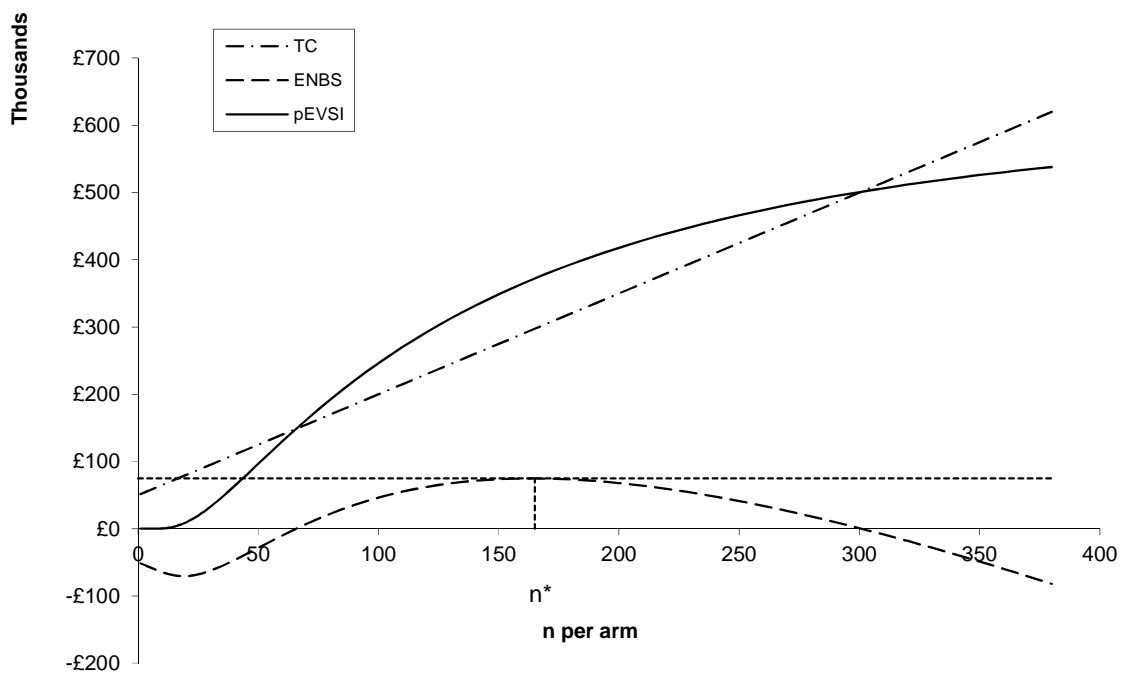
Example

Suppose the fixed costs of a trial totalled £50,000 and a variable cost of £250 per patient enrolled. A trial of size $n=100$ per arm would therefore cost (Equation [1-20]):

$$TC(100) = 50,000 + 200 * 250 + 100 * 1000 = £200,000$$

The ENBS of a trial of 100 patients in each arm is thus £246,247 - £200,000 = £46,247. As this is greater than zero, this trial would be worthwhile, however the calculations should be repeated for a range of values of n to identify the ENBS maximising n (denoted n^*). This occurs at a sample size of approximately 165 patients per arm (Figure 1.3).

Figure 1.3: Population EVSI (pEVSI), total cost of sampling (TC) and ENBS.



1.4.2. Numeric solution

The numeric solution is based around a decision model, e.g. decision tree or Markov chain, which combines together a vector of input parameters, θ to estimate the net benefit from each treatment. The decision rule is to choose option j which maximises the expected net benefit, based on the input parameter set, θ (Equation [1-3], expressed as a function of θ in Equation [1-22]).

$$D = \operatorname{argmax}_j E_{\theta} NB(j, \theta) \quad [1-22]$$

Expected Value of Perfect Information

The expected value of perfect information is the difference between the maximum net benefit with perfect information and that with current information (Equation [1-23]). The maximum expected net benefit with current information is simply the outcome of the model (Equation [1-22]). The maximum net benefit with perfect information is unknown, therefore the expectation is taken over θ .

Equation [1-23] is the EVPI per patient. As in Equation [1-6], this should be scaled up to the current and future population to provide an upper limit for the budget for future research in to the technology in question (Equation [1-24]), where N is defined as per Equation [1-7].

$$EVPI = E_{\theta} \max_j NB(j, \theta) - \max_j E_{\theta} NB(j, \theta) \quad [1-23]$$

$$PopnEVPI = EVPI \cdot N \quad [1-24]$$

Where:

j = intervention (e.g. 1= current treatment, 2=new treatment)

θ = input parameters to model

$NB(j, \theta)$ = net benefit of intervention j with parameter set θ .

Example

Suppose the current treatment for disease X is called 'Old', and patients currently treated have an annual risk of death of P_0 . A new treatment 'New' is developed which reduces the relative risk of death by RR_N , but is also more expensive. The decision question is whether to adopt 'New' in place of 'Old'. A systematic review and meta-analysis have yielded point estimates and distributions around P_0 and RR_N : assume current data comprises just one study of 200 patients randomised to Old and New, and that this estimated a 20% annual probability of death with 'Old', and a relative risk of death with 'New' of 0.75. Based on this information, the baseline probability of death (with 'Old') is characterised as a beta

distribution and relative risk as lognormal with parameters as per Equations [1-25] and [1-26].

$$P_0 \sim \text{Beta}(20,80) \quad [1-25]$$

$$\text{LN}(RR_N) \sim \text{Normal} - (0.29,0.097) \quad [1-26]$$

A simple economic model based on a two state Markov chain has been developed incorporating these as well as other data (such as the 'up front' cost of 'Old' and 'New', changes in subsequent resource use and costs associated with 'Old' and 'New' and quality of life estimates), over an 'appropriate' time horizon (Figure 1-4). The model combines these input parameters into estimates of the net benefit from each treatment (NB_O and NB_N) and a Monte Carlo simulation (probabilistic sensitivity analysis) is used to derive an expected net benefit with 'Old', $E(NB_O)$ and expected net benefit with 'New', $E(NB_N)$.

Figure 1-4: Structure of Decision Model

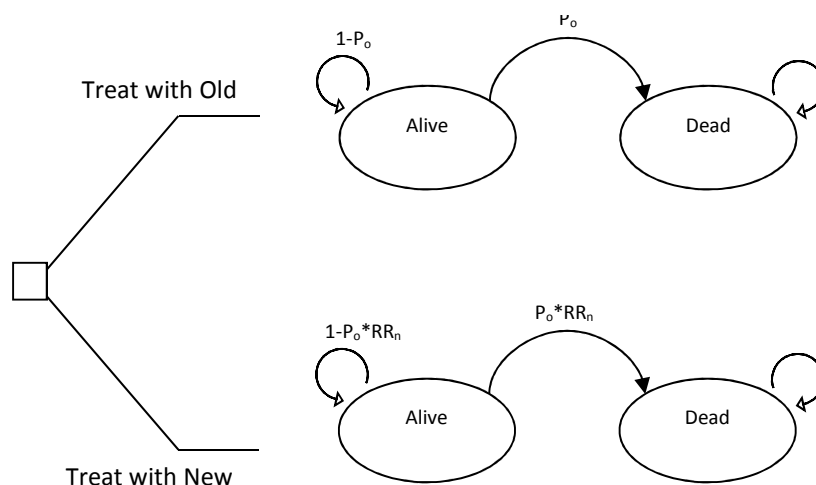


Table 1-1 shows the results of the PSA. (Note only five iterations are shown in this illustrative example: several thousand would typically be necessary, depending on the complexity of the model). The expected net benefit of 'Old' is £84,178 and of 'New', £92,153. The maximum expected net benefit (Equation [1-22]) is thus £92,153 and 'New' should be adopted. However, for iterations 2, 3 and 5, 'Old' had the highest net benefit, and therefore as 'New' is the one that is chosen, there would be an opportunity loss equal to the difference in net benefit (£16,677, £15,026 and £315 in each case respectively). The expected loss over all five iterations is £6,403, which is the expected gain from eliminating all uncertainty, i.e. the expected value of perfect information.

Table 1-1: EVPI Illustration

Iteration	NB _O	NB _N	D	Max	Loss
1	£67,913	£119,013	N	£119,013	£0
2	£110,199	£93,522	O	£110,199	£16,677
3	£77,624	£62,598	O	£77,624	£15,026
4	£68,291	£89,083	N	£89,083	£0
5	£96,863	£96,548	O	£96,863	£315
E(.)	£84,178	£92,153	N	£98,556	£6,403

The sum of £6,403 is the per-patient EVPI. This should be scaled up to the current and future population to provide an upper limit for the budget for future research in to the technology in question (Equation [1-24]). Thus if the incidence of disease X is 10,000 per annum, over a 10 year time horizon and discount rate of 3.5%, the population EVPI is:

$$£6,403 * \sum_{t=0}^9 \frac{10,000}{1.035^t} = £551m$$

The maximum budget for research into the cost-effectiveness of ‘Old’ versus ‘New’ should therefore be set at £551m. If a particular research project will cost more than the EVPI, then it will definitely not be cost-effective and the funds should be spent elsewhere (e.g. direct patient care or in an alternative research area).

Expected Value of Perfect Parameter Information

The EVPI is the overall maximum gain from eliminating all uncertainty within a decision model. As the numeric approach is based on a decision model incorporating many individual input parameters, it is natural to investigate the value of eliminating uncertainty in those individual parameters. The expected value of perfect information for a parameter (EVPPI) provides an upper bound to research expenditure in a particular parameter (Equation [1-27]).

The two expectations in the first term require nested iterations of the Monte Carlo simulation. This proceeds as follows:

1. Draw a value from the distribution of φ . This is a possible realisation of the 'true' value of φ .
2. Run the Monte Carlo simulation k times, drawing values from ψ (i.e. all other parameters in the model) for each simulation, whilst holding φ at the value drawn in step 1.

3. Record the expected net benefit from each treatment as the mean over the k simulations.
4. Repeat steps 1-3 n times.
5. Calculate the EVPPI as per Table 1-1 above, i.e. the difference between the expected maximum expected net benefit and the maximum expected net benefit.

Again, this should be multiplied by the incident population over an ‘appropriate’ time horizon to calculate the population EVPPI (Equation [1-28]). Note the sum of EVPPIs across all parameters will not generally equal the EVPI due to interactions between parameters.

$$EVPPI_{\varphi} = E_{\varphi} \max_j E_{\psi|\varphi} NB(j, \varphi, \psi) - \max_j E_{\theta} NB(j, \theta) \quad [1-27]$$

$$PopnEVPPI = EVPPI . N \quad [1-28]$$

Where:

$$\varphi \cup \psi = \theta$$

(φ is a parameter or subset of parameters of interest, ψ is all the others in set θ).

Expected Value of Sample Information

The expected value of sample information (EVSI) provides the sufficient condition as to whether undertake a particular trial by estimating the return from the trial given a sample size n. As per the explanations of the equations for EVPI and EVPPI, this can be expressed as the difference between the expected maximum expected net benefit with the new information and the maximum expected net benefit with current information (Equation [1-29]). Equation [1-30] shows the equation for groups of parameters within a decision problem. As before, the per patient EVSI then needs multiplying by the beneficial population, defined as N less those enrolled in the study.

$$E_D \max_j E_{\theta|D} NB(j, \theta) - \max_j E_{\theta} NB(j, \theta) \quad [1-29]$$

$$E_D \max_j E_{\psi, \varphi|D} NB(j, \varphi, \psi) - \max_j E_{\theta} NB(j, \theta) \quad [1-30]$$

Where:

θ = uncertain parameter

D = sampled value of θ from trial of size n

Example

Figures 1-5 and 1-6 show the steps in calculating the EVSI and ENBS for the parameter P_o . The Figures each show three tables labelled a-c. Table 'b' contains the output of the standard economic model PSA. Table 'a' contains the summaries of each of these, and Table 'c' will contain the EVSI estimated for each sample size.

With current information, the meta-analysis determined the prior distribution of P_o as $P_o \sim \text{Beta}(20,80)$. The EVPPI associated with this parameter is greater than zero (£339m) suggesting there is potential for additional information to be useful. Calculating the EVSI and ENBS of trials of different sizes will provide the sufficient condition for proceeding: the sample size that maximises ENBS (subject to $\text{ENBS} > 0$) is the efficient sample size.

First set a sample size for the proposed study, e.g. $n=10$. A value is then sampled from the prior distribution of P_o . This represents one possible realisation of the world. Suppose the value drawn is 0.24. So this is a world where the true population baseline mortality rate is 24%. This is filled in in cell (2,3) of Figure 1-5, table a. The results of the study (call this P_{os}) therefore must be a binomial random variable with mean 0.24 and sample size 10. I.e. $P_{os} \sim \text{Bin}(0.24,10)$.

A value is drawn from this distribution as a possible realisation of the study results. Suppose the result was 3. That is, one possible result is that 3 patients died and 7 survived to one year. The next step is to use Bayes' theorem to combine the prior distribution and the 'new' data to a (pre)posterior distribution.⁷ Call this P_o' . For the beta distribution this is simply:

$$P_o' \sim \text{Beta}(A+P_{os}, B+n-P_{os})$$

$$\Rightarrow P_o' \sim \text{Beta}(20+3, 80+10-3)$$

$$\Rightarrow P_o' \sim \text{Beta}(23,87)$$

This equation is entered in cell (2,1) of Figure 1-5, table b. The Monte Carlo simulation is then run on the model for a large number of iterations, each time sampling from the distributions of P_o' as well as the other model inputs, and the net benefit obtained from each treatment recorded in Figure 1-5, table b, cells (3,3) to (4,7).

After all the iterations, the expected net benefit from 'Old' is estimated at £98,453, and from 'New', £85,395 (Figure 1-5, table b, cells (3,2) and (4,2)). These figures are transferred to cells (3,3) and (4,3) of table a.

The next step is to sample from the prior distribution of P_0 again. Suppose the value this time is 0.18. This is recorded in Table a (Figure 1-6, Table a, cell (2,4)). The results of the hypothesised study are now a binomial random variable with a Binomial(0.18,10) distribution. Sampling from this distribution, suppose a possible study result is 1, that is 1 death and 9 survivors. Adding this new data to the prior yields $P_0 \sim \text{Beta}(21,89)$, which is the new (pre)posterior distribution. The model is then run a large number of times sampling from this new distribution, along with the remaining model inputs, and the expected net benefit from each treatment recorded. In this case the results are £72,814 and £84,599 respectively (Figure 1-6, Table a cells (3,4) and (4,4)). This process is repeated a 'large' number of times.

Figure 1-5: EVSI example

a. Iteration summaries

Run				
Average	Po ~ Beta(20,80)	NBo	NBn	Max
1	0.24	£98,453	£85,395	
2				
3				
4				
5				

c. EVSI

$$P_{os} \sim \text{Bin}(0.24, 10)$$



$$P_{os} = 3$$

n	EVSI
10	
20	
30	
40	
50	
...	

b. Model iterations

MODEL	Po' ~ Beta(23,87)	NBo	NBn
Mean		£98,453	£85,395
1	0.23	£118,651	£95,671
2	0.22	£118,873	£78,423
3	0.24	£73,682	£92,143
4	0.23	£104,719	£73,621
5	0.23	£76,339	£87,116

Figure 1-6: EVSI Example (continued)

a. Iteration summaries

Run		£81,324	£88,396	£91,008
Average	Po ~ Beta(20,80)	NBo	NBn	Max
1	0.24	£98,453	£85,395	£98,453
2	0.18	£72,184	£84,599	£84,599
3	0.19	£76,439	£94,157	£94,157
4	0.19	£75,368	£85,677	£85,677
5	0.21	£84,178	£92,153	£92,153

c. EVSI

n	EVSI
10	£2612
20	
30	
40	
50	
...	

$$P_{os} \sim \text{Bin}(0.21, 10)$$



$$P_{os} = 2$$

b. Model iterations

MODEL	Po' ~ Beta(22,88)	NBo	NBn
Mean		£84,178	£92,153
1	0.23	£100,584	£120,634
2	0.23	£67,527	£73,963
3	0.23	£108,340	£99,127
4	0.22	£57,498	£61,501
5	0.23	£86,941	£105,540

After running the iterations, the maximum expected net benefit from each run is recorded in the final column of Figure 1-5, Table a, and the mean of each column taken (recorded in the top row of Table a: £81,324, £88,396 and £91,008). The EVSI is the expected maximum net benefit with improved information about P_o less the maximum expected net benefit with current information. The former is simply the expectation of the final column of Table a

(£91,008). The latter is approximated by the maximum of the expectations of columns 3 and 4 (=max(£81,324, £88,396)). Thus the expected value of sample information on parameter P_0 from a study of size $n=10$ is:

$$£91,008 - \max(£81,324, £88,396) = £2,612.$$

The exercise is now repeated for a range of sample sizes. Table 1-2 shows the results for calculating EVSI for sample sizes of between 0-500 patients. As with EVPI and EVPPI, this is the per-patient EVSI, so this needs multiplying by the present and (discounted) future population of patients (column 3 of Table 1-2).

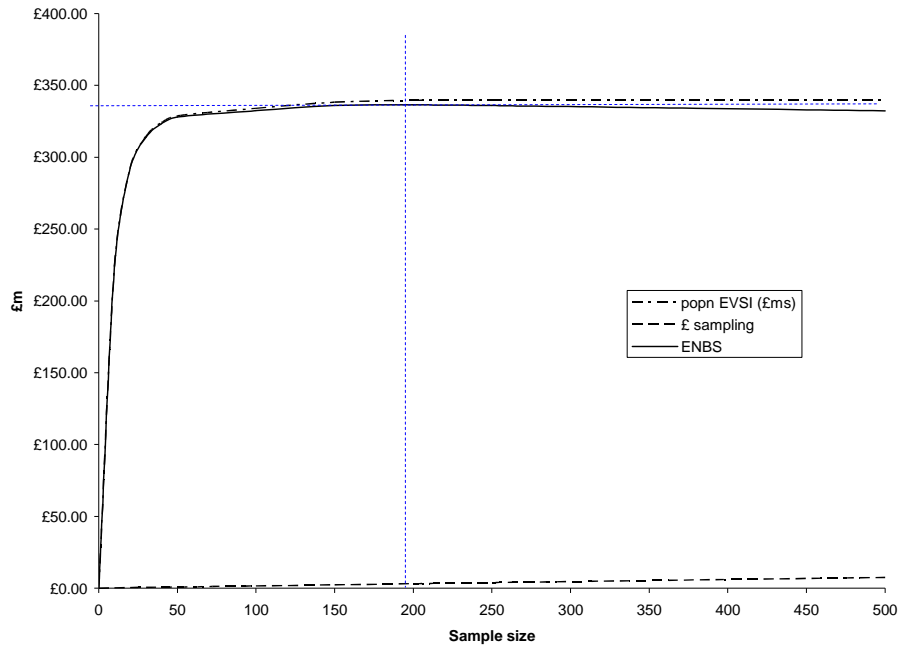
Expected Net Benefit of Sampling

This is estimated in exactly the same way as for the analytic solution. The costs of conducting the study are split into a fixed and variable component (Table 1-2, columns 4-6). The difference between EVSI and TC is the expected net benefit of sampling (Table 1-2, final column). The sample size at which this is maximised is the optimum sample size for the study, in this case one of approximately 195 patients (Figure 1-7).

Table 1-2: EVSI & ENBS illustration

n	Per patient EVSI	Population EVSI (£ms)	Cost of sampling (fixed costs, £ms)	Cost of sampling (variable costs, £ms)	£ sampling (£ms)	ENBS (£ms)
0	0	£0.00	£0.00	£0.00	£0.00	£0.00
10	£2,612	£224.83	£0.15	£0.15	£0.30	£224.53
20	£3,400	£292.66	£0.15	£0.30	£0.45	£292.21
30	£3,654	£314.52	£0.15	£0.45	£0.60	£313.92
40	£3,765	£324.08	£0.15	£0.60	£0.75	£323.33
50	£3,820	£328.81	£0.15	£0.75	£0.90	£327.91
100	£3,880	£333.98	£0.15	£1.50	£1.65	£332.33
150	£3,930	£338.28	£0.15	£2.25	£2.40	£335.88
200	£3,945	£339.57	£0.15	£3.00	£3.15	£336.42
250	£3,947	£339.75	£0.15	£3.75	£3.90	£335.85
500	£3,949	£339.92	£0.15	£7.50	£7.65	£332.27

Figure 1-7: EVSI, Cost of sampling and ENBS



1.5. Decision models vs Clinical Trials and the Cycle of Evidence Based Medicine

In Section 1.3, I described the general approach to economic evaluation, differentiating between the two implementations, namely piggybacked studies alongside clinical trials and decision models. I then introduced value of information analysis as a means to estimate the efficient sample size for a future study (Section 1.4). In this section I link these three approaches (decision modelling, clinical trials and value of information analysis) together into a comprehensive framework for decision making and research.

Firstly I discuss the relative advantages and disadvantages of trial-based and model-based economic evaluations. Whilst sometimes seen as mutually exclusive alternatives, I conclude that this is not the case, and that they form part of a cyclical, or iterative approach encompassing not only decisions as to whether to adopt a particular technology but research decisions too. Thus two distinct decision questions are posed: firstly, whether to adopt the new technology in question, and secondly, whether further research is required to reduce uncertainty.

1.5.1. Trials vs Models for economic evaluations

There is some debate as to the relevance and appropriateness of clinical trials as a tool for decision making. One of the key objections is that RCTs conducted as Phase III studies are typically designed to measure efficacy, not effectiveness. Phase III studies provide the bulk of evidence of effect in marketing authorisation applications and are intended to control for as many factors as is possible to establish whether the treatment itself is the reason for any difference observed between treatment groups. They therefore establish whether a treatment *can* work under ideal circumstances, known as efficacy. However, the controlled environment, strict inclusion/exclusion criteria and requirement to adhere to a strict protocol may result in a treatment situation and patient population very different from routine practice. The transferability of the results to 'real life' and thus suitability to inform decision making may therefore be brought into question.⁸⁷ An alternative trial design is a 'pragmatic' trial, designed with minimal exclusion criteria and treatment pathways that match routine care as closely as possible. These studies thus attempt to measure effectiveness, or whether the treatment *does* work under realistic conditions. They thus enhance the external validity (generalisability) of the results, but at the expense of internal validity (bias) due to the relatively uncontrolled^{xxii} nature of the study.

^{xxii} Uncontrolled in the lay sense rather than in terms of comparators.

The key advantage of RCT-based economic evaluations is that they generate evidence using well established best-practice principles: randomised controlled trials are generally considered the least-biased study design and so most likely to give the ‘correct’ answer (notwithstanding my comments about pragmatic trials above). However, one of the main drawbacks is that they do not necessarily ask the ‘right’ question. The information requirements from an economic evaluation for decision making include a statement and appropriate measurement of the objective function, an appropriate time horizon, and consideration of all relevant evidence.^{88,xxiii}

Cost utility analysis with outcomes measured in Quality Adjusted Life Years (QALYs) is considered the preferable form of economic evaluation by some agencies.^{89 90} This is because QALYs are a generic measure of overall quantity and quality of life, thus allowing comparisons of diverse interventions across broad disease areas in the same metric. The use of QALYs also presupposes the health system is interested in maximising the generation of ‘health’ with the resources at its disposal, which is not an unreasonable assumption.^{xxiv} Therefore the desirable objective function would be maximisation of net benefit as described in Section 1.3.8. However, the time frame of a clinical trial may not allow observance of the ultimate benefit of a treatment on a person’s quality and/or quantity of life. Therefore clinical trials frequently choose an interim or surrogate measure as the primary outcome for a trial.

For example, the hoped-for benefits of an anti-osteoporosis drug are ultimately an increase in quantity and/or quality of life (QALYs), mediated through a reduced incidence of fracture. A clinical trial measuring changes in QALYs would require a life-time follow-up period, possibly 30-40 years or more. This would be a highly impractical and expensive trial. Furthermore there is a desire to know whether the drug is likely to be (cost-)effective immediately (or as soon as is practical). Even a follow-up period sufficient to observe changing incidence of fracture may be unfeasible, and a surrogate marker, such as bone mineral density, used as the primary outcome. An economic evaluation conducted alongside this trial would not be estimating the desired objective function, nor would it be considering an appropriate time horizon.

^{xxiii} Sculpher et al.⁸⁸ include other requirements, in particular appropriate characterisation of uncertainty. I consider this later when incorporating Vol analysis into the overall framework.

^{xxiv} The distribution of the health gain (equity) will also be of concern to public health systems.

Critically, it is reasonable to propose that decisions should be based on ‘all relevant evidence’.⁹¹ This includes appropriate comparisons between all relevant treatment options, as well as taking into account current (prior) knowledge on the effects and costs of each: the results of a clinical trial do not exist in a vacuum, and setting it within the context of existing work is essential for interpretation. Indeed, it could be argued that it is impossible for a single clinical trial to collect all relevant cost and outcomes data required for decision making.⁸⁸ It is essential to compare ‘all relevant treatment options’ in order to avoid misleading conclusions. For example, treatment A may be cost-effective compared with B, but not with C. Thus C should be the preferred option, but this will not be known if C is excluded from analysis.

Thus there are limitations to the use of RCTs as a sole basis for decision making. There will always be diverse sources of evidence comprising randomised trials, observational studies, routine databases and expert opinion that individually are insufficient, but together may provide a sufficient estimate of the overall costs and consequences of the different options. These could be combined together informally by decision makers in a narrative manner. However, it may be desirable to formally assemble and structure the evidence sources to assist this process. In essence, a structured assembly and synthesis of the evidence is precisely the definition of a decision model, and so models may be seen as an essential component of the decision making process.⁹²

Thus whilst RCT-based analyses may be the best source of unbiased information about incremental outcomes and costs informing some parameters, the principles of evidence based medicine require that all relevant evidence be brought to bear on a particular decision question. This implies not only combining together the results of multiple RCTs, but, due to the limitations of trials (e.g. failure to compare all relevant treatment options, insufficient time horizon), incorporation of other evidence from diverse sources too. A structured combination of these evidence sources is the definition of a decision model. Thus the two implementations of economic evaluation should be seen as complements rather than substitutes, with trials providing estimates of parameters to incorporate into decision models.

1.5.2. Economic evaluation, Value of Information Analysis and the Iterative Approach to Decision Making

As I demonstrated in Section 1.4, value of information analysis uses the results of an economic evaluation to determine the optimal sample size for a new study on a particular parameter or group of parameters, whilst the results of those studies (not only clinical trials

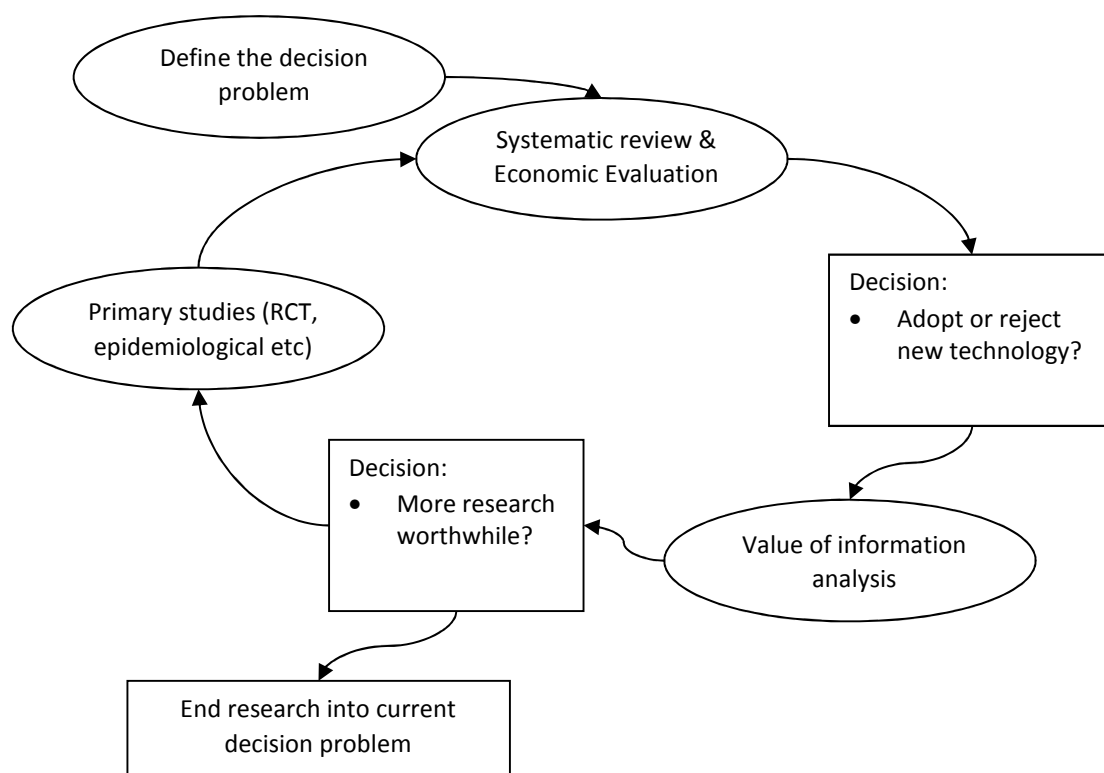
but epidemiological studies, database analyses etc) inform the development of a decision model. Thus we can define an iterative approach to evidence (or economics) based medicine, incorporating not only the adoption decision, but the research decision too (Figure 1-8).¹⁶⁹³⁻⁹⁵

Figure 1-8 begins with a definition of the decision question. This must clearly specify the question as stated in Section 1.3.1, defining (all) the different comparator treatments in a specific population, as well as the perspective of the analysis (e.g. health service or society). Following this, there should be a systematic review of all the relevant input parameters, critically including appropriate characterisation of uncertainty around those parameters. A decision model should then be developed combining these parameters into the objective function (net benefit). Probabilistic sensitivity analysis should be employed to propagate parameter uncertainty through to decision uncertainty. To inform the adoption decision, the most cost effective treatment option is then the option with the highest expected net benefit.^{xxv}

Following the adoption decision, value of information analysis should be conducted to estimate whether further primary research (whether a clinical trial, epidemiological study, database analysis or other study) would be worthwhile, and if so the efficient number of observations on which to collect the data. If the (expected) cost of a particular research project exceeds the expected benefit, then the project should not go ahead, and decisions should be based on the current level of uncertainty. However if the expected benefit exceeds the expected cost, then the project should proceed. The results of this study should then be combined with the prior information, updating the systematic review and economic evaluation, and the adoption decision revisited. Thus the cycle continues.

^{xxv} Decision uncertainty can be represented by a multivariate distribution of n-dimensions, where n is the number of comparators. Where there are only two comparators this simplifies into choosing the new treatment if the incremental net benefit (INB) is greater than zero, and the standard error around mean INB is the measure of the decision uncertainty.

Figure 1-8: The Cycle of ‘Economics’ Based Medicine



Source: Wilson & Abrams 2010¹

The iterative approach raises some interesting methodological points. The first and most important is that the adoption decision is based solely on expected values. This contrasts with the method by which clinical trial results are normally assessed: the purpose of such trials is usually to establish whether a treatment has any real effect *ceteris paribus*, and to rule out chance as the explanation of any observed difference in response rates. Hypothesis testing usually assigns a p-value of 0.05 as the cut-off for statistical significance. The full interpretation of a p-value is that if the experiment were repeated many times in the future, the p-value is the long-run relative frequency with which a difference at least as big as that observed in the original trial would be observed if there really was no difference between the two treatments. If this is sufficiently small, then chance may be excluded and a real treatment effect declared.

The first criticism of this approach is that the critical p-value of 0.05 is arbitrary and takes no account of the costs or consequences of drawing the wrong conclusions: the same value is usually adopted irrespective of the decision question. The second criticism is that basing decisions on statistical inference will not lead to maximisation of expected outcomes subject

to the budget in the long run. If a decision is based on a synthesis of all relevant evidence, then the conclusion of that evidence must be the best possible course of action given current knowledge, irrespective of the uncertainty around that. To take any other course of action is therefore perverse. Such an approach is consistent with the principles of statistical decision theory⁴ in which decision makers should be risk neutral.⁹⁶ Inference should therefore be seen as irrelevant to the adoption decision:³ the desire to establish whether a treatment effect is real or not is a separate activity from, and should not be confused with, making a decision based on the best evidence available.

The second methodological point is that the decision framework is based solidly within a Bayesian rather than frequentist statistical perspective. As stated above, in the traditional frequentist approach, probability is defined as the long run relative frequency with which an event is predicted to occur. The target parameter (e.g. relative treatment effect) is unknown (and unknowable) but certain, and sample information can be obtained as an estimate of the parameter.

In the Bayesian framework, probability is defined as a subjective belief about plausible values of the parameter. The parameter itself is uncertain and thus it is possible to speak of 'parameter uncertainty'. Formal Bayesian analysis always begins with a prior belief (which may be uninformative if starting from a position of ignorance). Prior belief is then combined with the data using Bayes theorem to form the posterior belief. The posterior belief is simply a weighted average of the prior and the data (likelihood).⁷ As the distribution around a parameter represents an individual's belief about likely values, Bayesian hypothesis tests can be used to make direct probability statements (i.e. the probability that the hypothesis is correct). This approach provides an elegant statistical technique capturing the learning process.

Decision models are described in terms of parameter uncertainty and are based within a Bayesian framework.^{xxvi} Value of information analysis is also intrinsically Bayesian, with its reliance on predicted posterior distributions (such that it is sometimes termed 'preposterior analysis').

The third methodological point is that value of information analysis is utterly dependent on the 'appropriate characterisation of uncertainty'. There are well established methods for pooling together odds ratios of treatment effects and appropriately estimating standard

^{xxvi} It is noteworthy that decision models developed as economic evaluations in the health care sector almost never elicit their parameters with a formal Bayesian analysis, stating priors, likelihood and resulting posterior.

errors (e.g.⁹⁷), but methods where there is no prior evidence whatsoever are less well developed. In these situations a decision model is reliant on eliciting beliefs about plausible values for the parameter from a panel of representative experts. Such techniques do exist,⁹⁸ but care must be taken to ensure the standard errors elicited in this way are ‘appropriate’ and fully represent current knowledge.

Finally, value of information analysis is not the only means by which research projects can be prioritised. Other techniques which are commonly used include subjective judgement, burden of disease, the degree of variation in clinical practice and ‘payback’ (a method with some features in common with value of information analysis, where the results of a proposed trial are categorised into positive, negative and uncertain with the costs and benefits from changes in policy under each of the three scenarios are calculated. Assigning probabilities to the likelihood of each scenario then allows computation of the expected net benefit of the trial).⁹⁹ Subjective judgement, burden of disease and the degree of variation in clinical practice do not take into account the opportunity cost of research and therefore are incomplete analyses. Whilst the payback method does take this into account, it divides the results of a trial into the three categories, presumably based on hypothesis testing. This runs counter to the principles of statistical decision theory and so will not necessarily lead to maximisation of expected net benefits. These issues are discussed further in Chapter 5.

1.6. Incorporation of EE & Vol / EBM into health policy in the UK and internationally

1.6.1. Economic evaluation in the UK, Europe and Globally

Within the UK, the principles of economic evaluation are incorporated most strongly within the work of the National Institute for Health and Care Excellence (NICE),¹⁰⁰ with a remit to provide advice on the clinical and cost-effectiveness of certain interventions to the NHS in England and Wales.^{xxvii} Its technology appraisals programme places a formal economic evaluation at the heart of its decision making process, conducted along the lines of that described in Section 1.3. The primary analytic perspective of any economic evaluation conducted by or for NICE is the NHS and personal social services (i.e. public sector).⁷²

Founded in 1999 as the National Institute for Clinical Excellence, a special health authority of the NHS, it faced criticism at first that far from preventing the adoption of cost-ineffective treatments, it was too reluctant to say 'no'.¹⁰¹ The first time that NICE rejected a treatment on the grounds of cost-effectiveness was in its 2002 appraisal of beta interferon and glatiramer acetate, two treatments for multiple sclerosis.¹⁰² The guidance stated that those patients who were currently receiving treatment should continue until such time as their clinician felt they should stop, but no new patients would receive the drug. However, there was an attempt to explore how the treatment could be provided on the NHS in a cost-effective manner. This was through the development of a risk sharing scheme. Despite running from 2002-2005, the results of the prospective cohort study were not published until 2009,^{103 104} and failed to show any beneficial effect of beta interferon. This finding was confirmed by a 2010 Cochrane review,¹⁰⁵ but then contradicted by another review the following year.¹⁰⁶ The explanation for the difference is that the Cochrane review was limited to the two pivotal RCTs for the drugs, whilst Oliver et al.¹⁰⁶ included a number of non-randomised observational studies and therefore may be at higher risk of bias. Despite criticism as a 'costly failure',¹⁰⁷ variants on the risk sharing scheme became the model for a number of other new and expensive therapies.

Most member states of the European Union have a public sector agency charged with providing health technology assessments for their respective health systems.¹⁰⁸ EUnetHTA acts as an association of national HTA agencies to develop methods for HTA across the members with particular regard to the generalisability of studies and adaptations to the

^{xxvii} Similar bodies exist within Scotland (the Scottish Medicines Consortium) and Wales (the All Wales Medicines Strategy Group). Note there is overlap in the remits of NICE and AWMSG. Northern Ireland does not have a specific HTA body, but tends to adopt the decisions of NICE.

different national health systems.¹⁰⁹ Collaborative work so far has resulted in development of a 'core' HTA model^{110 111} and a toolkit for adapting HTA reports to other settings,¹¹²⁻¹¹⁴ a handbook on HTA capacity building, and a database of studies for additional data collection on new technologies.¹⁰⁸

The equivalent global body, the International Network of Agencies for Health Technology Assessment (INAHTA) comprises 53 member organisations from 29 countries (as at 2013). It aims to minimise duplication of effort by disseminating summaries of technology assessments amongst its members and developing checklists of key HTA elements. Members also collaborate on a number of research projects investigating the impact of HTA, variations in findings, quality assurance for rapid reviews and education and training.^{115 116}

Approaches to health technology assessment and economic evaluation vary from country to country. The German health care system is based on a Bismarckian compulsory social insurance scheme. Two bodies are involved in the HTA process. The Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) is the body that produces technology assessments which are then appraised by the Gemeinsame Bundesausschuss (G-BA).

Prior to 2007, IQWiG had a remit to consider only the effectiveness of interventions. Following a change in the law that year the remit was broadened to cost-effectiveness too. The methods by which economic evaluations are to be conducted are broadly comparable to those undertaken by NICE except for some key differences.¹¹⁷ The primary analytic perspective is required to be statutory health insurees. Thus the costs of treatments to the insurer plus any co-payments must be taken into account in any analysis. On the outcomes side, IQWiG rejects the use of QALYs to make comparisons between broad disease areas on the basis of ethical and methodological concerns with utility elicitation techniques (namely the standard gamble and time trade off approaches).^{xxviii} They therefore limit all analyses to within-disease area comparisons. Furthermore an economic evaluation is only undertaken following establishment of an additional benefit on at least one outcome measure. Where an economic evaluation is conducted, IQWiG adopts an efficiency frontier approach where all treatment options are plotted on the cost-effectiveness plane.^{xxix} A new treatment must lie either on or beyond the efficient frontier in order to be approved at the manufacturer's desired price.

^{xxviii} IQWiG does not preclude the use of QALYs to make comparisons within a single disease area however.

^{xxix} By convention, IQWiG plots an effectiveness-cost plane, with costs on the X-axis and benefit on the Y, the reverse to convention in the UK.

The French health care system is, like the German system, based on social insurance. One of two committees of the Haute Autorité de Santé (HAS) advises the ministry of health on whether a particular drug or device should be funded by the public system. If approved, the Comité Économique des Produits de Santé (CEPS) negotiates the price with the manufacturers. Finally, the Union Nationale des Caisses d'Assurance Maladie (UNCAM – the representative body of insurance companies) determines the level of reimbursement offered under their insurance schemes.

HAS was given the remit to consider the cost-effectiveness of health care interventions under the Social Security Financing Act 2008. The required form of analysis is either cost-effectiveness or cost-utility analysis, with a preference for the latter with outcomes measured in QALYs based on results of the EQ5D or HUI3 instruments.⁹⁰ Where a cost-effectiveness analysis is conducted, outcomes are preferred to be reported in life years gained. HAS is less prescriptive than NICE or IQWiG about the analytic perspective, requiring 'a collective perspective that is sufficiently broad to take into account all stakeholders concerned by the treatments studied, in the French health system'.⁹⁰

In Australia, the state healthcare system is known as Medicare, and based around a Beveridge-style system funded out of general taxation along the lines of the NHS. The majority of healthcare costs are covered but patients are required to make co-payments for some services and pharmaceuticals. The Pharmaceutical Benefits Advisory Committee was set up in 1954 to make recommendations to the Minister for Health and Ageing about which drugs should be available to Medicare patients through the Pharmaceutical Benefits Scheme (PBS). Its remit has included assessing the cost-effectiveness of drugs since 1993, and was therefore probably the first state agency to do so.¹¹⁸

As in the UK and France, the preferred form of analysis for PBAC submissions is a cost-utility analysis except where the proposed drug is demonstrated to be therapeutically equivalent to its comparator, in which case a cost-minimisation analysis is acceptable.¹¹⁹ Cost-effectiveness, cost-consequences and cost-benefit analysis are also acceptable under certain conditions. The preferred perspective of the analysis is societal, including costs borne "by patients, government, health insurance agencies and any other part of society".¹¹⁹

In the USA, cost-effectiveness analysis is encompassed within the term ‘comparative effectiveness research’ (CER).^{xxx} There is no federal body in the US with the equivalent appraisal role of NICE, HAS, G-BA or PBAC, although insurance companies may consider the cost-effectiveness of treatments when deciding whether to include cover for them in their policies.

1.6.2. Value of information analysis in health policy

I am not aware of any systematic uses of value of information to formally determine research priorities within the UK health sector. However, NICE (and some other HTA bodies globally) have the option of ‘coverage with evidence development’¹²³ where a treatment is awarded a provisional acceptance subject to further research commissioned specifically to reduce decision uncertainty. Claxton and colleagues¹²⁴ sought to establish what assessments are required to inform an ‘only in research’ (OIR) or ‘approval with research’ (AWR) decision. They suggest an algorithm incorporating issues such as whether current evidence predicts an intervention to be cost-effective, whether there are significant irrecoverable costs associated with adoption, whether more research is worthwhile and if so, whether it is possible post-approval, whether other sources of uncertainty will resolve over time without formal research, whether the benefits of research outweigh the costs, and whether the benefits of approval outweigh the costs.

VoI is an obvious analytic tool which can be used to answer the majority of those questions, but more informal, narrative techniques are usually used. The potential use of VoI by NICE was previously examined in a pilot study in 2005.⁸⁵ The same research group also explored its feasibility within the HTA programme in the UK.⁸⁴

The pilot on behalf of NICE⁸⁵ examined the use of the technique with six case studies covering screening programmes and diagnostics, and drugs for both acute and chronic conditions. The scenario was that VoI analyses would be conducted following submission of technology assessment reports. As each technology assessment was already based on a probabilistic decision model, adapting them for use in VoI was relatively straightforward, and the authors were well able to conduct the analyses within the required timeframe (four weeks).

^{xxx} Precise definitions of CER vary: the American Institute of Medicine¹²⁰ definition makes no mention of cost whilst the American College of Physician’s¹²¹ includes relative effectiveness, safety and cost in its definition (both cited in ¹²²).

The key limitations were not with the principles of Vol themselves, but rather the framing of the initial decision question, particularly the inclusion of all relevant alternatives. The authors point out that whilst in the cases considered exclusion of relevant alternatives may not affect the adoption decision,^{xxxii} it has the potential to bias substantially the value of information and research recommendations. Other issues the authors highlighted included appropriate handling of bias from incorporation of evidence from non-randomised study designs, inclusion of ‘unrelated events’, appropriate specification of priors in the absence of evidence (requiring elicitation from experts), exploring uncertainty associated with alternative decision model structures and resource constraints limiting the ability to systematically search and synthesise evidence on all model parameters, not just the treatment effect.

Notwithstanding the comments above, the authors did raise some issues with Value of Information analysis. They identified issues over estimating the appropriate population who will benefit from the information, estimating the value of information for correlated parameters, for patient subgroups and incorporating a ‘value of implementation’ into the analysis too. These limitations are discussed in detail in Chapter 3.^{xxxiii}

The second pilot explored the use of Vol as part of the NHS Health Technology Assessment research programme.⁸⁴ Three case studies were considered: a screening topic (also considered in the report for NICE⁸⁵), and treatments for one chronic and one acute condition. Three core tasks were required for each: 1) construction of a decision model, 2) probabilistic sensitivity analysis and 3) value of information analysis. The authors reported that they were able to complete the analyses within the required timeframe (six person-weeks of work spread over 10-12 weeks of time) such that recommendations could be made to the HTA Board as to where the best returns from research in those disease areas were likely to be.

The authors recommended a need for clear definition of the research problem at the outset, the need to ‘down-weight’ evidence of an inferior quality within a model (an issue raised in the NICE analysis too),⁸⁵ and suggested the Vol approach could be considered alongside the ‘vignette’ approach currently used.

^{xxxii} In general, exclusion of relevant alternatives can lead to erroneous adoption decisions (see Section 1.3.2)

^{xxxiii} I do not consider the Value of Implementation in this thesis. This is a technique for valuing the return on efforts to fully implement current best practice, rather than the return on research efforts: it may be more efficient simply to disseminate current information and encourage uptake of best practice rather than reduce decision uncertainty. Both can be considered simultaneously.¹²⁵

Despite these pilots, neither NICE nor the HTA commissioning board formally adopted Vol to assist their research prioritisation, both describing it as ‘interesting and potentially useful’.¹²⁶ A major barrier to implementation appears to have been lack of comprehension of the technique: most of the respective board members were unfamiliar with the principles of cost-effectiveness, probabilistic analysis and evidence synthesis, let alone value of information. Board members were thus understandably reluctant to adopt a technique with which they were unfamiliar. A further problem however may be the separation of the research and adoption decisions:¹²⁶ methods for the adoption decision have developed considerably whilst those for the research decision still lack transparency. Whilst conceptually separate decisions, they are intrinsically linked (see Section 1.5.2), and should be made simultaneously. This would require a considerable reorganisation of the current R&D management structure in the NHS: If NICE makes the adoption decisions, it should also be the principle decision maker for allocation of research funds too.

1.7. Statement of study problem

In Sections 1.2 to 1.5 I outlined the entire decision making framework within which this thesis is set. Briefly, the entire approach to research into, and adoption of, new technologies should be seen as an iterative process whereby all existing evidence relevant to the decision question is identified and synthesised via systematic review into a decision model. The adoption decision should be based on the point estimate expected incremental net benefit, irrespective of uncertainty, and the uncertainty used to estimate the expected return from further research via Vol analysis. Once new evidence is gathered, the systematic review and decision model should be updated and the adoption and research decisions revisited.

Section 1.6 provided a brief overview of how elements of the framework have been incorporated into decision making processes in a number of countries.

The purpose of this section is to return to the original question of the thesis and to outline a number of pertinent issues to be explored in the following chapters.

The question of this thesis is to consider how much detail is required in economic evaluations alongside clinical trials to optimise evidence for decision making.

By 'level of detail' I refer to the different approaches to costing from 'top-down' gross costing to 'bottom-up' micro costing. These are explained in detail below. By 'optimising evidence for decision making' I mean the level of detail that will maximise the expected return on investment in research. Value of information analysis is a technique compatible with this objective (Section 1.4).

1.7.1. Level of detail in costing in an economic evaluation.

Unlike most clinical outcome measures, total cost per patient is a composite outcome measure of a number of different components. Section 1.3.4 described some approaches to deciding which components should be included. The general approach is to include all those items thought relevant to the decision question. Items common to both arms of a study can be safely excluded as they will cancel out in the incremental analysis, thus collecting and including those data is a wasted effort. I also mentioned issues relating to protocol driven costs and the merits of inclusion and exclusion.

However, each component can typically be disaggregated into a number of smaller and smaller sub-components. At one extreme, a micro-costing (or 'bottom-up') approach is possible, where every activity and consumable is measured and recorded. For example, the

cost of a hospital admission for a surgical procedure can be estimated by recording staff time, units of blood, anaesthetics and other drugs, swabs and other consumables, time spent in theatre, ICU and ward, nurse and medic time attending to the patient prior to discharge, plus appropriate allocation of overheads and capital costs. At the opposite extreme, the cost for the entire episode can be approximated based on national average costs, such as the NHS Reference Costs¹²⁷ (top-down or gross-costing).

The appropriate level of detail is driven by two key issues, firstly the study question itself⁷⁰ but also the cost of data collection. A comparison of two surgical techniques will require micro-costing of the relevant procedures, whereas hospital admissions due to side effects from drug treatments may be sufficiently approximated with a per-diem or per-admission unit cost. A recent review of the methodological issues in costing health services¹²⁸ observed that a costing exercise should itself be cost-effective, based on detailed, comprehensive and representative resource use and unit cost data, be accurate and precise, and be at minimal risk of potential errors and biases (such as measurement and valuation bias, case-mix and service mix bias and site selection bias).¹²⁹⁻¹³² Critically, the authors¹²⁸ point out that these objectives may compete against one another: a cost analysis needs to be provided subject to the research resources available. A micro-costing approach may be the most theoretically robust method, but it may not be practical or feasible in all cases, and a less precise costing method may be substituted. However, this 'inferior' costing method may lead to a biased estimate of cost and thus incremental net benefit, therefore increasing the risk of drawing the wrong conclusion. For pragmatic reasons, most economic evaluations employ a mixture of the two approaches.⁵²

Thus this thesis aims to explore whether, given two (or more) different approaches to estimating (some component of) cost per patient in an economic evaluation alongside a clinical trial, it is possible to estimate the incremental value of the 'superior' process and compare this with the incremental cost to judge whether it is worth collecting data using the superior process. Extending this idea a stage further, it may be more efficient to collect a proportion of the data using one process and the remainder with the other.

In the next chapter, I review the existing literature on efficiency of data collection approaches. I firstly explore whether value of information analysis or any other economics based method^{xxxiii} has been used to help inform the design of clinical trials (beyond sample

^{xxxiii} Defined as any approach that takes into account both the expected cost and benefits of different design choices.

size estimation). Secondly I review the literature on the cost of data collection (where possible focusing on the cost of resource use data collection). Finally I review the data comparing different approaches to gathering the same resource use data.

In Chapter 3 I explore the methods of Value of Information further, exploring recent theoretical developments in the area and criticising the different approaches to calculation. I then focus on a particular practical limitation of such analyses where correlations between input parameters in a decision model are unknown, and frequently assumed zero. This may lead to a biased estimate of the variance of incremental net benefit and thus biased estimates of the value of information.

In Chapter 4 I build on the current applications of value of information by extending the principles to the comparison of different data collection processes in order to answer the question of this thesis. This has been explored theoretically⁴ but I am not aware of any applications of this technique in the health care field to inform the design of clinical trials^{xxxiv} or indeed in any other field. Using an example dataset, I show how the technique can be used to assist decisions as to whether it is preferable to collect exact drug cost data on every patient (including brand, dose and duration of every prescription item), or whether an aggregate cost per patient based on class of medications will suffice.

Finally, in Chapter 5 I discuss the strengths and weaknesses of the entire approach, implications for practice and potential avenues for future research.

^{xxxiv} Although Shavit and colleagues²⁰ explored a related technique with similarities to the payback approach (see Chapter 2).

2. Literature review

2.1. Introduction

Chapter one provided an introduction and background to this thesis, in particular explaining the principles of economic evaluation and value of information analysis. Before exploring the suitability of value of information analysis to address the study question in more detail, I firstly review the literature to answer the following questions:

1. Have value of information analysis or similar economics-based prioritisation processes been used to inform the design of clinical trials other than for sample size estimation?
2. What evidence is there concerning the cost of collecting resource use data alongside clinical trials and/or the cost of conducting clinical trials themselves?
3. What evidence is there comparing alternative approaches to collecting and/or measuring the same resource use data?

The methods section explains the approach to each of the three reviews. The results of each are then presented, followed by a discussion of the issues raised.

2.2. Method

Literature searches based on key terms for methodological studies can be of very low specificity, returning many thousands of irrelevant hits. An alternative approach is a ‘citation pearl growing’ method¹³³ (cited in Dolan et al. 2005⁶⁰) where key ‘core’ manuscripts are identified. Searching then continues by review of title and then abstract of articles in reference lists. The reference lists of any subsequently identified articles are also searched and so on until no further references are identified. The search then proceeds in the opposite direction where manuscripts citing the manuscripts so far identified are reviewed by title and abstract. This is possible using the forward citation feature of Web of Knowledge.¹³⁴ Citations of those citations are reviewed and so on until no further manuscripts are identified. The search is then repeated in the backwards direction (i.e. search of reference lists) on the newly identified manuscripts, then forwards on any further papers identified and so on until no new manuscripts are identified.

For a methodological review it is difficult to define tight inclusion and exclusion criteria a priori. The general inclusion criteria for each review are defined in each subsection below. Citations of articles not indexed by Web of Knowledge (WoK) could not be determined therefore linkage was limited to backwards searches of these articles.

2.2.1. Use of Vol or similar techniques to inform study design.

Inclusion criteria: Any study reporting the use of some concept of efficiency or cost-effectiveness in the design of a clinical trial.

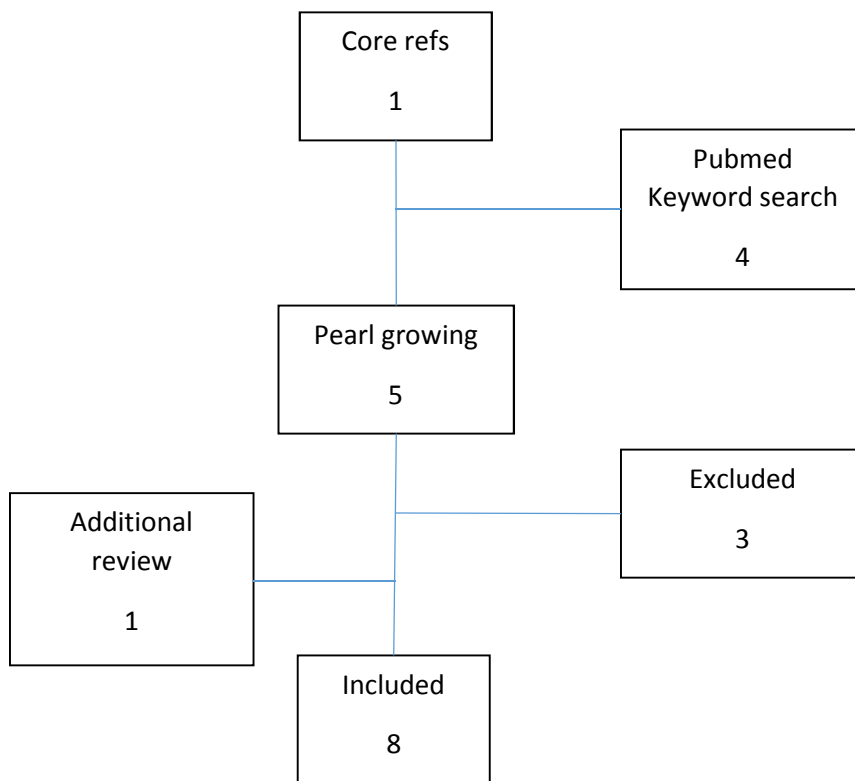
Exclusion criteria: Any study reporting some concept of efficiency or cost-effectiveness exclusively to inform sample size at the outset of a study.^{xxxv}

Core manuscripts: At time of writing I was aware of one article (Shavit et al. 2007²⁰) employing an economics-based approach to informing the design of a study. In order to broaden the base for the core articles, I conducted a scoping search using key terms from this paper and words from the title and abstract. This yielded a further four potentially relevant articles.

The pearl growing technique yielded an additional five potentially relevant studies. Following review of the full text, seven of these ten articles were considered suitable for inclusion in the review. After conclusion of the review, a new systematic review of the methods and applications of value of information analysis was subsequently identified and included in this section, thus eight studies were included (Figure 2-1).

^{xxxv} Studies using economics to inform sequential designs are included

Figure 2-1: Flowchart, review 1



2.2.2. The Cost of collecting resource use data alongside clinical trials and/or the cost of clinical trials themselves.

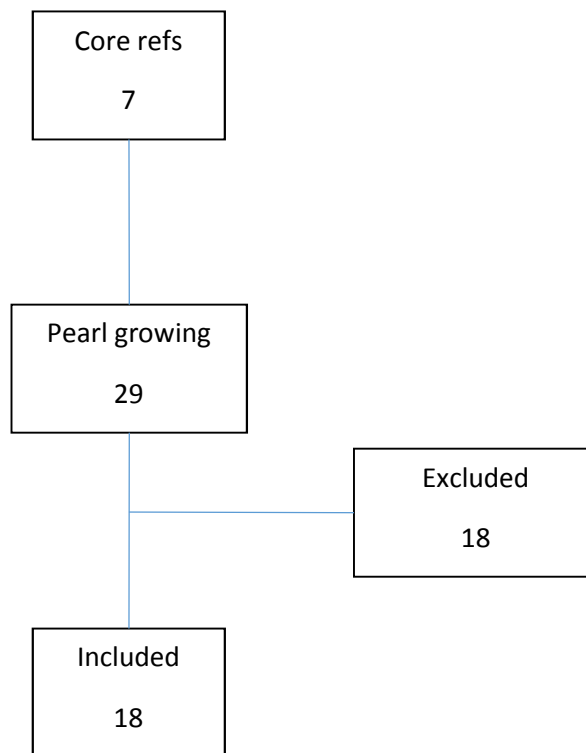
Inclusion criteria: any study reporting or commenting on costs of research within the context of a clinical trial.

Exclusion criteria: None

Core manuscripts: I identified seven potentially relevant studies from my own sources.

The pearl growing technique identified a further 29 studies for inclusion. Review of full text led to a total of 18 studies considered suitable for inclusion in this review (Figure 2-2).

Figure 2-2: Flowchart, review 2



2.2.3. Comparisons of alternative data collection techniques

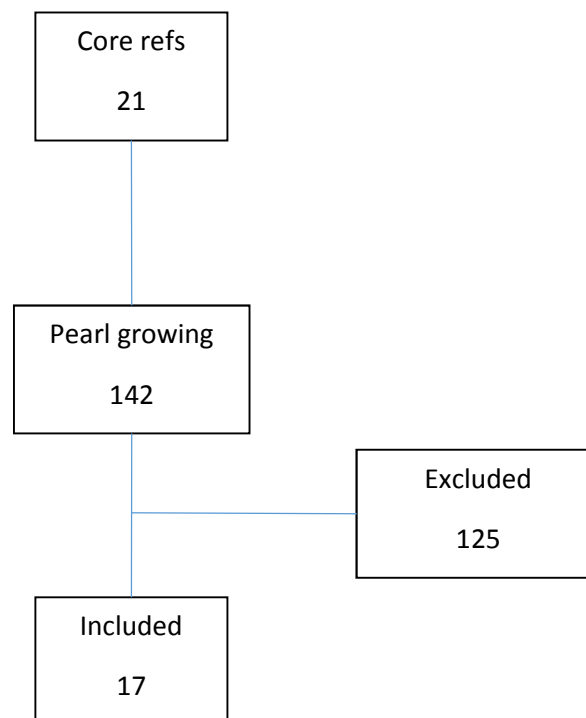
Inclusion criteria: Any study comparing two or more approaches to collecting the same resource use data or data from which resource use could be estimated or inferred, for example self-report of acute events such as MI or fracture. Studies could be RCTs, sub-analyses conducted within RCTs or free standing prospective or retrospective analyses.

Exclusion criteria: Studies not reporting comparisons of alternative data collection techniques. Studies published prior to 1990.

Core manuscripts: I identified 21 source studies from my library which potentially fulfilled the inclusion criteria.

During searching using the 'pearl growing' technique, it became apparent that there were a very large number of potentially valid studies. Therefore searches were limited to a time period of 1990-present (2013). Even then the search yielded 142 potentially relevant manuscripts. A full appraisal and meta-analysis of each of these is beyond the scope of this thesis, thus the titles and abstracts were reviewed and those chosen for inclusion were a) systematic reviews of comparisons of data collection techniques and b) representative examples of comparisons of data collection methods in specific areas and of particular interest. Seventeen studies were therefore included in the review (Figure 2-3).

Figure 2-3: Flowchart, review 3



2.3. Results

2.3.1. Use of Vol or similar techniques to inform study design.

A total of eight manuscripts were identified. These included one systematic review of methods and applications,¹³⁵ and nine reporting to use some concept of economics to assist in the design of a clinical trial (other than sample size determination) including approaches to reducing the cost of enrolling patients into the trial,¹³⁶ structural decisions around trial design^{20 137} (e.g. sequential trial designs,^{138 139} and nested case control vs full cohort analysis¹⁴⁰), and using routine versus *de novo* data.¹⁴¹

Steuten et al.¹³⁵ undertook a systematic review of methods and applications of value of information analysis, grouping the methodological literature into four categories: (1) rationale and basic principles of Bayesian decision analysis, (2) the potential role of Vol in the regulatory process for health technologies, (3) development or optimisation of mathematical methods and (4) additional uses and adaptations to the core Vol methods. The latter category is of particular interest to this thesis. However, of the twelve studies they identify, only one uses Vol to specifically inform the design of a clinical trial (other than sample size), and is reviewed below (Griffin et al.¹³⁹).^{xxxvi}

Moons et al.¹³⁶ describe an approach to reduce the cost of recruiting patients into a clinical trial. Depending on the nature of a trial, it may be more appropriate to spread required diagnostic assessments over several visits. For example some measures such as blood pressure or cholesterol levels may require repeat testing due to day to day variability within a particular patient to minimise confounding by regression to the mean. Furthermore, beginning with simple non-invasive tests before following up with more expensive and invasive diagnostics at a later date may 'weed out' those not eligible for the study prior to exposure to invasive testing, hence reducing exposure to these tests and also reducing cost. Moons et al. describe their approach to obtaining the maximum number of patients randomised with the minimum number of screening examinations, illustrated with the example of a large trial for a cholesterol lowering drug.

The assessment for inclusion in the trial comprised up to five visits spaced one month apart at which cholesterol was measured as well as a battery of questionnaires ranging from a short screening questionnaire at visit 1 to a more detailed one including informed consent at visit 4. Using data on 2200 out of a cohort of 6544 men who had completed the selection period, the authors generated a multivariable logistic model to predict inclusion at each screening visit. Cut-off probabilities were

^{xxxvi} The others concern the value of implementation,^{125 142 143} (a concept not addressed in this thesis: see discussion of Chapter 5), option value of delaying a decision,¹⁴⁴ allocation of research funds,¹⁴⁵ industry perspectives,¹⁴⁶⁻¹⁴⁸ calculation of EVPPi / EVPSi,¹⁴⁹ criticism of CEACs in relation to Vol,¹⁵⁰ and analysis using the 'cost-disutility' plane.¹⁵¹

assigned according to the values which maximised the area under the ROC curve. Using knowledge of the cost of each screening visit, they then estimated the cost per patient successfully randomised in the remaining 4344 potential enrolees and compared this with the predicted cost per patient successfully randomised using the prediction model to exclude potential enrolees earlier in the screening process. In this case, the model was predicted to reduce trial recruitment costs by \$52 per patient randomised (3.6%).

The authors discussed whether their approach was likely to compromise either the internal or external validity of the resulting study. As the selection is conducted prior to randomisation, there should be no threat to the internal validity. However it is conceivable for the prediction model to change the resulting population that finally gets entered into the trial. The authors tested for this in their own application but found no evidence of a shift in baseline characteristics. In conclusion, using data from the first patients screened for inclusion to a clinical trial to predict factors associated with subsequent dropout, which are then used as a further screening tool for subsequent patients has the potential to reduce the cost per patient successfully enrolled in a trial.

An editorial by Falagas and Bliziotis in 2007¹³⁷ made the case that whilst some research questions require specific study designs, other questions can be answered by a number of different approaches. Whilst there have been comparisons of the outcomes of different designs addressing the same question, at the time of writing only two (Hak et al.¹⁴⁰ and Shavit et al.²⁰; see below) explicitly considered the cost-effectiveness of one design compared with another. Falagas & Bliziotis¹³⁷ argue that a well-designed study may fail to answer its question due to being under-powered as a result of resource limitations. Therefore consideration of the cost-effectiveness of alternative designs is essential. Failure to do so risks inappropriate allocation of research resources ultimately slowing down clinical progress.

As mentioned by Falagas and Bliziotis,¹³⁷ Hak et al. in 2004¹⁴⁰ considered the cost-effectiveness of two alternative study designs to estimate the risk of hospitalisation due to influenza or pneumonia or death in an elderly population receiving influenza vaccination. They did not explicitly estimate the costs and benefits of the two study designs. Instead their *a priori* position was that a nested case-control study was less expensive than a full cohort study. Their purpose therefore was to determine whether a case-control study nested within the full cohort study would provide similar estimates of the effectiveness of vaccination. The probability of an outcome (hospitalisation or death) was estimated from the data (either an entire cohort of 20,000 elderly persons followed up over six 'flu seasons or subset comprising the nested case-control). The coefficient against vaccination status on a logistic regression of outcomes (hospitalisation or death) was taken as the treatment effect.

The authors found that the predicted results did not differ substantially and therefore concluded that a nested case-control study may be a cost-effective alternative to an entire cohort analysis.

Publishing three years after Hak et al.¹⁴⁰, Shavit and colleagues²⁰ developed a process to choose between alternative study designs (namely prospective RCT vs retrospective cohort study) to collect additional evidence to reduce decision uncertainty. The authors argue that the key difference between different study designs is the degree of bias inherent in their approach. By quantifying the bias, the estimated treatment effect yielded from different studies can be estimated. The expected opportunity loss of the adoption decision plus the cost of the trial itself allows the total cost of the different study designs to be estimated. The 'net information benefit' is calculated by subtracting the net benefit of the retrospective study design from that of an RCT.

In an hypothetical example, the authors present a situation where value of information analysis has been used to determine the optimal sample size of an RCT to reduce decision uncertainty around whether a new treatment is preferred (i.e. is cost-effective / yields a positive incremental net benefit) compared with current treatment. An alternative to the RCT is a retrospective database analysis of a national disease registry. Five potential sources of 'inherent bias' are identified (representativeness of the sample, selection bias, follow-up period, 'real-life reflection', and accuracy of records) and their likely impact on the observed treatment effect expressed as a normal distribution representing the percentage deviation of the sampled estimate from the true value. Monte Carlo simulation was used to generate an overall mean and distribution of bias from each study type.

The next steps appear to comprise predicting the results of the two study designs, then subtracting the predicted effectiveness of the superior intervention yielded from the RCT by the predicted effectiveness estimate of the superior intervention yielded from the retrospective study. This yields the expected incremental health gain of one study design over the other.

The incremental cost of one study design compared with the other is estimated likewise: the total cost of each is the expected loss due to uncertainty (estimated via the loss function on the prior: see Figure 1-1 in Section 1.4.1) plus the cost of the relevant study itself. Dividing the incremental cost by the incremental effect of one study design over the other estimates the extra cost for every extra unit of outcome, or rearranged into an incremental net 'information benefit' (directly analogous to calculating an ICER and incremental net benefit as described in Section 1.3.8). If this is positive the RCT should be preferred, if negative the retrospective study is preferred.

A sequential trial design is one where the data are analysed after the results of every patient are measured (or after every group of n patients, in which case it is known as a group sequential design).¹⁵² As such it offers the opportunity to assess stopping rules thus minimising unnecessary exposure of patients to the inferior treatment and reducing the total cost of the trial. They therefore provide an obvious means for improving the efficiency of clinical trials.

Thatch & Fisher¹³⁸ describe an approach to estimation of the optimal sample size for a two-stage group sequential trial with a focus on minimising the cost. Importantly, they define the cost of different sample sizes as not only the direct cost of the trial but also “the gain or loss associated with the outcome”. Their approach begins with a standard power calculation based around prior belief of the treatment effect. After completion of each stage of the trial the prior belief is updated with the data from the trial to form the posterior belief. This then becomes the new prior to estimate the optimal sample size for the next stage.

The authors provide a conceptual and worked example from the perspective of a drug company, where the relevant costs are the cost of sampling and the financial gain or loss to company profits as a consequence of a positive or negative result. The approach is to find the optimal sample size for each stage of the sequential trial that maximises the probability of reaching a statistically significant result for the minimal cost.

Griffin et al.¹³⁹ explore the implications of sequential trial designs within a formal value of information analytic framework. They observe that some input parameters within a decision analysis may have a low EVPI when analysed individually, but have a higher EVPI when considered as part of a sequential trial design. Furthermore, collecting information about one or a group of parameters will change the EVPI of the remaining parameter(s). Therefore the overall expected value of a sequence of research studies should be considered as an alternative strategy which may be more efficient than considering individual parameters in isolation.

Given two parameters of interest, Griffin et al.¹³⁹ state that there are four possible research strategies: no research, parameter 1 alone, parameter 2 alone or parameter 1 and 2. With a sequential design, a further two options arise: research parameter 1 then 2 or 2 then 1. As the number of parameters increases the number of possible strategies increases exponentially, thus making for a potentially very computationally expensive procedure if the analysis is conducted by simulation. Indeed the authors do not present a worked example in their manuscript. However, an efficient solution may be calculable by assuming (normal) parametric distributions for the parameters as described in Chapter 1. The value of information of the four strategies comparing simultaneous collection of data can be analysed in one equation determining the optimal sample

sizes for each parameter within a single study: where the optimal sample size is zero, then no research should be conducted on that parameter. Whilst complicated analytically, estimation of the optimal sample sizes for varying sequential designs should be calculable with minimal computer processing time.

Cohen and colleagues¹⁴¹ undertook an analysis of the marginal valuation a mock grants committee placed on *de novo* data collected alongside an RCT compared with data extracted from routine sources. Whilst Shavit et al.²⁰ focused on eliciting the expected bias associated with different study designs, Cohen and colleagues attempted to elicit the intrinsic valuation of data from alternative sources using two methods (willingness to pay [WTP] and implied values [IV]) from a mock grants committee. The authors began with an exercise in which four clinical questions were addressed with four RCTs. They then devised approaches to estimate the outcome measures used in the RCTs with routine data. In each case the RCTs were substantially more expensive than the analyses conducted using routinely available data, but the results were inferior (for example some outcome measures observed in the trials could not be estimated from routine data, data completeness varied from site to site and some cost summaries comprised different components compared with their RCT-based counterparts).

Thus the incremental cost and ‘quality’ of RCTs are in the NE quadrant of the cost-effectiveness plane when compared with analyses based on routine data. The authors therefore set about determining the mock grant committee’s value of the ‘better quality’ RCT data over the routine using the two methods (WTP and IV). The committee was able to generate results using both methods, with higher values measured for the studies using the ‘designed’ rather than ‘routine’ data sources.

2.3.2. The Cost of collecting resource use data alongside clinical trials and/or the cost of clinical trials themselves.

I was not able to identify any studies exploring the cost of resource use data collection *per se*. Within the 18 included studies, nine reported some aspect of the cost of conducting research, particularly with a view to comparing the treatment costs of trial enrolees with non-trial enrolees,¹⁵³⁻¹⁶¹ whilst the remaining nine explored methods of reducing the cost of a trial with minimal impact on quality of the outcomes.¹⁶²⁻¹⁷⁰

Descriptions of the cost of conducting research

In the US, there is evidence that insurers are reluctant to permit their policy holders to take part in clinical trials due a perception that such patients are more expensive to treat. A number of studies

have explored whether patients enrolled in clinical trials are more costly than those receiving routine care.

For example, Evans et al.¹⁵³ compared the cost of two cancer drug trials with the routine treatment patients would receive. Patients in one of the trials cost more than routine treatment, but this was due to increased survival and hence administration of more cycles of treatment, whilst in the other the cost per patient was lower. Kilgore and Goldman¹⁵⁴ found that participation in US cancer clinical trials was associated with “a modest increase in prescription drug utilisation and costs” but did not find evidence to suggest an increase in out of pocket expenditures. Bennett et al.¹⁵⁵ in a small (n=70) pilot study observed similar 6-month treatment costs for cancer patients enrolled in the treatment arm of a trial compared with controls, whilst a 5-year follow-up of 61 patients enrolled in phase II or III cancer trials compared with 61 matched controls¹⁵⁶ suggested a modest increase in costs (maximum 10%) in trial enrolees, concluding that “clinical protocols may add relatively little to [the already high cost of treating cancer].”¹⁵⁶ A larger study of 264 patients enrolled in a trial of an acute MI treatment protocol compared with 335 matched controls did not find any significant difference in initial hospitalisation cost.¹⁵⁷ Finally, in HIV medicine, an examination of drug costs prior to, during and following enrolment in clinical trials found that drug costs were lower during the trial, but rose on conclusion of the study.¹⁵⁸

In conclusion, there is little evidence to suggest that patients enrolled in clinical trials systematically incur higher treatment costs than those undergoing routine care, although the generalizability of these results is unclear and costs will vary on a case by case basis. However, these studies only consider any additional treatment costs of patients (e.g. additional monitoring or more intensive drug regimens). They do not include any research protocol driven costs, specifically researcher time to collect, record and analyse data. Writing in 2003, Emanuel et al.¹⁵⁹ sought to estimate the cost to a centre of enrolling and following up a patient according to a mock trial protocol. They estimated the cost of a trial involving 20 subjects with 17 office visits each would cost approximately \$6,000 per patient, excluding overhead costs, of which one third is non-clinical costs (although an accompanying editorial suspected this \$2000 was an underestimate¹⁶⁰). A contemporaneous Canadian study (Roche et al. 2002¹⁶¹) collected timing data from 83 clinical research associates (CRAs) on 41 distinct research-related tasks. Significant predictors of time input included phase of the clinical study and sponsor, with early phase and industry sponsorship associated with higher time input from CRAs.

Methods to reduce the cost of clinical trials

There is some concern in the literature about the cost of clinical trials and the need to explore options for minimising cost. In a 2010 editorial, Sargent et al.¹⁶² point out that cancer trials are increasing in complexity and hence cost, and there is thus a need to “distinguish between necessary complexity... [and] unnecessary complexity [such as] excessive adverse event reporting, on site monitoring and eligibility criteria”. Multinational trials add another dimension of complexity and cost to the mix. Demol & Weihrauch¹⁶³ argue that due to differences in health care systems, medical cultures and treatment strategies, enrolling centres in different countries is more costly than enrolling the equivalent number of centres in just one country. For such a trial to succeed requires “a well-coordinated multidisciplinary team and an effective project management.”¹⁶³

Urban et al.¹⁶⁴ describe how the predicted cost of a very large randomised controlled trial (the women’s health trial; WHT) examining the impact of reduced fat intake on the risk of breast cancer with 10 year follow-up was reduced from \$200m to \$95m by estimating research cost as a function of various features of trial design. Approaches to reducing cost were then explored.

Five design parameters were identified: a) the sample size, b) proportion of the sample allocated to intervention, the number of women c) initially contacted and d) subsequently screened, and e) the number of clinical centres involved in recruitment and follow-up of patients. Five components of cost were also identified, broadly relating to the five parameters, comprising a) establishing and maintaining each centre, b) recruiting participants, c) screening d) administering the intervention and e) follow-up of participants. Overall costs were estimated based on a consensus approach (adapted Delphi), and divided into fixed costs, recruitment costs, screening costs, variable intervention costs and variable research costs.

From these estimates, the authors searched for the most efficient trial design that minimised research cost whilst maintaining 80% power for the trial. Examples of changes that were made following the cost analysis included seeking out bulk contracts for printing and mailing of invitation letters, use of less expensive staff grades for routine operations such as telephone calls to follow up contacts, conducting initial follow-ups as a group session, eliminating two follow-up visits, and analysing only a sample of rather than all dietary records provided by participants. They conclude by stating that their analysis cost less than 1% of the total direct costs of the trial, yet halved the overall cost.

Thornquist et al.¹⁶⁵ adopted the same approach to minimise the cost of another large scale prevention trial (Carotene and Retinol Efficacy Trial; CARET), measuring the effect of these two compounds on incidence of lung cancer in two high risk populations (asbestos-exposed workers and

heavy smokers). The design was a placebo-controlled RCT of six years follow-up in approximately 18,000 individuals.

Three issues arose whilst the trial was underway. Firstly, in order to achieve the target person-years of exposure, the researchers discussed whether to increase the follow-up time with the currently recruited patients or to open a new trial centre and enrol more patients. Secondly, given a decision to open a new centre, the timing of its opening needed to be determined, and finally, throughout the trial, there was a need to improve adherence of patients to the comparator treatments.

A cost function was devised using the same methods as Urban et al.¹⁶⁴ Four alternative designs, a base case and three alternatives incorporating the three issues were then costed, and the predicted outcomes estimated. The authors estimated both the total costs of each of the options and also the expected outcomes (in terms of person-years of follow-up and (expected) number of outcomes (lung cancer cases and death) observed). They report that extending the follow-up of existing patients instead of opening up a new study centre would be less expensive, but according to the data reported, would lead to a lower overall person-years exposure. Due to the make-up of the population though, a higher number of cases of lung cancer and deaths would be expected. In their report however, the authors concentrate solely on cost without mentioning the difference in outcomes observed (the data themselves are reported in a table). They state that ultimately, whilst “the cost analysis was informative[, it] did not drive the decision”. Interestingly the difference in outcomes did not drive the decision either, instead the authors explain that extending the study was considered to be too much to ask participants, and that opening a new study centre would act as an ‘insurance policy’ reducing the risk of failure to recruit sufficient participants.

The decisions analysed in the Urban¹⁶⁴ and Thornquist¹⁶⁵ articles are picked up by Allison et al.¹⁶⁹ who describe methods of study design that consider statistical power and cost simultaneously, principally by introducing a budget constraint and maximising the power subject to the constraint and vice versa (minimising the cost subject to a given power). Approaches they consider include selecting optimal cut-points in screening tests; optimal allocation between different treatment arms; choosing between increasing the sample size and replicating measurements on existing subjects and using covariates as explanatory factors.

They conclude that the approaches may have application in a variety of settings, depending on the research questions. An area they briefly mention is ‘measurement quality’, the example they cite being the use of more highly trained raters or observers (e.g. interviewers to elicit data). They

suggest that 'better quality' (and more expensive) data may lead to an increase in statistical power for a given sample size (and thus be cost-neutral^{xxxvii}).

More recently in 2005 Eisenstein et al.¹⁶⁶ managed to reduce total predicted costs of two cardiovascular trials by over 40% by modelling costs under a number of assumptions and performing sensitivity analysis (or 'stress testing'), principally focusing on reducing management complexities. Similarly in 2008, the same lead author¹⁶⁷ sought to review current practices in clinical trials and to identify areas where costs could be reduced without compromising scientific validity. The authors applied the principles to a proposed 'mega-trial' with a predicted cost of \$427m comprising 20,000 patients at 1,000 sites. Introduction of electronic data capture and changes to proposed site management structure led to a 59% reduction in predicted trial cost, potentially rising to 90% if "an even more streamlined trial design than has typically been considered for regulatory submissions in the past" was adopted.

Where trials are struggling to recruit, increasing the number of subjects allocated to the treatment arm of a trial may increase interest in a study and hence recruitment due to a prior perception amongst the public that the treatment arm will be more beneficial. However, such a tactic requires a larger overall sample size to maintain the same statistical power with consequent impacts on recruitment time and trial cost. Vozdolska et al.¹⁶⁸ explored the impact of alternative allocation ratios on statistical power and cost of recruitment: the unequal allocation between arms requires a greater overall number of participants, which of course may take longer and cost more to enrol. Investigating a number of example scenarios, they find that if allocation between arms is on a 1.5:1 ratio, recruitment is time neutral only if the recruitment rate increases by at least 4%. Where patients are allocated to arms on a 2:1 ratio, the recruitment rate needs to rise by 12% in order not to extend the recruitment period. However the requirements for cost-neutrality are much higher, where a 13% and 47% increase in recruitment rate are needed under each scenario respectively. The authors recommend that such a tactic should only be considered where there is clear evidence that unequal allocation will increase recruitment rates.

Finally, an alternative approach to conducting clinical trials is presented by Vickers & Scardino.¹⁷⁰ Their proposal is to integrate randomised trials into routine clinical care, based on the principle that "the clinical experience of the patient and doctor should be indistinguishable from routine care, whether or not the patient is randomised."¹⁷⁰ They give a number of examples of where their technique could be applied, which may well reduce the cost of recruiting patients to a study,

^{xxxvii} The authors use the phrase 'situations in which power and costs can be optimised'.

although the main drivers of this appear to be the use of electronic data capture and integration of medical records databases with research databases.

2.3.3. Comparisons of alternative data collection techniques

As stated in the methods, over 100 potentially relevant manuscripts were identified comparing alternative data collection techniques. The included sample comprises six systematic and other reviews¹⁷¹⁻¹⁷⁶ and 11 primary studies considered to be of specific interest and relevance to this thesis.^{55 177-186}

Systematic reviews

A total of six reviews were identified including two on the accuracy of self-reported history of cancer screening tests,^{172 173} three of self-reported health service use in general,^{171 174 176} and one studying the effect of questionnaire design on recall of pharmacological treatments.¹⁷⁵ It is notable that none of these reviews explicitly considered the costs of the alternative methods, most simply stating as fact that data collection is an expensive process and self-report may be a convenient alternative to other more systematic approaches.

Rauscher and colleagues¹⁷³ undertook a meta-analysis of the accuracy of self-reported cancer screening histories, calculating the sensitivity and specificity for self-reported history of mammography, clinical breast exam (CBE), Pap smear, prostate specific antigen (PSA), digital rectal exam (DRE), faecal occult blood test (FOBT) and colorectal endoscopy compared with 'documented screening history'. The literature search covered the period 1966 to 2005, identifying 55 potentially relevant articles. Ten were excluded on the basis of their being conducted outside the USA and 16 were excluded as accuracy measures were not (fully) calculable. Thus analysis was conducted on 29 studies.

The authors' results suggested highest sensitivity for recall of mammogram, CBE and Pap smear (0.95 to 0.93) and lowest for PSA and DRE (0.71 & 0.75). Specificity was highest in endoscopy, FOBT and PSA (0.9 to 0.73) and lowest for CBE, Pap smear and mammogram (0.26 to 0.61). Amongst different population groups, the authors report that studies enrolling mostly Black and Hispanic respondents tended to report lower accuracy of recall (both sensitivity and specificity) than studies enrolling mostly White respondents. With regard to the US context, the authors conclude that when the results of their meta-analysis are applied to screening rates estimated from the National Health

Interview Survey, raw results may overestimate the overall prevalence of screening across the USA, and underestimate the disparities between different ethnic groups.^{xxxviii}

Howard et al.¹⁷² also report a meta-analysis of the accuracy of recall for cancer screening tests compared with medical record data, focusing on Pap smear and mammograms only. Thirty-seven studies were included in the meta-analysis. Most were US-based, but six were from other countries (Australia, Canada & Sweden). Studies included age ranges from 18 upwards (Pap smear) and 40 upwards (mammography), different ethnic backgrounds (mostly White, African-American or Hispanic) and different socioeconomic status.

Point estimate sensitivity and specificity of recall of Pap smears was 0.947 and 0.474, and for mammography 0.949 and 0.618. The authors were not able to account for between study heterogeneity when adjusting for population characteristics (ethnicity and socioeconomic status), length of recall and quality of medical record, thus they were not able to infer any differences in recall amongst different population groups. However, overall results are consistent with Rauscher,¹⁷³ suggesting that women tend to over-report their history of Pap smear and mammography screening.^{xxxix}

Gama et al.¹⁷⁵ explored the impact of questionnaire design in more detail, focusing on recall of pharmacological treatments. Searching Pubmed, EMBASE and the Cochrane Library from inception to 2007, eight studies of drug use comparing different questionnaires or means of questionnaire administration were identified.

^{xxxviii} Rauscher et al.'s¹⁷³ analysis excluded studies conducted outside the USA to answer a US-specific question, i.e. whether a national health survey based on self-report is likely to provide accurate data on cancer screening rates. This limits the generalisability to other settings; in particular, it would be of interest to explore whether there were any systematic differences by country. It is interesting to note that the sensitivity was highest and specificity lowest for female screening tests (mammogram, CBE, Pap smear), whilst sensitivity was lowest and specificity highest for male screening tests (PSA & DRE). The higher sensitivity in females suggests females are better able (or willing) to recall screening tests, yet the lower specificity suggests females are also more likely to over-report screening. This raises interesting hypotheses as to the differences in attitudes towards screening between males and females, a discussion of which is beyond the scope of this thesis. The authors suggest that known underreporting amongst men is a likely explanation, but found no evidence of this in their analyses. They also suggest a lower specificity (over-reporting) for Pap smears could be due to women mistaking a routine gynaecological examination without a test for one that includes one. An alternative explanation the authors consider for the observed low specificity of CBE and DRE is poor reporting for these exams in clinical records. They note that CBE and DRE are not reimbursed well, reducing the incentive to record tests accurately, thus the observed specificity may be biased downwards.

^{xxxix} As per Rauscher, Howard et al.¹⁷² suggest that over-reporting of Pap smears in particular may be due to confusion of any examination with a screening test: mammography may be less susceptible to over-reporting due to the nature of the test itself, requiring a specific visit to a radiology unit rather than a general gynaecological unit (where different investigations including the Pap smear take place) leading to less risk of confusion in recall. They also suggest that over-reporting may be due to the social desirability of screening, and that the wording of the questionnaire may have some influence over the false positive rate.

Studies were set in the USA (two studies), Netherlands (two studies), Switzerland, Canada, Mozambique and Finland (one in each). The studies each varied substantially in purpose and approach: five explored the effect of prompting respondents for specific drugs or indications, two examined the effect of question or response order, and one explored the use of memory aids to assist recall. Four studies recruited members of the general population, whilst the others studied women post-partum (two studies), university students and hypertensive patients. Half of the studies collected information by face to face or telephone interview whilst the other half relied on self-completed questionnaires. Recall period varied from current consumption of drugs to those used during a previous condition.

The included studies reported that asking for specific indications or drug names and the use of pictures or lists of medicines following an indication-oriented question leads to a higher reported prevalence of drug use compared with open-ended questions. Asking questions on specific drugs followed by an open-ended question was associated with higher recall of use of the specific drugs than when the open-ended question preceded the specific. Finally the order in which a list of drugs is presented appears to affect response with options appearing at the top of the list more likely to be selected ('primacy effect').

In discussion, the authors state that the design of a questionnaire is one of several known sources of bias in surveys. In addition to the effects described above, Gama et al.¹⁷⁵ identify a number of other known limitations, namely that excessively lengthy surveys can induce respondent fatigue resulting in inaccurate responses, interviewer administered questionnaires lead to more reliable results than self-administered, and that open-ended questions should, in general, be avoided. Given the variability in responses according to the means by which questions are asked, the authors recommend that journal editors should request full access to questionnaires used to allow assessment of the validity of the instrument.

Whilst Gama et al.'s¹⁷⁵ study shows there is a difference in reported prevalence of use of different drugs according to the manner in which the question is asked, they do not appear to show whether there is a difference in overall accuracy – that is, whilst the sensitivity of one approach may be higher than another, the specificity is unknown. Indeed the word 'specificity' does not appear at all in the review. Thus it is not possible to know whether the increase in reported prevalence of use of drugs is true omission or a false-positive. This has obvious implications for economic evaluations leading to potential overestimation of the cost of drugs (although if the absolute bias in each arm of a study is equal the estimate of the incremental cost will be unbiased).

Evans and Crawford¹⁷⁶ conducted a review of the literature to provide guidance as to when patient self-report data can provide reliable estimates of resource use and when their use can lead to erroneous cost-effectiveness conclusions. In general, the authors noted that under-reporting of events was more common than over-reporting. Recall appeared to be better over shorter time periods and with more salient events (e.g. surgical procedures vs a routine GP visit). However, medication recall was less accurate compared with, say, hospitalisation. Interestingly (and perhaps unsurprisingly) the perceived social acceptability of a condition also influences patient reporting.

The authors conclude that with respect to cost-effectiveness, the major concern is with validity rather than bias. For example, over long recall periods there may be some vagueness about specific resource use episodes. This may affect the mean cost per patient estimated in each arm, but in general if such vagueness is equally spread between treatment groups, it should not affect the incremental cost-effectiveness ratio.

Bhandari et al.¹⁷⁴ aimed to establish the accuracy of self-reported health care utilisation, identifying a number of factors that affect the reliability of such data. Building on and complementing the Evans and Crawford review,¹⁷⁶ the authors identified 42 studies. The factors having the greatest impact on accuracy of self-report data are the sample population and their cognitive abilities, the recall time frame, type of health service use, the frequency of use, design of the questionnaire, mode of data collection and the use of memory aids or probes. In order to assist analysts, they suggest considerations to be taken into account to mitigate the effects of each of these on accuracy of recall.

Finally, Evans and Crawford, in a second review considered the accuracy of data collection methods in prospective economic evaluations,¹⁷¹ focusing on three areas: the use of self-report data, surrogate respondents and mode of administration of questionnaires ('recall', 'proxy' and 'mode' respectively). They discuss how each leads to different levels of internal validity within a prospective study as well as the measurement validity (defined as "the ability of a measure to assess what it was designed to assess"¹⁷¹). The authors recommend an *a priori* assessment of the likely effects of different approaches to measurement on the validity and bias of the results. They cite an example of a study of a terminal disorder where non-response due to death can be anticipated and data loss minimised by proxy reporting, but as proxy completion may not be accurate, data can be collected by both patient and proxy, and the proxy responses adjusted for any systematic difference between the two.

In summary, Evans and Crawford¹⁷¹ provide a useful categorisation of issues to assess the accuracy of data collection methods: 'recall', 'proxy' and 'mode'. Systematic reviews suggest there is some

evidence of recall bias when extracting resource use data from patients compared with medical records, the extent of which may vary by gender and ethnicity. Questionnaire design also has the capacity to affect responses, with evidence of primacy effects and prompting shown to alter recall.

Primary Studies

I now review a number of primary studies not covered by the systematic reviews described above and selected for their particular interest and methodological importance. Studies cover areas including the use of cost diaries to measure resource use,^{177 178} comparisons of the reliability of self-report versus medical records specifically in a UK setting,^{55 179 180} and comparisons of top-down vs bottom-up costing approaches^{181 182} as well as other levels of aggregation in cost analyses, including collection of a restricted data set to approximate full data.¹⁸³⁻¹⁸⁶

Resource use diaries versus questionnaires

The development, feasibility and construct validity of a cost diary to measure direct and indirect costs have been explored following use in two randomised controlled trials.¹⁷⁷ The authors cite evidence¹⁸⁷⁻¹⁹¹ in support of (prospective) diaries compared with (retrospective) questionnaires, with reduced recall bias and higher completion rates. Their diary was intended for continuous completion, with each diary filling a single page, divided into four columns, one for each week of the month. Patients were sent six at baseline and again at six months, with a request to return those completed to date every three months in a pre-paid envelope. Therefore the period of completion was a year. Non-responders were prompted with a telephone call.

As keeping a diary continuously for 12 months may be seen as somewhat onerous, the authors also simulated the impact of partial collection (e.g. for a period of 2 weeks every 2 months) from the full data, with extrapolation to fill in the missing periods, and compared these estimates with the full data. They also compared the self-reported data with insurance company records as a validity check.

Over both RCTs, 85% of patients completed at least one diary (covering a 4 week period), and 68% of diaries were returned, yielding an average of 32 weeks of data per patient (out of a maximum of 52). Fifty percent of patients completed all diaries but 15% completed none (mainly due to withdrawal from the trials). There was no significant difference observed in cost estimates based on complete data versus partial completion, although the correlation between the estimates was only moderate. This could be explained by low statistical power. When comparing self-reported data with insurance company records, there was agreement between some visits (e.g. specialist contacts), but quite some disparity in others (physiotherapist visits, where there was a 10-fold difference between self-report and insurance company records: 5.4 vs 0.55 visits per patient). The authors explain this as

due to half of the patients consulting a physiotherapist weekly which for some reason was not on the insurance company records.

In conclusion, for chronic diseases the authors recommend collecting data for an abbreviated period of time rather than for a whole period such as a year or more. However, they caution against this approach for acute conditions where resource use may well be time dependent (and indeed in chronic disease with seasonal effects).

Merlo and colleagues¹⁷⁸ compared the use of a self-administered questionnaire with a personal diary for use of hormone therapy in 16,060 women aged 45-73 years in a prospective cohort study in Malmö, Sweden. They consider that a personal diary may be more appropriate than a questionnaire for collecting these data, but that feasibility and cost may lead to a decision to use a questionnaire. For the purposes of calculating sensitivity and specificity, they assume a personal diary represents the gold standard.

The questionnaire comprised an open-ended question asking respondents “which medicines do you use on a regular basis?” The diary was a structured 7-day form including an open-ended question on drug use. In both cases, participants were asked to list current hormone use.

Overall agreement between the two methods was high (kappa 0.84, $p < 0.001$), although reported prevalence of hormone use was higher with the questionnaire than with the diary, except in the case of women also taking anxiolytics, hypnotics and opiates where recorded prevalence was higher with the diary. Agreement was lower amongst women in their 50s compared with those in their 40s, but interestingly was highest amongst women aged 70-73. Agreement was also lower for those with high alcohol consumption.

If the diary is assumed to be the gold standard, the sensitivity and specificity of the questionnaire is high at 0.977 and 0.845 respectively. The authors conclude that a simple self-administered questionnaire comprising a single open-ended question on drug use is reliable and highly valid compared with a diary in order to ascertain current hormone use.

An obvious criticism of the method is the assumption of a personal diary as a gold standard for use of hormone therapy. The authors cite evidence in support of this¹⁹² but arguably the most reliable approach would involve biochemical measurements (blood tests or similar) to detect levels of the drugs in the body. However depending on the half-life of the drug in the body, and variability between individuals themselves, such an approach may not be a sufficient proxy to measure strict adherence to every drug consumed. Furthermore very regular monitoring would be inconvenient to the patient, as well as time consuming and expensive. Whilst the authors do not explicitly consider

the incremental cost of a diary approach compared with the questionnaire, knowledge of the relationship between the two methods (i.e. sensitivity and specificity of one compared with the other) could in principle be used to estimate the impact on decision uncertainty in a model to then estimate the incremental gain from the diary over the questionnaire. This idea is explored and developed in Chapter 4 of this thesis.

Questionnaires vs medical records: UK evidence

Several studies have compared cost estimates obtained from patient self-reporting (e.g. questionnaire) with those from hospital or primary care records in the UK.^{55 179 180}

Patel et al.¹⁷⁹ found generally good agreement between self-completed postal questionnaires (the client service receipt inventory, CSRI¹⁹³) and GP records when asking about the number of GP contacts over the previous six months in a sample of primary care attendees (3.03 vs 2.99 per patient according to GP records and CSRI respectively, correlation 0.76). However agreement between the cost of those visits was poorer, due to differences in reported average visit length.

Byford et al.⁵⁵ compared patient self report health service use (CSRI¹⁹³) at 6 and 12 months with GP records in an RCT of treatments for adults with recurrent deliberate self harm. The authors found agreement for overall *incremental* cost observed in the trial to be high, but at a lower level of aggregation, agreement was poorer. In particular, the mean cost per patient estimated using the CSRI was significantly higher than that based on GP records (£2185 vs £1371, 95%CI of difference: £423, £1204). The authors observe that GP records generally provide more reliable information on primary care contacts than patient self-report data, but less reliable information on other care contacts. They conclude by cautioning against the use of primary care records as a source of data on hospital and community services use.

The period over which patients are asked to recall health service contacts is, *a priori*, likely to affect reliability, with memories fading over longer recall periods (recall bias). Petrou et al.¹⁸⁰ sought to establish whether the accuracy of self-reported health service use was a function of the duration of the recall period (as well as the prominence of the service contact). They also sought to identify any socioeconomic or clinical factors that may influence the accuracy of self-report data.

Data comprised the control arm of an RCT of an intervention in primiparous women at risk of postnatal depression, the control comprising routine care for the patients (n=82). Resource use data were collected via face to face interviews at 4 and 12 months postpartum. The interviews comprised structured closed-ended questions delivered by a trained interviewer covering primary care and community midwife contacts, attendances at A&E, outpatient contacts and inpatient admissions.

The first interview related to the previous four months, and the second to the previous eight. The primary, secondary and other provider care records of the control arm patients were obtained and the cost estimates collected using both methods compared.

The authors found that self-reported resource use may be unreliable for longer recall periods and for heavy service users. Particularly salient events such as attendance at outpatient and inpatient admissions tend to be more accurately recalled, although they caution against over interpretation of their findings due to the small numbers of events observed in their sample.

Top-down vs bottom up costing

As I explained in Chapter 1, two key costing approaches are top-down and bottom-up. Chapko and colleagues¹⁸¹ explicitly compared the two approaches using veterans' association (VA) data in the USA. They compared data for 14,915 patients at 72 facilities in 2001 in terms of total annual cost per patient plus the cost of specific services (for example clinic visits, inpatient admissions). Cost was calculated using two approaches. The bottom-up approach extracted data from the US Department of Veterans' Affairs Decision Support System, comprising local costs for specific items. The top-down approach was calculated using a costing system devised by the VA Health Economics Resource Center. Agreement between the two systems was variable, with a correlation on 0.85 for total annual cost, but 0.24 and 0.77 for the cost of outpatient and inpatient admissions respectively. Agreement between the systems increased with the level of aggregation. The authors concluded that those conducting cost analyses needed to carefully consider the purpose and relative merits of different approaches when choosing an appropriate method.

Tan et al.¹⁸² compare bottom-up micro-costing with top-down gross costing for hospital services (appendectomy, normal delivery, stroke and AMI in 2005). They define three approaches: bottom-up microcosting is the identification, measurement and valuation of all patient specific resource items with hospital specific unit and overhead costs. Top-down micro-costing is as per bottom-up microcosting, but national average unit costs are applied in place of local ones. Finally gross costing is based on inpatient length of stay and hospital specific unit costs alone.

The authors found that top-down approaches tended to lead to higher cost estimates than bottom-up. For three of the interventions there was no significant difference. However, cost estimates for appendectomy were significantly higher using the gross costing method. They recommend the use of bottom-up costing for the largest components of total cost, for example labour costs for particularly labour intensive procedures, or components of hotel costs for admissions with a particularly long length of stay (e.g. stroke).

Other levels of aggregation in costing studies

Geue et al.¹⁸³ compared five different methods to calculate the cost of a 'continuous inpatient stay' (CIS) with a view to exploring whether regression models based on each method yielded a difference in the significance and magnitude of explanatory variables. This would have implications for the econometric modelling of costs in future studies. The methods compared are (1) application of the English HRG tariffs (comprising a fixed cost plus a *per diem* component for excessively long stays); (2) application of the Scottish HRG tariff (based on a fixed cost alone); (3) *per diem* costing; (4) per episode costing with a fixed and variable component; and (5) per episode costing using a fixed component alone.

When comparing HRG-based and per-diem based costing approaches, the authors did observe some difference in estimated effect size for regressors. They also observed more variability in per-diem approaches (to be expected given the larger possible range of values for length of stay compared with the number of episodes of care). They point out that costing based on a pure per-diem approach ignores the nature of a hospital stay, characterised by "fixed costs being independent of length of stay and variable costs varying with LoS". Alternatively, the per diem costs in the first day or two of an admission are frequently much higher than the remaining days as medical interventions reduce in intensity, leaving only 'hotel costs'.

The authors found that whilst general conclusions such as elderly and male patients being more costly seem to hold irrespective of the costing approach, the magnitude of the cost differences is dependent on the costing method. They state this has implications for studies exploring the impact of ageing or end of life care costs. Finally, the authors recommend the first method (as followed by the English HRG system) as the most appropriate (disease specific costings incorporating a fixed and variable component).

Chapko et al.¹⁸⁴ considered the question of which resource use items needed to be tracked in an economic evaluation alongside a clinical trial, and whether centre specific or national average unit costs would be preferable. They recalculated costs per patient in an RCT of adult day health care (ADHC) vs treatment as usual in patients at risk of admission to nursing home in the US veterans' administration (VA). The original method used in the trial comprised a mix of local and national costs. Costs were then recalculated using facility level costs and national costs and the results compared. Using facility level costs resulted in higher estimates of per patient costs in each arm of the trial, but the bias appeared to be approximately equal in each arm, thus the estimate of the incremental cost was very similar irrespective of the method employed. After examining the costs in more detail they conclude that the services to be tracked include the intervention itself, the services

which account for 'a good deal of the variance in total cost', and any specific smaller items which are likely to be affected by the intervention. In terms of the sensitivity of the results of an economic evaluation, in their case study there was very little difference in incremental cost irrespective of the costing method, suggesting an ICER would be insensitive to the costing approach. However, they advise conducting sensitivity analysis to establish this.

Knapp and Beecham¹⁸⁵ also considered whether collecting less resource use data affected the accuracy of resulting cost estimates, given the expense of collecting the data. The authors apply a 'reduced list' costing based on the CSRI¹⁹³ in two case studies in mental health: psychiatric reprovision in the community for long-stay hospital residents and caseloads of community psychiatric nursing teams. They find that concentrating on the top five cost items, between 91 and 94% of the total costs are predicted. These results appeared robust when simulating data collection in the two case studies, as well as in three further applications. However, they caution against the use of the reduced list when (i) an analysis specifically requires full costing, (ii) the impact of a policy change on 'seemingly peripheral services' is relevant to the decision, (iii) when inter-individual differences in cost are of primary concern.

Whynes and Walker,¹⁸⁶ explored the use of Knapp and Beecham's¹⁸⁵ reduced list approach in an acute care setting (colorectal cancer surgery). Their original study comprised follow-up of 360 patients treated for CRC. Data on the number and nature of pre-operative diagnostic procedures, length of hospital and ICU stay, duration of surgical procedure, blood use and other diagnostic requests were all extracted from patient notes covering a period of three years from index procedure (or until death if sooner). Unit costs were based on local centre costs where possible, with other sources (e.g. national costs) used where necessary.^{xi} Of particular interest to this thesis, the authors estimate the researcher resource requirements to collect the data at one year WTE.

The authors' objective was to "identify a formulation of crude costs which best approximates [the detailed costing approach described above]". Replicating Knapp and Beecham's¹⁸⁵ approach, Whynes and Walker¹⁸⁶ found it to generate far less accurate predictions of total cost: a crude costing approach based on length of stay multiplied by a daily rate produced much better predictions, with only 12% of the variation between the crude and 'true' cost remaining unexplained. They conclude that the reduced costing method did not fare well in this example, and that a crude costing approach based on mean costs per specialty appears adequate with large samples and when 'the frame of reference is the aggregate'¹⁸⁶ (that is, the interest is in total cost per patient rather than some more detailed component).

^{xi} This work pre-dates the NHS reference costs

2.4. Discussion

This review set out to explore the literature with a view to answering three questions of pertinence to this thesis. Issues arising in the literature relating to each question are discussed in turn below. I conclude by drawing together the three areas, considering whether the existing literature does indeed provide adequate answers to the questions, and thus where the gaps are.

2.4.1. Use of Vol or similar techniques to inform study design.

Whilst some studies mention the importance of cost-effectiveness in the design of a clinical trial,¹³⁷¹⁴⁰ very few studies use a formal economics-based technique to choose between different aspects of study design. Only four studies appeared to do this. Two were based on the principles of value of information analysis,²⁰¹³⁹ one sought to elicit the willingness to pay of a grants committee for 'better quality' information¹⁴¹ and one appeared to take into account some concept of the opportunity cost of research to inform research design.

Hak et al.¹⁴⁰ compared a nested case-control study with a full cohort analysis. An obvious limitation of this study is the fact that it was an ex post study: the case-control data were drawn from a sub-sample of the entire cohort. Thus in order to conduct the case-control study, the entire cohort study needed to be conducted too. As the two options are not mutually exclusive it is not possible to realise any increase in efficiency of research funds from this particular example! The authors do point out though that given their results, such case-control studies could be considered as a valid alternative to RCTs, their logic being the apparent lack of bias of the smaller subset compared with the full cohort study. However, this may be pushing the interpretation of their results further than is justified given the analysis: they did not compare an RCT design with case-control study.

Shavit et al.'s²⁰ approach appears to be a logical means by which the expected value of one study design can be compared with another. However, the lack of a fully worked out numerical example hampers comprehension of the approach. Correspondence following publication of the manuscript challenged the use of two thresholds: one for a willingness to pay for a 'unit' of information in the research decision and another for a willingness to pay for a unit of health outcome in the adoption decision.¹⁹⁴ The author (Grandjour) suggested that the same threshold should be used for both research and adoption decisions. This is logical if it is assumed the sole purpose of information is to reduce decision uncertainty thus increasing the expected health gain associated with the decision. However in response Shavit and colleagues disagree,¹⁹⁵ arguing that the societal willingness to pay for information may well differ from that it is willing to pay for health gain, the reason being that information for its own sake may have consumption qualities (stemming from a desire to feel 'in control') as well as production qualities (use in reducing decision uncertainty in health care). This is an interesting area for further research, but beyond the scope of this thesis. Grandjour also suggests

that the analysis take account of decision makers' risk aversion. The arguments in favour of risk neutrality in public sector decision making are fairly persuasive,⁹⁶ however health care decision making, particularly in the NHS is extremely risk averse. I discuss this further in Chapter 5.

A key aspect of Shavit and colleague's²⁰ approach is the quantification of the likely bias associated with the different study designs. The means by which the expected impact is assessed is not explicitly stated but appears to be narrative approach based on the authors' opinions. This is reasonable as in a real application of this process, such information could be elicited from experts, by review of the literature or use of pilot studies. Thus Shavit and colleagues have a potentially viable method to assist the choice between overall trial designs to provide evidence for decision making.

There are three important criticisms of Thach & Fisher's¹³⁸ approach as a means of maximising the expected net benefits of a research project. These relate to the stopping rules being based on hypothesis testing, powering on expected treatment effect rather than the minimal clinically important difference and ignoring the opportunity cost of making the wrong decision.

The authors point out that their approach focuses on frequentist hypothesis testing to determine the stopping rule at each stage of the trial. Given the obvious linkage with Bayesian statistics (updating prior beliefs about expected treatment effect), they state that a frequentist interpretation of the analysis may be sub-optimal. However they point out that their approach "mixes both Bayesian and frequentist thought processes and may be considered from either point of view". They state that their analysis can be adapted to consider either point of view: a frequentist interpretation would use the prior distribution of treatment effect to inform sample size and then make inferences on the observed data alone. A Bayesian interpretation would formally combine the prior and observed data into a posterior from which to make inferences.⁷

A criticism of the approach though is whether hypothesis testing is an appropriate means to determine a stopping rule in any case. As stated in Chapter 1, Claxton³ is critical of hypothesis testing on the grounds that it does not lead to maximisation of expected net benefits. Nevertheless, Thach & Fisher's¹³⁸ approach may have greater acceptability to regulatory agencies as it sits within a more familiar approach based on hypothesis testing.

Cohen et al. argue that research output has "multidimensional... output[s which] make it impossible to specify a single outcome measure and hence derive and ICER [for a particular research design compared with another]". If one takes the extra-welfarist point of view and argues that the sole purpose of research is to reduce decision uncertainty and increase expected health gain associated

with the decision, then Cohen et al.'s argument does not hold: the expected value of a research project is unidimensional in terms of the increase in expected health gain it will yield. Thus it would be possible to calculate an incremental cost-effectiveness ratio for one trial design compared with another.

In summary, only one study (Shavit et al.²⁰) appears to have employed the principles of value of information analysis to inform qualitative choices between study designs. Their approach appears reasonable, relying on prior beliefs about sources of relative bias between, for example observational and randomised trials. However their work tackles a more macro question, addressing overall trial type. The starting point of my thesis is where a randomised controlled trial has already been decided upon, and more micro-design questions are of concern, namely how best to collect a particular data item.

2.4.2. The Cost of collecting resource use data alongside clinical trials and/or the cost of clinical trials themselves.

Despite much discussion of the expense of resource use data collection in the health economics literature (e.g. ^{196 197}), my review did not identify a single study attempting to measure this aspect of data collection at all. Instead, studies focused on overall costs of trials and strategies to contain such costs. Careful examination of predicted costs and efficient trial management appears to have had quite a dramatic effect on the expected costs of some trials, with 50% or more reductions appearing feasible, particularly with very large 'megatrials' such as the women's health trial.¹⁶⁴ However, a second attempt to use the same methods was less successful: faced with a choice between adding another trial centre and extending the follow up of existing recruits to increase power at minimum cost, extension of follow-up appeared to be more efficient. The investigators ultimately decided on the former course of action, citing the need to hedge against risk as justification.¹⁶⁵ This provides an example of the limitations of technical solutions: they cannot incorporate all elements of interest, and so should be used to guide to assist, rather than make, decisions.

Vickers & Scardino¹⁷⁰ present an interesting method based on the idea of incorporating experimentation into routine clinical practice. This is an appealing idea (subject to ethical boundaries and informed consent). However, the authors appear to focus on the treatment costs of a trial, rather than the associated overheads: the resources required to develop protocols, design data collection instruments, analyse the data and write up reports are substantial and so, depending on the nature of the trial, the scope for savings will vary.

Finally, and of most relevance to this thesis, Allison et al. in their study of approaches to maximise statistical power for minimum cost, conclude that 'better quality data', for example use of better

trained observers, may increase the statistical power of a study (presumably by reducing variability due to measurement error), and '[optimise].. power and costs'.¹⁶⁹ Unfortunately they do not elaborate on this, but in essence this is the question posed in this thesis, that is, what is the added value of 'better quality' data?

2.4.3. Comparisons of alternative data collection techniques

The primary purpose of this review was to explore the evidence of any systematic differences in resource use estimates / costs measured using alternative data collection techniques / processes. I uncovered a large number of studies comparing different data collection techniques. Whilst some informally referred to cost or cost-effectiveness none attempted to formally value cost or the additional gain (or loss) associated with using one method compared with the other. For example Merlo et al.¹⁷⁸ referred to the expense associated with a daily diary to record hormone use compared with an open-ended questionnaire to be completed once only. They did not elaborate on what these costs would be and who bore them. In this case a daily diary places substantial burden on the participants themselves (possibly leading to incomplete data), as well as requiring considerable analytic and data entry effort, thus adding to research costs. As I state above, knowledge of the relationship between the data gathered using each method could be used to value the relative gain from one method over the other (this idea is developed in Chapter 4).

The systematic reviews of comparisons of data collection techniques reveal an overall picture of the relative reliability of self-reported resource use compared with official records, concluding that surveys based on self-report are broadly valid but may overestimate the proportion of the population screened for cancers.^{172 173} If an analyst must rely on self-reported health service use, then the design of the questionnaire can also influence the responses received. Gama et al.'s¹⁷⁵ systematic review found that beginning with a prompted list of drugs followed by an open ended question leads to a higher reported prevalence of drug use than if questions are asked the other way around, and that within the prompt list, the order in which drugs are presented to respondents affects the responses received. However, it was not possible to judge the overall accuracy of responses from the Gama review as the specificity of different approaches was not reported: a higher prevalence may simply be a false positive. Of interest is Gama's observation that excessively lengthy surveys can lead to 'respondent fatigue'. This will affect the completion rate of the survey, thus whilst a more detailed survey may yield 'better quality' data, the response rate will fall. Such a survey is also more costly to administer and analyse.

Amongst studies conducted in the UK, using self-reported questionnaires compared with medical records does not appear to lead to a substantive difference in incremental cost, thus the resulting

incremental cost-effectiveness ratio will not be biased. However the total cost per patient may be affected. This is a consistent theme in the literature: the divergence between self-report and medical records increases with increasing disaggregation of cost.⁵⁵ Extending this, studies comparing top-down costing with bottom-up approaches found greater agreement with top-down approaches,¹⁸¹ although top-down approaches also tend to lead to higher overall per patient cost estimates.¹⁸² Unsurprisingly the reliability of self-report also decreases with increasing recall period.¹⁸⁰

In attempting to replicate Knapp and Beecham's¹⁸⁵ reduced list costings approach, Whynes and Walker¹⁸⁶ found it generated far less reliable estimates of cost than the formers' original analysis. However, Whynes and Walker do not consider whether the reduced cost approach is 'good enough' for purpose. As observed by Knapp and Beecham,¹⁸⁵ if the objective is to obtain accurate resource use costs for a particular cohort, then the reduced approach may be inappropriate. However, if the objective is to inform a decision as to whether to adopt an intervention, the costs form one input into incremental net benefit. There are then two issues to consider. Firstly, as stated above, whilst absolute costs may be invalid, this is not necessarily true for incremental cost. Secondly, even if the estimate of incremental cost is biased, the issue is whether it is sufficient enough to change the probability of making the 'wrong' decision. A rational approach would be to quantify the expected change in expected loss as a result of using the inferior or superior process and to compare this with the incremental cost of collecting one or the other data sets. This idea forms the basis of my approach in this thesis, and is developed in Chapters 3 and 4.

2.5. Conclusion

The purpose of this review was to establish the extent to which (1) economics-based techniques have been used to inform the design of clinical trials (other than sample size), (2) the costs of collecting (resource use) data within clinical trials have been analysed and (3) different approaches to collecting the same data had been compared.

My review found very few examples of the use of economics-based techniques to inform the design of clinical trials: whilst several studies claimed to consider both costs and 'outcomes' (i.e. validity/bias of the trial result) of different aspects of trial design, only one study reported an approach attempting to explicitly value the costs and consequences of alternative designs, this being a comparison between a prospective and retrospective study.²⁰

Despite frequent reference in the health economics literature to the cost and burden of resource data collection alongside clinical trials, I was unable to identify any studies exploring or estimating the cost of collecting such data. However, there were a number of studies comparing the treatment

costs of patients enrolled in trials compared with those receiving routine care, the conclusion being that treatment costs were not necessarily higher for such patients.^{153-161,xii} Other studies focused on organisational and management techniques to reduce the total cost of very large trials, in some cases with considerable success.¹⁶⁴

Finally, there is a substantial literature comparing alternative approaches to measuring the same data. In general, different approaches lead to similar but not identical cost estimates, although the generalisability of such comparisons is uncertain. For example, use of a 'reduced list costings' approach was able to predict the majority of cost in one case,¹⁸⁵ but not another.¹⁸⁶ However, as stated above, this does not mean that the 'inferior' process should be abandoned: if it is substantially less expensive to collect, it may still be preferable if the reduction in research cost outweighs the reduction in 'outcome' or 'quality' of the resulting parameter estimates (however defined). The remainder of my thesis provides an approach to measuring and valuing the added quality of one data process compared with another, and thus when taking into account the relative cost of each, whether one yields a higher net benefit compared with the other.

^{xii} The driver for these analyses being a reluctance for insurance companies in the USA to meet the treatment costs of patients enrolled in trials due to a perception of higher cost.

3. A critique of Value of Information Analysis

3.1. Introduction

In this section, I explore the methods and limitations of value of information analysis as currently applied in health care research. I consider three specific areas: a criticism of value of information analysis in general, followed by separate consideration of issues specific to the analytic and numeric solutions. Finally I consider a particular limitation in more detail, that is, the unknown correlation structure between evidence combined from different sources.

As outlined in Chapter 1, the two main approaches to calculation are the analytic and numeric solutions. The analytic approach requires knowledge of the mean and standard error of incremental net benefit. This approach has been used to calculate value of information statistics using data from randomised controlled trials, where the source data for calculating incremental net benefit is incremental mean cost and outcome, the valuation of a unit of outcome, the standard deviation and error of each increment and the covariance between the two.⁸² The numeric (simulation) solution is more commonly associated with decision models, where Monte Carlo simulation is used to build up an empirical distribution of incremental net benefit (e.g.^{84 85}).

Limitations of the analytic solution are 1) the requirement for incremental net benefit to be normally distributed and 2) that if based on a single source of evidence it is not necessarily consistent with the principles of evidence based medicine,⁹¹ potentially excluding other relevant evidence. The numeric solution (typically based on a decision model) overcomes these limitations, but with the consequence that the correlation structure between model input parameters is usually unknown, and whilst prior beliefs about correlations can be incorporated, frequently they are ignored. The structure of the model also leads to additional uncertainty (that is, different model structures may be equally valid yet yield different results). The final section of this chapter explores the impact of ignoring correlation on the value of information statistics with four empirical examples. I find that ignoring correlation between parameters can have consequences for the optimal sample size, particularly where parameters are very strongly positively correlated or where the

variable cost of sampling is low relative to the EVSI. Finally I consider possible solutions to take account of correlation in a decision model.

3.2. Criticism of methods

In this section, I first discuss issues relating to value of information analysis as a whole before discussing aspects specific to the analytic versus numeric approaches. The issues relevant to Vol itself are defining the beneficial population, the assumption of independence between the adoption and research decisions, the impact of multiple jurisdictions and the analytic perspective.

Issues specific to the analytic approach are the required assumption of normally distributed parameters and the use of 'all relevant information'. Those relating only to the numeric solution are structural uncertainty and the characterisation of parameter uncertainty, and correlation between input parameters.

3.2.1. Issues relating to Vol in general

Defining the relevant patient population

The value of additional research into a decision question is a function of not only the current, but also future patient population estimated over an 'appropriate' time horizon. This is usually taken as the sum of the (discounted) incidence of the disease. Depending on the nature of the disease and treatments, it may also be appropriate to include the current stock of prevalent disease to this. Whilst it may be possible to estimate the future incidence and prevalence of the disease with a reasonable degree of certainty, it is far from clear exactly what an 'appropriate' time horizon is, and the value of information is extremely sensitive to the time horizon selected.¹⁹⁸

One approach would be to adopt an infinite time horizon. This will yield a finite value of information for any positive discount rate, and so may provide an upper limit of the value of information (although this ceases to be true once uncertainties associated with technological change and future prices are incorporated¹⁹⁸). However, an infinite time horizon is not particularly plausible. An alternative would be that the time horizon should reflect the 'effective lifetime of the technology'. However, this is itself an unknown parameter for which further information could be sought.⁸³ The most appropriate horizon would be one that equates to the time over which the decision question remains relevant. In other words, the time to the next major development in the disease area rendering the

current decision question obsolete. 'Horizon scanning' of new technologies in early stage development may be a means to estimate this.

A review of applied studies employing Vol techniques found studies tended to use a horizon of either 10 or 20 years, with no clear justification in either case.¹⁹⁸ Philips et al.¹⁹⁸ argued that adopting a single cut-off is essentially an approximation representing a more complex process of changes to the decision problem through time, where changes in relative prices, information and development of new technologies each affect the value of information attributable to different model parameters to differing degrees.¹⁹⁸ They do not, however, recommend simultaneous modelling of all these aspects for pragmatic reasons. Instead they recommend consideration of the information needs of decision makers: in essence the analysis has to be sophisticated enough to incorporate all relevant influences, but simple enough to be delivered within a reasonable timeframe and comprehensible to those decision makers. The decision to pursue additional research will be made, with or without formal analysis and the purpose of decision analysis is to improve the quality of the decision (increase the probability of a decision being the 'correct' one), and not to capture in minute detail every nuance of the decision problem. The overriding principle is the same for all decision modelling: that the analysis has to be fit for purpose.^{xiii}

Those patients who participate in a study will not normally be able to benefit from the information obtained from that study. Thus when calculating population EVSI, the population is defined as $(N-2n)$ where n is the number of patients in each arm of the study. However, this is not necessarily the case. For example, study participants may benefit from the information from that study where the disease is characterised by well-defined periods of relapse and remittance, and the treatment provides symptomatic relief of relapses. Therefore when multiplying the per patient EVSI to the population level, the population should be adjusted for this.³ Furthermore, there will be a delay between any decision to carry out research and the results being acted upon, which carries an opportunity cost borne by the entire patient population, requiring further adjustment to the population EVSI.¹⁹⁹ Similarly, patients enrolled in the trial randomised to the 'inferior' arm incur an opportunity cost equal to the foregone incremental net benefit per patient. The impact of these issues on the overall EVSI depends on the size of the patient population relative to those enrolled in the trial. For a common disease such as asthma or diabetes, trial enrollees will comprise a very small proportion of the total population. However, for rarer diseases, accounting for

^{xiii} A consequence of this is that making a model fit for purpose may make it less transferable to another purpose.

the opportunity cost of trial enrollees may affect the optimal sample size calculations substantially.

Independence of the adoption and research decisions

The iterative approach to evidence-gathering and decision making requires the important conceptual split between the decision to adopt a new technology and the decision to pursue further research (to inform a future revision of the adoption decision).

Whilst separate, the adoption and research decisions are not truly independent of one another as (1) if the adoption decision is delayed whilst new research is underway, there will be an opportunity cost to those who could have benefitted if the technology does indeed have a positive incremental net benefit (and vice versa),¹⁴⁴ and (2) if there are considerable costs associated with reversing a decision.²⁰⁰

For example, suppose current information suggests a new technology has a positive mean incremental net benefit, but with sufficient uncertainty to warrant further research. The recommendation from this would be to adopt the technology for the present, and gather new evidence. If, after gathering the new evidence and incorporating it with the prior, expected INB becomes negative, the decision should be reversed at that point. But if the cost of reversing that decision is greater than the benefit from reversal (e.g. retraining of staff, construction of new facilities), the expected value of “adopt and research” was actually zero (it cannot affect the adoption decision): the optimal decision would have been “delay and research”, even when current evidence suggests a mean positive incremental net benefit.²⁰⁰ It is important therefore to include the cost of reversal in the analysis.¹⁴⁴

Multiple jurisdictions

Information is a public good: once in the public domain it is non-rival and non-excludable meaning consumption by one individual or group neither diminishes consumption by another, nor can that individual group prevent the other from consuming it. In the absence of other benefits from research (e.g. employment maintenance and prestige), this would lead to free riding as there is no reason for one jurisdiction (e.g. a state research funder) to pay for research when another can do so. Therefore whilst the EVSI may suggest a particular study should be carried out, it may be strategically optimal to wait for another jurisdiction to undertake the research instead, depending on the transferability/ generalisability of the results to the local jurisdiction. This could lead to a sub-optimal (Nash) equilibrium with a failure to carry out research that would be beneficial to both jurisdictions. Alternatively,

there may be a global optimal allocation of patients across jurisdictions in a particular trial, dependent on the relative costs and benefits in each location.²⁰¹

Analytic perspective

As discussed in section 1.3.1, a critical issue when defining the study question for an economic evaluation is the analytic perspective. This is just as true for value of information analysis: the analytic perspective determines whose costs should be included. In 1.3.1, I stated that the preferred perspective is all of society (usually defined as a particular legal jurisdiction, i.e. country). It is important that value of information analyses are consistent in the perspective they adopt. For example, where a Vol analysis is based on an economic evaluation with a societal perspective relevant to the UK, the relevant population by which the per patient EVPI and EVSI should be calculated would be based on the incidence and prevalence across the whole of the UK.

Where (as is often the case and indeed, mandated, for the reference case in NICE appraisals⁷²) the perspective is limited to just the public sector, the Vol will ignore uncertainty in broader societal costs such as out of pocket costs, carer time and lost productivity to the economy (although, by defining the perspective as public sector only, these costs are not relevant to the decision in any case).

Related to this issue is the use of centre specific versus national average unit costs. As I stated in Section 1.3.6, the choice between using nationally representative unit costs and centre specific ones is dependent on the study question: if the purpose is to inform policy across the UK, then it may be reasonable to apply national average unit costs for inputs such as GP or hospital attendances (e.g. ^{127 202}). However, this risks underestimating the true variability in unit costs from centre to centre: collecting centre specific data would allow estimation of not only the mean costs nationally, but also the spread around those means. This though would be extremely resource intensive, but could well be unnecessary as published datasets such as the national schedule of reference costs¹²⁷ include measures of uncertainty around the national average which should be incorporated into sensitivity analyses and thus value of information analyses.

3.2.2. Issues specific to the analytic solution

In this section I consider the requirement of the analytic solution for normally distributed parameters, and the consideration of 'all relevant evidence'.

Assumption of normality

The analytic solution described in Chapter 1 requires distributions of incremental net benefit (and indeed of any component parameters) to be normally distributed. The central limit theorem states that the sampling distribution of the mean of a parameter will be approximately normal, irrespective of the distribution of the parameter itself.²⁰³ Eckermann et al.²⁰⁴ thus argue the analytic solution provides a quick and simple approach to calculating value of information statistics avoiding the complexities of the numeric simulation method.^{xliii} However such a justification risks conflating frequentist and Bayesian statistical techniques: Within a frequentist framework, the 'true value' of a parameter (such as mean incremental net benefit) is unknown and a fixed constant, but can be estimated from samples drawn from the data (in this case conducting a clinical trial reporting mean incremental net benefit). As the number of observations increases, the sampling distribution of mean incremental net benefit tends towards normal. The implication is that large sample procedures can be used to make inferences about the population mean even when the distribution of the population mean is unknown (i.e. hypothesis testing).

In a Bayesian framework, the distribution around the mean of the parameter represents belief about likely values, therefore there is no reason why this should (or indeed should not) be normal.

Coyle and Oakley²⁰⁵ in their review of a number of approaches to calculating the expected value of perfect parameter information (EVPI) include one based on the unit normal linear loss integral (UNLLI) similar to that described in Chapter 1. The difference however is that whilst they require the target parameter(s) for the EVPI analysis to be normally distributed and linear in incremental net benefit, no such requirement is placed on the remaining parameters.

Use of all relevant evidence

The analytic solution does not naturally incorporate all relevant information. The source of the prior distribution of INB is not specified by Willan & Briggs,⁸⁶ but it is natural to assume that it may be from a pilot trial with concurrent resource data collection. Alternatively it may be from a systematic review (or rather, meta-analysis). However there are difficulties with this. Systematic reviews and meta-analyses such as those produced by the Cochrane

^{xliii} Eckermann and colleagues invoke Occam's razor to justify their preference for the analytic solution, although arguably the numeric solution is not more complicated but merely time consuming: performing sufficient simulations to adequately characterise uncertainty may in practice require many months of computer processing time, and thus may only be practical with the use of high performance cluster processors.

Collaboration²⁰⁶ focus on randomised controlled trials to estimate an incremental treatment effect. To my knowledge, there are no meta-analyses of trials reporting incremental net benefit (or even of costs alone). Therefore it is unclear how the cost data should be combined with trial outcomes data to form a measure of incremental net benefit within an analytic solution. Doing so enters the realm of decision modelling.

Decision models start from the principles of evidence based medicine by allowing incorporation of 'all relevant evidence'. This would include not only randomised controlled trials, but other data sources too such as database analyses, reference cost indices and even expert opinion where necessary. However, the use of data from such disparate sources means that the correlation structure of the input parameters is unknown (see section 3.3).

3.2.3. Issues specific to the numeric solution

In this section I discuss the issues of computational burden, the characterisation of structural and parameter uncertainty, and the lack of correlation between input parameters.

Computational burden

The numeric solution to Vol can be somewhat “computationally expensive”. Whilst improvements in the power of computers may partially solve the problem, there is a tendency for programmers to develop more sophisticated models as a result, thus counteracting the increase in modern computer processor speed. Alternative shortcuts have been proposed including linear approximations of non-linear models, meta-models and search algorithms, which may provide an appropriate compromise between computational speed and loss of accuracy.^{11 145 207-209} However computational burden remains a challenge to simulation methods to estimate Vol due to it being costly in both time and resources.

Structural uncertainty and characterisation of parameter uncertainty

The validity of the value of information approach to research prioritisation rests on two critical assumptions. Firstly, that the structure of any decision analytic model on which it is based is correct, and secondly that the uncertainty around each of the parameter inputs is appropriately characterised.

The first point is a question of ‘structural uncertainty’, a type of uncertainty which conventional sensitivity analyses do not commonly address, other than through a range of scenario analyses from which the decision maker is invited to choose the most plausible.²¹⁰ An alternative is to employ a model averaging approach based either on model fit to the data, or by adding parameters to the model to represent the choice between the alternative

scenario analyses. Each iteration of the probabilistic sensitivity analysis then selects one of the scenarios based on some distribution of the likelihood of each.²¹⁰

The second point relates to the use and combining of evidence, i.e. systematic review and meta-analysis. Economic evaluations should make use of “all appropriate evidence”.²¹¹ This is to ensure consistency with the principles of evidence-based medicine, defined as:

*“...the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients”.*⁹¹

Although in this case the concern is with the population level, i.e. policy decision making, rather than an individual patient, the statement is equally valid for informing economic evaluations upon which to base policy decisions.

The question though is what constitutes appropriate evidence. The Cochrane Handbook for Systematic Reviews of Interventions states that the primary difference between a systematic as opposed to narrative review is the “pre-specification of eligibility criteria for including and excluding studies in the review”,²¹² defined as a statement of the clinical question and specification of the types of study that will be included. The Handbook suggests a review should seek ‘all rigorous studies (e.g. randomised trials) of a particular comparison of interventions’. For estimating measures of the effect of health care interventions, they suggest a focus on randomised trials as the study design least prone to bias.

The result is that systematic reviews, such as those produced by The Cochrane Collaboration, focus on establishing the best estimate of one particular parameter, or group of related parameters (i.e. beneficial and adverse effects). They provide a valuable input into decision analytic models, but are highly unlikely to provide data for every parameter included in a model.²¹³ For example, a systematic review of early versus delayed laparoscopic cholecystectomy for acute cholecystitis reported statistics relating to risk of peri- and post-surgical complications, conversion to open procedure and mortality.²¹⁴ A subsequent economic evaluation drew heavily on this review, but in addition required data on the probability of a patient becoming symptomatic, prognosis of pancreatitis, as well as resource use, unit costs and utilities.²¹⁵

To be classed as using *all appropriate evidence* these model inputs too should be based on systematic reviews.¹ However these are simply not available for every parameter (especially resource use estimates), and as a last resort, model inputs are sometimes based simply on

author estimates.^{xiv} It is therefore unclear whether the probability distribution assigned around such point estimates adequately reflects parameter uncertainty.

Incorrect specification of parameter uncertainty will lead to incorrect estimation of the value of further research. The only practical solution is to ensure that care is taken to fully characterise parameter uncertainty. For example replacement of 'author estimates' with a formal elicitation technique,²¹⁸ or use of Bayesian multi-parameter evidence synthesis, where data on a number of related parameters from a number of sources are synthesised together.²¹⁹ Ultimately, a 'comprehensive decision analytic modelling' technique is desirable where systematic review, parameter estimation, sensitivity analysis and economic evaluation are carried out within one single modelling framework.²²⁰

Correlation

Due to the nature of decision models, drawing on diverse data from numerous sources, the analyst does not have information on the correlation structure between the parameters. A common approach, as observed by Briggs et al.⁷⁴ is to simply ignore correlation and assume independent inputs, but this may not be realistic. There are two sources of correlation: that between parameters observed in the same individual and that induced between parameters by the structure of a model. I refer to ignorance of the former in this case. For example, patients with poorer outcomes are likely to consume more health care resources. Patients who consume more of one type of resource (e.g. hospital admission length of stay) may also consume more of another (e.g. long term care). In a linear model, ignoring correlation should not bias the point estimate of incremental net benefit (although it may do in a non-linear structure), but could either over or underestimate the variance. As the value of information statistics are a function of variance of INB, the value of additional research too could be over or underestimated. I provide a fuller discussion of this in Section 3.3.

3.2.4. Summary

In this section I have outlined a number of issues relating to value of information in general, and issues specific to the analytic and numeric solutions. A key issue with the analytic solution is that it is inconsistent with the axiom that decisions should make use of 'all relevant data'. The exact definition of this is a moot point, but all definitions are likely to include the desire to incorporate data from more than one source, if only for pragmatic reasons.⁹² Published examples using the analytic solution⁸² have drawn data from one

^{xiv} Expert opinion is considered the lowest grade of evidence in a hierarchy topped by well conducted systematic reviews, followed by RCTs, observational studies (in the order of cohort studies, case control studies and then cross-sectional surveys) and finally case reports^{216 217}

source only, where information on all input parameters (i.e. cost and outcomes) are simultaneously observed in the same sample. The correlation structure between those parameters is therefore also observed. However, once more than one data source is employed the data need to be combined in some decision model, and the correlation structure is unobserved. The next section explores the implications of this for value of information in detail.

3.3. Ignoring correlation between model input parameters

3.3.1. Introduction

Stochastic decision models (i.e. Monte Carlo simulations of Markov or other model structures) are recommended to estimate the cost-effectiveness of interventions, and to appropriately characterise uncertainty around the mean^{72 92}

In such models, input parameters (for example response rates to treatments, resource quantities and utilities) are commonly assumed to be independent as “...analysts usually have no data on the covariance structure and so choose not to model covariance...”.⁷⁴ This is due to the nature of decision analytic modelling: a single trial is unlikely to provide all necessary data for decision making. It is therefore necessary to draw on a variety of summary data from numerous sources.^{88 92} Even for those modelling studies drawing on one primary study for effectiveness data, resource use data are frequently extracted from other sources, and any correlations between resource use and other model inputs (e.g. response rate) are therefore unknown. Excluding correlations implicitly assumes the correlation between parameters is precisely zero.

In the remainder of this section, I first define correlation (as the Pearson linear correlation coefficient), before exploring methods by which correlation between input parameters may be incorporated in decision models. I then examine some decision models reported in the literature and consider how they have incorporated correlations or dependencies between input parameters. In the following Section (3.3.2) I explain the methods to compare the impact of ignoring correlation before presenting results and discussion in Sections 3.3.3 and 3.3.4 respectively.

Definition of correlation

Correlation is a measure of the strength of the relationship between two parameters. Whilst regression analysis attempts to estimate the relationship between two (or more) parameters, correlation analysis attempts to estimate the strength of that relationship.²²¹ In other words it is a measure of how often high values of one are associated with high or low values of the other.

The most common statistical measure of correlation is Pearson's (linear) correlation coefficient. This is the covariance divided by the product of the standard deviations, and is usually denoted ρ_{xy} (Equation [3-1]). What is true for individual observations is also true for parameter means, where the correlation coefficient between the mean of two parameters is defined as the covariance between the means divided product of the standard errors (where

the covariance between the means and standard errors are the sample covariance and standard deviations divided by n and \sqrt{n} respectively). In both cases, the division by n or \sqrt{n} cancels out and the equation reduces to equation 3-1.

$$\rho_{xy} = \frac{cov_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad [3-1]$$

The advantage of the correlation coefficient over the covariance is that it is normalised to between -1 and +1, whilst covariance is a function of the units of measurement. Independent parameters have a correlation coefficient of zero. However, it is not true to say that zero correlation implies independence as the correlation coefficient only measures linear correlation (the variables may have some other functional dependency). The exception is where both variables follow a bivariate normal distribution. That is, for any value of y, x is normally distributed, and for any given value of x, y is normally distributed.

I distinguish here two sources of correlation. Firstly, the relationship between two parameters can be estimated where both are observed in the same patient. For example, within a clinical trial it can be observed that patients who have a better health status may have lower health care costs. Alternatively at baseline, it may be observed that heavier patients also tend to have higher blood pressure. Secondly, there are correlations induced by the structure of a decision model. Interestingly the model structure can induce both observable and unobservable correlations:

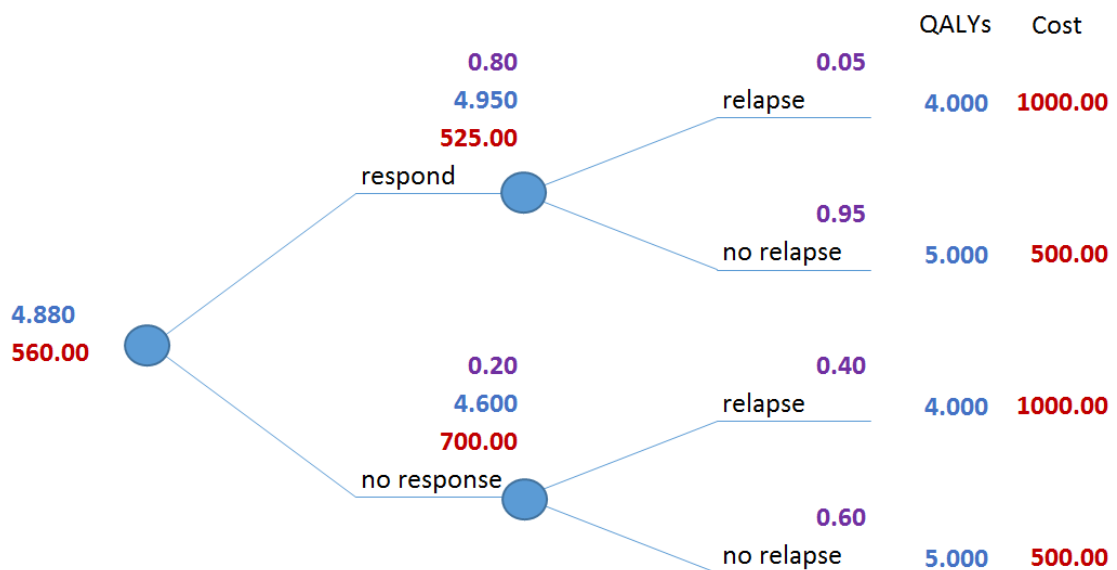
Potentially observable correlations can be induced within a model as shown in Figure 3-1 below. This shows part of a decision tree where a patient given a treatment has an 80% probability of response (e.g. as measured by some biochemical marker). If they respond, they have a 5% probability of relapse, compared with a 40% probability of relapse if they do not respond. QALYs and costs are assigned at the terminal nodes as shown. Rolling back the tree, the expected cost and QALYs gained for a patient receiving this treatment are £560 and 4.88 respectively.

A probabilistic sensitivity analysis can be conducted by assigning probability distributions to each parameter, in this case the probability of response, probability of relapse given response, probability of relapse given no response, cost and QALYs of relapse and non-relapse respectively. Even when all input parameters are assigned independent probability distributions, a correlation between expected cost and QALYs is induced due to the structure of the model. For example, as the probability of relapse rises, expected cost will rise and

expected QALYs will fall simply because the model is structured such that those relapsing incur higher cost and lower QALYs.

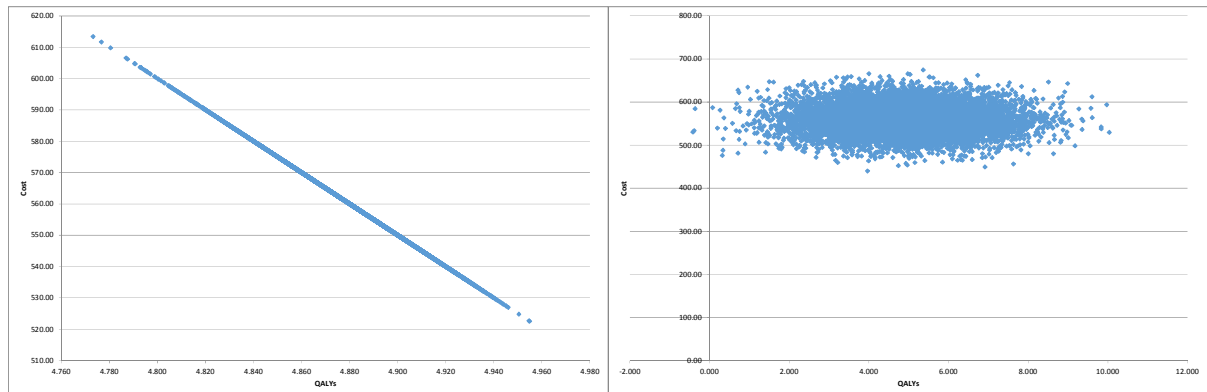
Note that the degree to which this is expressed in the correlation between final cost and QALYs accrued per patient depends on the proportion of overall cost and QALYs the model section represents, as well as uncertainty in mean cost and QALY decrement: if there is very little uncertainty in cost or the utility decrement, the observed correlation will be stronger than otherwise. For example, assigning probability distributions to probabilities and assuming QALYs and Cost are known with certainty means that as the probability of relapse or response is varied, expected cost will always rise as expected QALYs fall and vice versa, thus inducing perfect negative correlation between the two (-1, Figure 3-2a). Assigning an arbitrarily 'large' standard error around QALYs and cost 'dilutes' this effect, yielding a correlation coefficient close to zero (-0.017 in this example, Figure 3-2b).

Figure 3-1: Example decision tree (part)



Probabilities are illustrated in purple, QALYs in blue and costs in red.

Figure 3-2: Scatterplot of sampled costs and QALYs with (a) fixed costs and QALYs, probabilities allowed to vary, and (b) with uncertainty in cost, QALYs and probabilities.



Unobservable correlations are those *between* patients. For example, if a particular patient does better than expected on treatment A, it may be reasonable to assume they will also do better than expected on treatment B. This is true of means too: if the mean response to treatment A is higher than expected, then it may be reasonable to believe that the mean response to treatment B will also be higher than expected. There is no way to observe this as it is not possible to give the same (group of) patient(s) both treatments separately at the same time. However a decision model can impose a correlation structure between the means. For example, Wilson et al.¹⁵ estimated the cost effectiveness of early versus delayed laparoscopic cholecystectomy (LC). One of the inputs was the conversion rate to an open procedure with early or delayed LC. The contingency table is shown in Table 3-1, using data extracted from a Cochrane systematic review.^{214,xlv}

Table 3-1: 2x2 contingency table for conversion to open procedure

	Convert to open?	
	Y	N
Early	46	178
Delayed	52	166

These data were modelled as independent beta distributions. However, it may have been more appropriate to model the probability of conversion with Early LC as a function of the baseline probability and the odds ratio of conversion.^{xlvi} Figure 3-3 shows the results of 10,000 draws from the respective independent beta distributions, yielding as expected, a correlation coefficient very close to zero. Figure 3-4 shows 10,000 draws with the imposed

^{xlv} Note the data include a non-informative prior. See Appendix to Wilson et al.¹⁵ for details.

^{xlvi} Where $P(\text{conversion with early LC}) = \exp(\text{logit}(P(\text{conversion with delayed})) + \text{Log Odds Ratio})$

structural correlation, yielding a correlation coefficient between the two parameters of approximately 0.57.^{xlvii} In particular note the range of conversion rates with ‘early LC’ is much higher than in the uncorrelated case. Incorporating these ‘structural correlations’ often leads to induced correlations between mean incremental cost and QALYs gained.

Figure 3-3: Scatterplot of 10,000 uncorrelated draws

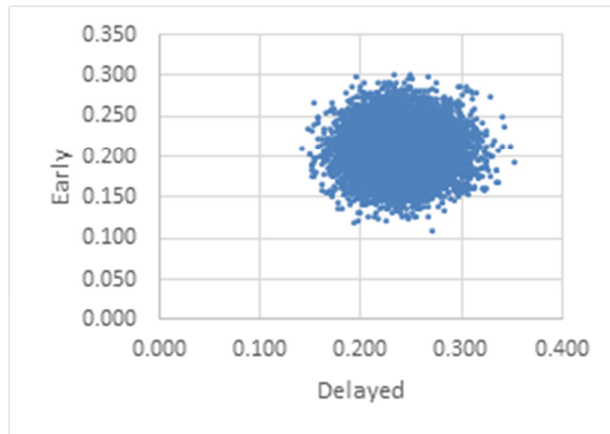
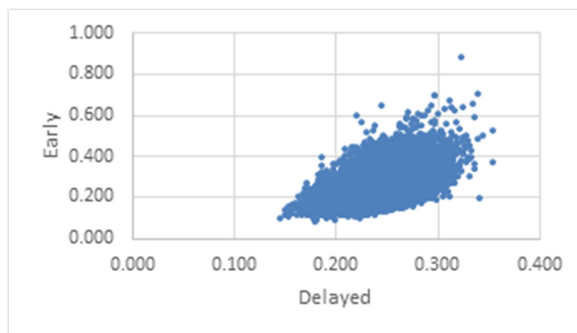
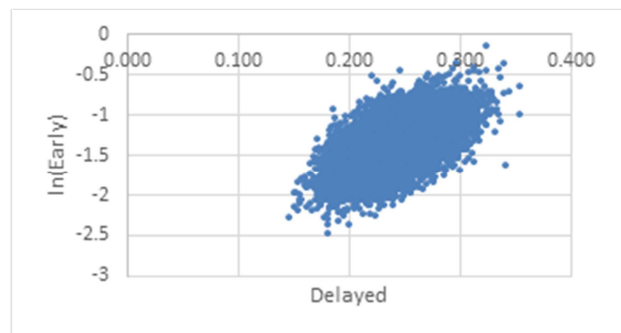


Figure 3-4: Scatterplot of 10,000 correlated draws (a) natural units and (b) logarithmic

(a)



(b)



Methods to incorporate correlation in decision models.

There are a number of approaches which explicitly take correlation into account within decision modelling. These are the Cholesky decomposition, Gibbs sampling (an

^{xlvii} This is the correlation coefficient between $P(\text{conversion with delayed})$ and $\log(P(\text{conversion with early}))$, to linearise the relationship.

implementation of Markov Chain Monte Carlo (MCMC) methods), and Copula equations. Each assumes that the complete multivariate distribution is known. Such a distribution is typically extremely awkward to solve analytically, and so the three methods are means of making correlated draws from the joint distribution. Repeated sets of draws can then be used as inputs into a probabilistic decision model. These are discussed in turn below.

The Cholesky decomposition

Briggs et al.⁷⁴ describe a method for making correlated draws from a set of multivariate normal parameters known as the Cholesky decomposition. This requires some estimate of the variance-covariance matrix (either from data or based on prior belief) and all parameters are (assumed to be) normally distributed, such as is generated from a standard OLS regression analysis. Where parameters are non-normal, appropriate transformations can be made such that they more closely approximate the normal. A number of models make use of this technique for at least some model inputs where data on correlations are available (e.g. ²²²⁻²²⁴).

The method is as follows:

Define a matrix, V, as the variance-covariance matrix. The Cholesky decomposition of V is T, such that T is a lower triangular matrix which, when multiplied by its transpose, TT gives V. A vector of correlated draws from the parameters (call this x) can be generated as per equation [3-2], where y is the vector of parameter means and z is a vector of independent standard normal variables. An example with two parameters, x₁ and x₂, is shown in Box 3-3.

$$x = y + Tz \quad [3-2]$$

Box 3-3: Bivariate example of Cholesky decomposition

$$\text{Let } T = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \therefore T^T = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$$

$$TT^T = V:$$

$$\begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} v_{x_1} & Cov_{x_1, x_2} \\ Cov_{x_1, x_2} & v_{x_2} \end{pmatrix}$$

$$\begin{pmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{pmatrix} = \begin{pmatrix} v_{x_1} & Cov_{x_1, x_2} \\ Cov_{x_1, x_2} & v_{x_2} \end{pmatrix}$$

$$\begin{pmatrix} a & 0 \\ b & c \end{pmatrix} = \begin{pmatrix} se_{x_1} & 0 \\ \rho_{x_1, x_2} se_{x_2} & \sqrt{1 - \rho_{x_1, x_2}^2} se_{x_2} \end{pmatrix}$$

Using the Cholesky equation, the correlated parameters are defined as :

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

Multiplying out the matrices yields:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mu_1 + az_1 \\ \mu_2 + bz_1 + cz_2 \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mu_1 + se_{x_1}z_1 \\ \mu_2 + \rho_{x_1,x_2}se_{x_2}z_1 + \sqrt{1 - \rho_{x_1,x_2}^2}se_{x_2}z_2 \end{pmatrix}$$

Markov Chain Monte Carlo & Gibbs Sampling

Gibbs sampling is based on the Metropolis-Hastings algorithm.²²⁵ The Metropolis algorithm is used to sample from complex distributions which would be awkward or impossible to integrate analytically, and can therefore be considered an improvement on the Cholesky decomposition as it relaxes the assumption of normally distributed parameters. Briefly, a starting value is chosen, θ_0 . A new candidate value is then picked at random from some (arbitrary) 'jumping' or 'proposal' distribution. This is denoted θ^* . The density of the target distribution is then evaluated at θ^* and θ_0 . The probability of accepting θ^* is then the minimum of the ratio of these densities and 1 (Equation [3-3]). If θ^* is accepted, then θ_t is set to equal θ^* , else θ_t is set to equal θ_{t-1} . This is repeated an arbitrarily large number of times. The values $\theta_0, \theta_1, \dots, \theta_n$ thus form a Markov chain as the probability of observing a particular value of θ_t is only dependent on θ_{t-1} . As the chain progresses, whilst initial values may be at extremes of the density, eventually, after a 'burn in' period of k iterations, the distribution of the θ s will be approximately equal to the target distribution, that is, the chain will converge on the distribution of θ .

Hastings generalised the method by allowing for non-symmetric jumping distributions (the Metropolis-Hastings algorithm; Equation [3-4]).

$$\alpha = \min \left[\frac{p(\theta^*)}{p(\theta_{t-1})}, 1 \right] \quad [3-3]$$

$$\alpha = \min \left[\frac{p(\theta^*)q(\theta^*, \theta_{t-1})}{p(\theta_{t-1})q(\theta_{t-1}, \theta^*)}, 1 \right] \quad [3-4]$$

Where:

α = probability of accepting θ^*
 $p(\theta)$ = target distribution evaluated at θ
 $q(x,y)$ = jumping distribution.

Note that where the jumping distribution is symmetric, the second term in Equation [3-4] cancels out (i.e. $q(\theta^*, \theta_{t-1}) = q(\theta_{t-1}, \theta^*)$, and the algorithm is the same as the Metropolis algorithm). For example, if a uniform or normal distribution is chosen as the jumping distribution, the Metropolis algorithm is appropriate. However for a chi-squared or gamma jumping distribution, the Metropolis-Hastings algorithm must be used.

Gibbs sampling is a special case of the Metropolis-Hastings algorithm where θ^* is always accepted. For a complex multivariate distribution, initial values are set for all the parameters. The sampler then draws a new value for the first parameter, conditional on the values of all the others. This process is repeated for all the other parameters until a new set of values is drawn for all parameters. This is termed one scan of the sampler. This is repeated a large number of times and after a sufficient burn-in period, the chain will converge on the target multivariate distribution.

Cooper and colleagues²²⁰ extend the MCMC approach to incorporate meta-analysis of trial data and economic evaluation into one process termed 'comprehensive decision analytical modelling', using software such as WinBUGS.²²⁶ For example, to estimate the cost-effectiveness of taxanes for second-line treatment of advanced breast cancer,^{220 227} a meta-analysis of trial data was combined with a Markov chain model. The advantage of this approach is that it "makes full allowance for any potential inter-relationships between model input parameters, as the joint distribution of costs and effectiveness, conditional on the model, prior beliefs and the data, is used to make inferences".²²⁰ However, this is only possible where trial data allow correlations to be observed.

For example, Table 3 of Cooper et al.²²⁰ reports marginal probabilities of a number of side effects. These could be entered into a decision model as stated as marginal probabilities, and sampled from independently in probabilistic sensitivity analysis. However, it is reasonable to suppose that a patient suffering from diarrhoea may also be more likely to suffer vomiting, i.e. there may be a positive correlation between the two. Data informing

risks of side effects is primarily from one trial (Chan et al. 1999).²²⁸ Therefore using the MCMC sampling algorithm will allow correlated draws from the data, which are then 'plugged' straight into the decision model.

Copula equations

Copula equations are functions that link multivariate distributions to the marginal distribution of each parameter. A detailed exploration of copulas is beyond the scope of this thesis, but briefly, Nelsen²²⁹ defines copulas as follows:

Consider the bivariate case, with two random variables X and Y . Their respective marginal distributions are shown in Equations [3 5] and [3 6], and the joint distribution in Equation [3 7].

$$F(x) = P(X \leq x) \quad [3-5]$$

$$G(y) = P(Y \leq y) \quad [3-6]$$

$$H(x, y) = P(X \leq x, Y \leq y) \quad [3-7]$$

For every value of x and y , $F(x)$, $G(y)$ and $H(x,y)$ are each defined on the interval $[0,1]$, so every value of (x,y) corresponds to a point $[F(x),G(y)]$ in a unit square, and also a value of $H(x,y)$. Therefore every point $[F(x),G(y)]$, which is in the range $[0,1],[0,1]$ corresponds to a point $H[x,y]$ in the range $[0,1]$. The copula is the equation that maps $[F(x),G(y)]$ to $H(x,y)$.

The advantage of the copula is that instead of modelling the links between complex marginal distributions, the problem is reduced to specifying the dependency between a series of uniform distributions bounded between $[0,1]$.

Most examples of their use are in the financial literature, for example, to price collateralised debt obligations,²³⁰ although their popularity may have waned somewhat since the banking crisis of 2008. I identified one use within a medical context, comparing use of a copula with simple Pearson's correlation coefficient to predict the change in aortic ejection fraction following surgery.²³¹ Using pre- and post-operative data on 20 patients, the authors estimated post-operative ejection fraction as a function of pre-operative using OLS and also using copulas. On both measures of concordance (Lin's concordance measure) and accuracy, the copula-based model performed better than the correlation-based prediction model. The authors concluded that copula-based models are an appropriate alternative to conventional correlation-based approaches, as the conventional approaches are not suited to parameters with skewed (i.e. non-symmetric) distributions.

Examples of economic evaluations incorporating structural correlations and/or dependencies between input parameters

Of recent economic evaluations published in the journal 'Pharmacoeconomics' employing decision analytic modelling and probabilistic sensitivity analysis, a number had access to individual patient level data from one (or more) primary source(s) and so incorporated the correlation structure between (some) variables, most usually the treatment effect. For example:

- Reed et al.²³² modelled the cost-effectiveness of imatinib versus interferon alpha (IFNa) in patients with newly diagnosed chronic phase chronic myeloid leukaemia, drawing on a meta-analysis of two RCTs. In the model, survival in the treatment (imatinib) arm was modelled as survival in control (IFNa) multiplied by a 'calibration constant' corresponding to a percentage reduction in the hazard ratio. From the report it is unclear from where resource use data were drawn.
- Ramsey and colleagues²³³ developed a Markov model drawing on data from a single trial of atorvastatin vs no statin in the primary prevention of cardiovascular (CV) events in type 2 diabetes (T2DM). Risk of a CV event with atorvastatin was expressed as a hazard ratio relative to baseline (no statin), using data observed in the CARDS trial. The hazard ratio was assigned a normal distribution, and the baseline risk was that observed from an epidemiological study of the United States T2DM population. Costs were estimated from a separate source.

However, other studies maintained independence between parameters:

- Bojke et al.¹⁹ estimated the cost effectiveness of pharmacotherapy versus surgery (laparoscopic fundoplication) for gastro-oesophageal reflux disease (GORD). Key variables (e.g. outcome of surgery, risk of complications, probability of stable maintenance on medical management) were estimated from a fixed-effects meta-analysis of the literature, and incorporated into the model as independent beta distributions. Resource use was estimated from a survey of five hospitals involved in a concurrent comparative trial (the REFLUX trial), and resulting cost inputs were modelled as independent gamma distributions.
- Teerawattananon and colleagues²³⁴ report a comparison of three treatments for cytomegalovirus retinitis (eye infection) in HIV/AIDS patients: systemic (either oral or intravenous) ganciclovir ("O/IV"), intravitreal injection ("IVI"), or intraocular implantation ("IMP"). The model is a decision tree populated with data from a

systematic review. All inputs are modelled independently. For example variables "U2" (unilateral infection, risk of complications with IVI) and "U9" (unilateral infection, risk of complications with IMP) are modelled as independent beta distributions.

The first point to note is that even in the first two cases, dependencies between variables are limited to treatment effect between intervention and control (response in one arm is modelled as a function of response in the other): other inputs are modelled as independent distributions. Secondly, a crucial difference between the Reed²³² & Ramsey²³³ papers and the Bojke¹⁹ and Teerawattananon²³⁴ papers is that the first two are modelling one drug versus another, and thus the model structure between the two arms is identical. For the latter two however, comparing surgical and medical interventions, the model structure is of necessity different between the arms.

Where model arms are of identical structure, treatment effect of one drug can be modelled as a function of the treatment effect on the other, imposing a structural correlation between parameters: $P_n=f(P_o)$, where P_n is response on new, P_o response on old. For example, a commonly used relationship is $P_n=P_o.RR$ where RR is relative risk. Where the structure between the arms is fundamentally different (e.g. drug vs surgery) it is less clear exactly where and how relationships between parameters should be incorporated.

These heterogeneous data are also a key drawback of decision models as the interactions between the input variables (that is, the variance-covariance matrix) are unknown and probably unknowable for two reasons: firstly, as data are from different samples, there is no correlation observed between the parameters, yet it is reasonable to believe that high values of one (e.g. cost) may be associated with, for example, low values of another (e.g. utility / health-related quality of life / QALYs). Secondly, where parameters are observed in the same sample, analysts may not have access to the individual patient data, and so rely on what is reported in the published manuscript, which occasionally fails to provide the relevant statistics.

For these reasons, most decision models assume independence between input parameters. This has implications for the value of information, as the Vol is a function of the variance of incremental net benefit (INB). Ignoring correlations between input parameters may over or underestimate the variance of INB, thus over or underestimating the Vol.

3.3.2. Method & Data

In order to focus on the issue of correlation and exclude other impacts on value of information, I propose a very simple decision model. That is, where incremental net benefit, b , is simply a function of two parameters, incremental cost and outcome, denoted ΔC and ΔE , and that these are observed simultaneously in one dataset (i.e. a within-trial economic evaluation). I also assume that b , ΔE and ΔC are normally distributed. Mean incremental net benefit (b) and variance are calculated as per Equations 1-4 and 1-5 in Chapter 1, repeated here as Equations [3-8] and [3-9]. λ is the willingness to pay for a unit of outcome. The outcome in each example is QALYs, and λ assumed to be £30,000. Note the covariance is expressed as the product of the standard errors and correlation coefficient.

$$\hat{b} = \lambda \Delta \hat{E} - \Delta \hat{C} \quad [3-8]$$

$$v(\hat{b}) = \lambda^2 v(\Delta \hat{E}) + v(\Delta \hat{C}) - \lambda 2 \sigma_{\Delta \hat{E}} \sigma_{\Delta \hat{C}} \rho_{\Delta \hat{E}, \Delta \hat{C}} \quad [3-9]$$

This simple setup with its assumptions of normally distributed parameters allows analytic solutions to Vol on b , ΔE and ΔC to be calculated without concern for the parameters 'behind' ΔE and ΔC (e.g. health state utilities, unit costs, resource quantities etc), or the structure of a decision model. Both the implications of structural uncertainty and the random 'noise' of numeric solutions are therefore avoided. However, it could be argued that this approach oversimplifies the issue. I therefore provide one further example calculating value of information statistics on a previously published decision model¹⁵ both assuming independence between all parameters, and with structural correlations induced wherever possible, by, for example expressing response rates under one strategy as a function of the baseline response rate and the relative risk.

For the first three example datasets, I calculate the Vol statistics and optimal sample size for further studies. I then recalculate the Vol statistics assuming a correlation coefficient of zero between ΔE and ΔC . This simulates developing a decision model drawing on separate data sources where the correlation is unknown. I then show how the Vol statistics vary as the correlation coefficient is varied between its logical limits of -1 and +1. For the decision model I calculate the Vol statistics as reported in the original analysis (which assumed independence between all parameters), before repeating the calculation with structural correlations incorporated.

Data

Data are drawn from three recently reported clinical trials and one decision model based analysis:

1. The BEfriending and Cost of CARing trial (BECCA);²³⁵
2. The Effectiveness of Leukotriene receptor antagonists in the EVAluation of Asthma Therapies and for health Economics trial (ELEVATE);²³⁶
3. The Conventional ventilation or Extra corporeal membranous oxygenation for Severe Adult Respiratory failure trial (CESAR).²³⁷
4. A cost-utility and value of information analysis of early versus delayed laparoscopic cholecystectomy for acute cholecystitis.¹⁵

The BECCA trial is a study of the cost-effectiveness of a befriending intervention on the quality of life of carers of people with dementia.^{50,235} Data are the total NHS cost and QALYs gained over 15 months of 105 intervention and 113 control subjects. The reference case analysis reported incremental analyses adjusted for baseline cost and utility with 95% CIs calculated via a non-parametric bootstrap (thus allowing for non-normality of the data). Missing data were handled using multiple imputation.²³⁸ For the purpose of the analysis described here, I reanalysed the data, adjusting for baseline values and missing data as previously, but calculating 95% CIs around increments parametrically (i.e. assuming normality). Costs for a future trial are estimated based on original trial budgets, adjusted to reflect actual expenditure and expressed in 2010 prices.

The second example is based on data from a pragmatic randomised controlled trial of leukotriene receptor antagonists (LTRAs) in asthma patients treated in primary care (the ELEVATE trial).^{236, 239-241} The full trial comprised the same experiment on two groups of patients: those at 'step 2' and 'step 3' of the British Thoracic Society's asthma treatment algorithm. The data used here relate to the less severe 'step 2' patients, and are QALYs gained and cost to society over two years (discounted at 3.5%) from 324 patients randomised between LTRAs and inhaled corticosteroids. As in the previous example, missing data were handled in the analysis by means of Rubin's multiple imputation method.²³⁸ Note that budgets for the two experiments were not reported separately. For the purpose of estimating the cost of a new trial, I assumed that the fixed costs were constant, but variable costs were allocated across all 629 patients enrolled in both trials.

The third dataset is a trial of Extra Corporeal Membranous Oxygenation (ECMO) versus conventional ventilatory support for severe adult respiratory failure (the CESAR trial).²³⁷ This trial reported cost and QALYs gained on 84 and 79 intervention and control patients

respectively. The primary end point of the study was 6 months, but cost and QALY data were modelled to a lifetime horizon. The lifetime data are used for this analysis.

The decision model is an analysis of two strategies for treating patients presenting with acute cholecystitis.¹⁵ Conventional management is to delay removal of the gallbladder for several weeks until the initial inflammation has subsided as it was thought to be associated with a higher risk of intra-operative complications. However, a Cochrane review²¹⁴ suggested this was not the case and if delayed, patients were at higher risk of complications such as recurrence of acute cholecystitis, biliary colic, obstructive jaundice and pancreatitis.

The decision model was designed to establish the cost-utility of early versus delayed laparoscopic cholecystectomy. The study drew heavily on the Cochrane review as well as other sources for additional data with which to populate the model. All parameters such as probabilities were modelled as independent beta distributions rather than assigning any 'structural linkages' (modelling one probability as a function of another as previously described). The EVPI and EVPPI were calculated. These showed that at a threshold of £20,000 - £30,000 per QALY, there was only value in additional research to reduce uncertainty in health state utilities. The EVSI of proposed study designs was not calculated. I therefore repeat the analysis calculating the EVSI and optimal sample size (n^*) for a study eliciting health state utilities, with assumptions as to the cost of the research clearly stated. Following this I assign the following structural dependencies calculable from the data reported in the Cochrane review²¹⁴:

- Probability of conversion to open procedure
- Probability of inter-operative complications (bile duct injury, bile leak, 'other')

In the original analysis, each of the three probabilities of conversion (risk of conversion during 'early' LC, during an elective delayed procedure and during an elective emergency procedure) were modelled as independent beta distributions. Structural dependencies were then induced by defining the probabilities under the comparators as functions of the baseline probability via the odds ratios (Table 3-2).

For inter-operative complications, the probabilities of each of the four possibilities (bile duct injury, bile leak requiring ERCP, other complication and no complication) were originally modelled with a Dirichlet distribution. Structural dependencies were induced by expressing the probability of any complication with ELC as a function of the probability with DLC and

the odds ratio, and using Dirichlet distributions to model the type of complication (bile duct injury, bile leak and 'other', Table 3-3).

Table 3-2: Conversion rates to an open procedure

	Converted to open	Successful laparoscopic procedure	P
Conversion if delayed and emergency admission (DLC+Em)	19	23	45.24%
Conversion if operated early (ELC)	46	178	20.54%
Conversion if delayed and operated electively (DLC+EI)	52	166	23.85%
	DLC+Em:ELC+EI	ELC:DLC+EI	
OR	2.6371	0.8250	
Ln(OR)	0.9697	-0.1924	
V(Ln(OR))	0.1214	0.0526	

Table 3-3: Risk of inter-operative complications

	ELC	DLC
Bile duct injury	2	4
Bile leak	8	1
Other complication	21	17
No complication	195	198
OR (ELCvsDLC)*	1.4381	
Ln(OR)	0.3634	
V(Ln(OR))	0.0879	

* OR calculated as the ratio of odds of any complication, i.e. $[(2+8+21)/195]/[(4+1+17)/198]$

I repeat the EVPI and EVPPi calculations as well as calculating the EVSI and n* for the same utility elicitation exercise and compare the results.

3.3.3. Results

BECCA

Summary statistics from the data are in Table 3-4, and net benefit calculated assuming a threshold of £30,000 per QALY. The point estimate ICER is £117,039 and incremental net benefit -£1490 (variance £6.1m). On this basis, the decision would be to reject the intervention. The 95% confidence ellipse of incremental cost and QALYs is shown in Figure 3-5, and the CEAC in Figure 3-6. The distribution of incremental net benefit is shown in Figure 3-7).

Expected value of perfect information, of sample information and expected net benefit of sampling

Using the unit normal loss integral method, the per-patient EVPI at a £30,000 threshold is £414.36 (Table 3-5). The per-patient EVSI of a trial of, for example, sample size $n=100$ in each arm gathering data on b is £188.85. As the sample size approaches infinity, the EVSI of b approaches the EVPI of b (Figure 3-8). The per-patient EVSI cannot be used to make a research funding decision. This should be based on the expected net benefit of sampling (ENBS), which is the per-patient EVSI multiplied by the total (present and future, discounted) patient population, less the cost of sampling (i.e. the cost of the trial). The ENBS maximising point is where the marginal gain from the last patient enrolled is equal to the marginal cost of sampling.

For the purpose of this example, I assume the beneficial population is approximately 770,000 patients (based on a 10 year time horizon, discounted at 3.5%). Research costs for the BECCA trial were based on the original budget categories included in the grant application, adjusted to actual cost and uprated to 2010 costs. This equates to a fixed cost of £469,731 and variable costs of £2,131 per patient (see Appendix A Tables A-1 to A-3 for details).

Under these assumptions, the population EVPI is £318.8m. The population EVSI of a trial of $n=100$ in each arm is £145.3m. A trial of 100 patients in each arm would cost £469,731+2*100*£4,262=£1,044,861, thus the expected net benefit of sampling is £144,274,000. The trial of $n=100$ per arm, collecting b (incremental net benefit) as the outcome would therefore yield a positive net benefit. Figure 3-9 shows the population EVSI and ENBS for a range of sample sizes. The ENBS-maximising sample size is 2279 per arm.

In summary, current evidence is in favour of rejecting the befriending intervention, but there is sufficient decision uncertainty that a further trial is warranted, enrolling 2279 patients in each arm.

Impact of ignoring correlation

Setting the Pearson correlation coefficient, $\rho_{\Delta C, \Delta E}$, to zero is analogous to assuming independence between input parameters: in principle they could have been drawn from different sources.

If $\rho_{\Delta C, \Delta E}$ is set at zero, the variance of incremental net benefit becomes simply the sum of the variances of ΔE and ΔC , which may be either an over or underestimate of the true variance, depending on whether $\rho_{\Delta C, \Delta E}$ is positive or negative (see Equation [3-9]).

In this example, $\rho_{\Delta C, \Delta E}$ is negative, so ignoring it will underestimate the variance of INB at £4,994,210 (Table 3-6), compared with £6,097,911 in the base case (Table 3-5).

As EVPI is a function of the variance, the EVPI is similarly underestimated at £337.79 per patient, compared with £414.36 in the base case (Table 3-6). As $\rho_{\Delta C, \Delta E}$ is increased from -1 to +1, both the variance of incremental net benefit and EVPI fall (Figure 3-10). This follows through to impact on the estimated EVSI of a study of a given sample size, and hence also the EVSI-maximising sample size. In this case, if $\rho_{\Delta C, \Delta E}$ is assumed equal to zero, the efficient sample size of a study is 2329 patients per arm, rather than 2279, a 2% overestimate (Table 3-6).

Plotting for $\rho_{\Delta C, \Delta E}$ between -1 and +1, the optimal sample size varies between 2146 for $\rho_{\Delta C, \Delta E} = -1$, is at a maximum where $\rho_{\Delta C, \Delta E}$ is approximately 0.4 ($n=2380$), before dropping to zero at values of $\rho_{\Delta C, \Delta E} > 0.95$. At this point, the ENBS is negative at all sample sizes (Figure 3-11).

In conclusion, in this data set, the point estimate of $\rho_{\Delta C, \Delta E}$ is approximately -0.26. Ignoring this correlation leads to a slight overestimate of the efficient sample size for a future study estimating b . However, the optimal sample size is much more sensitive to high positive values of $\rho_{\Delta C, \Delta E}$.

Figure 3-5: BECCA Cost-effectiveness plane and 95% confidence ellipse

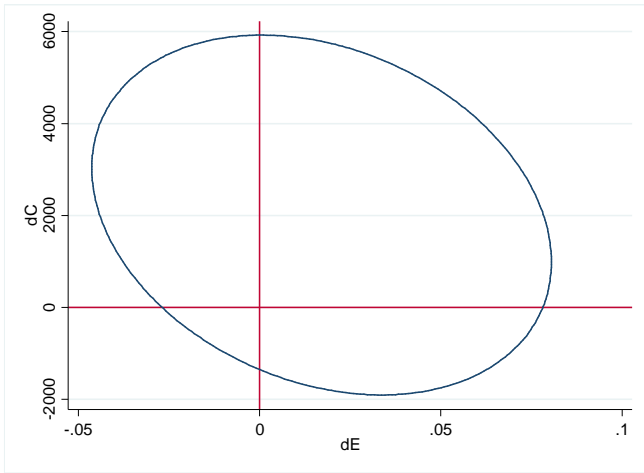


Figure 3-6: BECCA Cost Effectiveness Acceptability Curve

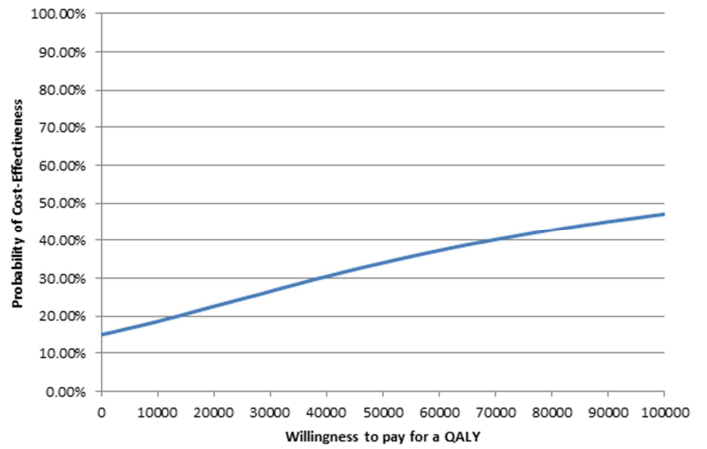


Figure 3-7: BECCA Plot of Incremental Net Benefit @ WTP = £30,000

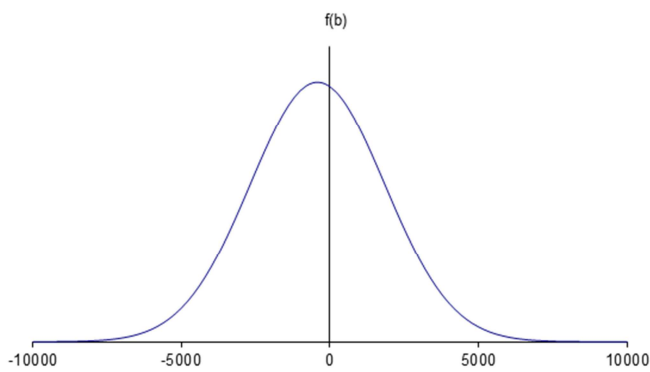


Figure 3-8: BECCA per patient EVSI and EVPI

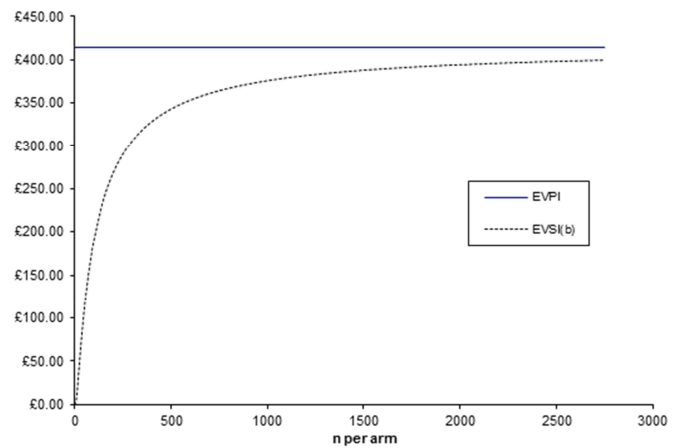


Figure 3-9: BECCA EVSI and ENBS

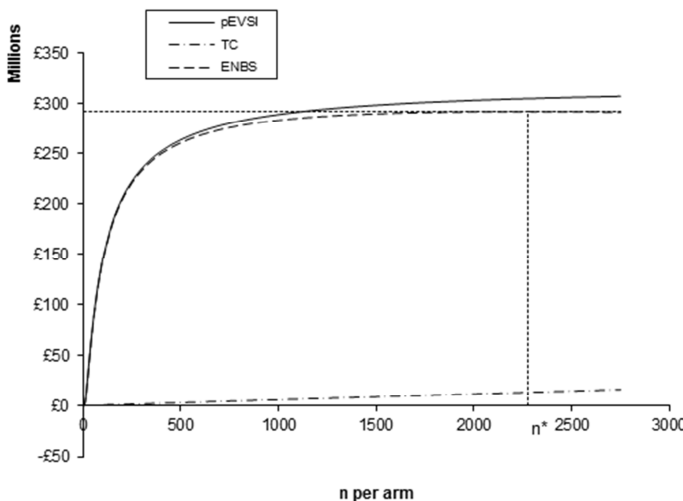


Figure 3-10: BECCA variance of INB and EVPI as a function of correlation coefficient

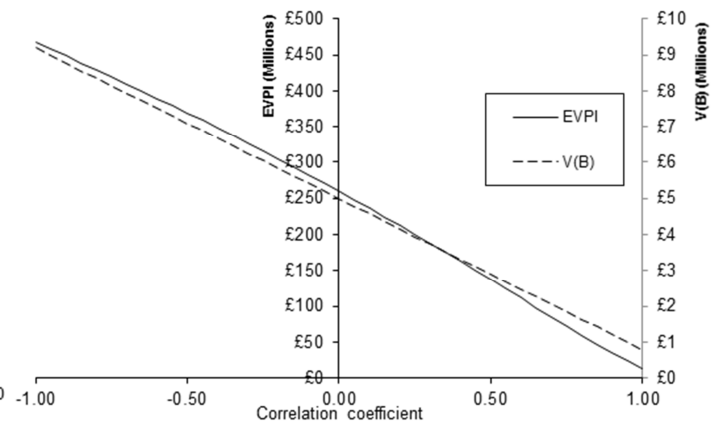


Figure 3 11: BECCA Optimal sample size (n^*) of a new study measuring INB, and ENBS at n^* as a function of the correlation coefficient

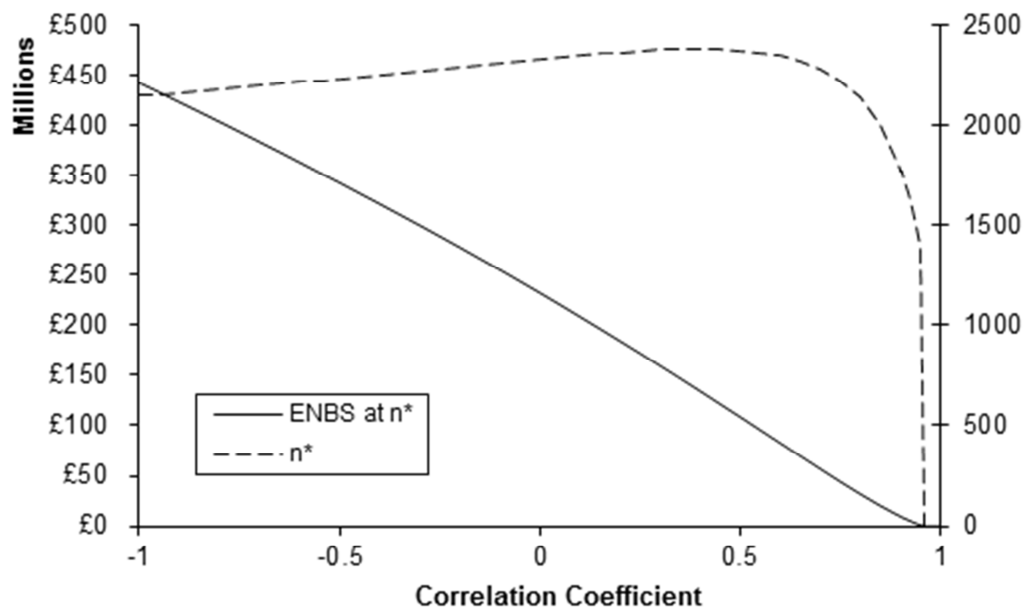


Table 3 4: BECCA Summary Statistics

	Intervention	Control	Increment
N	105	113	
\hat{E}_j	0.946	0.929	$\hat{\Delta}E = 0.017$
\hat{C}_j	£13,740	£11,737	$\hat{\Delta}C = \text{£}2,003$
NB_j	£14,644	£16,133	$\hat{\Delta}NB = \hat{b} = -\text{£}1,490$
$S^2(E_j)$	0.060	0.068	
$S^2(C_j)$	£253,411,414	£140,659,473	
$S^2(NB_j)$	£360,990,075	£260,589,328	
$V(\hat{E}_j)$	0.00058	0.00070	$V(\hat{\Delta}E) = 0.00127$
$V(\hat{C}_j)$	£2,413,442	£1,435,301	$V(\hat{\Delta}C) = \text{£}3,848,743$
$\rho(\hat{E}_i, \hat{C}_i)$	-0.226	-0.315	$\rho(\hat{\Delta}E, \hat{\Delta}C) = -0.263$
$V(NB_j)$	£3,438,019	£2,659,892	$V(\hat{\Delta}NB) = V(\hat{b}) = \text{£}6,097,911$

Table 3 5: BECCA Vol Statistics

Parameter	Per patient	Population*
$\hat{\Delta}NB (= \hat{b})$	£1,490	£1,490
V(b)	£6,097,911	£6,097,911
EVPI	£414.36	£318,840,000
EVSI _b (n=100 in each arm)	£188.85	£145,319,000
Trial of:	n=100 per arm	n=n* per arm
N	100	2279
EVSI	£145,319,000	£305,240,000
Fixed cost	£469,731	£469,371
Variable cost	£426,176	£9,713,000
Total cost	£1,044,861	£13,577,000
ENBS	£144,274,000	£291,663,000

*769,484 potential beneficiaries.

The first section of table shows the per patient and population level Vol statistics. The second section shows the population level statistics for trials of size n=100 and size n=n, i.e. the sample size that maximises ENBS.*

Table 3 6: BECCA Vol statistics, rho=0

Parameter	Per patient	Population*
$\hat{\Delta}NB (= \hat{b})$	£1,490	£1,490
V(b)	£4,994,210	£4,994,210
EVPI	£337.79	£259,928,000
EVSI _b (n=100 in each arm)	£124.77	£96,012,000
Trial of:	n=100 per arm	n=n* per arm
N	100	2329
EVSI	£96,012,000	£245,916,000
Fixed cost	£469,731	£469,371
Variable cost	£426,176	£9,926,000
Total cost	£1,044,861	£13,865,000
ENBS	£94,967,000	£232,052,000

*769,484 potential beneficiaries.

The first section of table shows the per patient and population level Vol statistics. The second section shows the population level statistics for trials of size n=100 and size n=n, i.e. the sample size that maximises ENBS.*

ELEVATE

Summary statistics for the ELEVATE data are in Table 3-7 with net benefit calculated assuming a threshold of £30,000 per QALY. The point estimate ICER is -£2,291 (control is dominant) and the incremental net benefit -£2,952.30 (variance £1.57m). On this basis, the decision would be to reject the intervention. The 95% confidence ellipse of incremental cost and QALYs is shown in Figure 3-12, and the CEAC in Figure 3-13. The distribution of incremental net benefit is shown in Figure 3-14.

Expected value of perfect information, of sample information and expected net benefit of sampling

Only a small proportion of the distribution is to the right of the Y-axis, suggesting a small probability of being 'wrong'. This is reflected in the small per patient EVPI of £3.87 (Table 3-8). The per-patient EVSI of a trial of, say, $n=100$ patients in each arm with b as the primary outcome is £0.01. EVPI and EVSI for various sample sizes are plotted in Figure 3-15.

I assume the beneficial population of 6,787,000 patients (based on a 10 year time horizon, discounted at 3.5%). Research costs for the ELEVATE trial were based on the original budget categories included in the grant application, adjusted to actual expenditure and uprated to 2010 costs using the CPI. I estimate a fixed cost of a new trial of £1,305,000 with variable costs of £289 per patient (see Appendix A Tables A-4 to A-6 for details).

On this basis, the population EVPI is £26.276m, and the population EVSI of a trial of $n=100$ in each arm is £84,521 (Table 3-8). A trial of 100 patients in each arm would cost £1,305,470 + £289*2 = £1.658m, yielding an expected net benefit of sampling of -£1.6m. As this trial has a negative net benefit, it would not be efficient to undertake (Table 3-8). However, as the sample size increases, the marginal gain exceeds the marginal cost, thus larger trials would be efficient: the point at which the marginal gain is equal to the marginal cost is at 1822 patients enrolled into each arm (Figure 3-16). This is thus the ENBS-maximising sample size.

In summary, current evidence is in favour of rejecting the intervention, but a new trial with a sample size of 1822 patients in each arm would be justified on economic grounds.

Impact of ignoring correlation

As the correlation coefficient between ΔC and ΔE is negative, setting it to zero will underestimate the variance of b . The EVPI per patient is then underestimated at £3.69 (versus £3.87), and the population EVPI £25.035m (versus £26.276m). The optimal sample size for a trial is now 1783 (versus 1822), or a 2% underestimate (Table 3-9). Calculating n^*

for all values of ρ yields Figure 3-17, with a steady almost linear (although slightly concave) decline in n^* as ρ rises from -1 to +1.

In conclusion, in this data set, the point estimate of $\rho_{\Delta C, \Delta E}$ is approximately -0.13. Ignoring this correlation leads to a slight (2%) underestimate of the efficient sample size for a future study estimating b .

Figure 3-12: ELEVATE Cost-effectiveness plane and 95% confidence ellipse

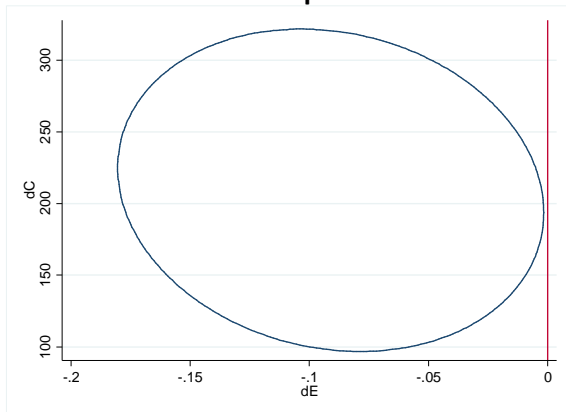


Figure 3-13: ELEVATE Cost-effectiveness acceptability curve

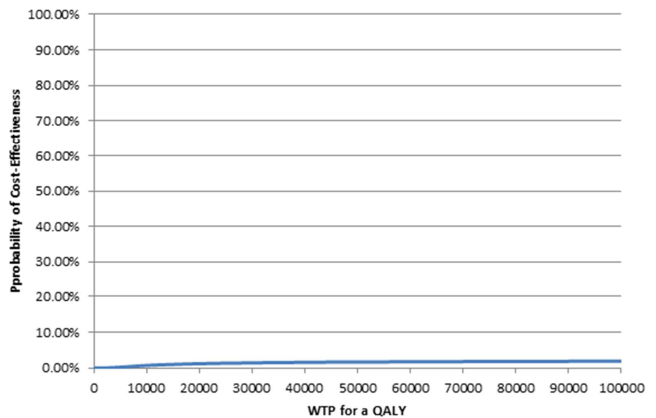


Figure 3-14: ELEVATE Plot of Incremental Net Benefit

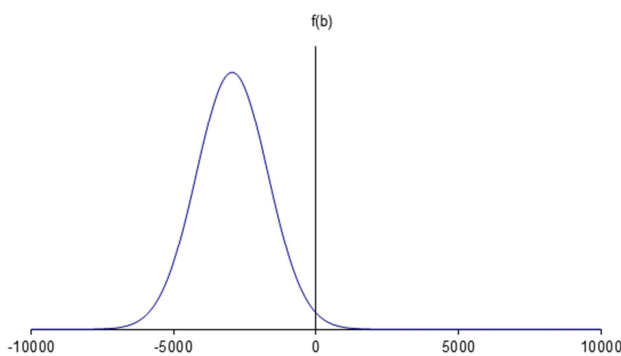


Figure 3-15: ELEVATE per patient EVPI and EVSI

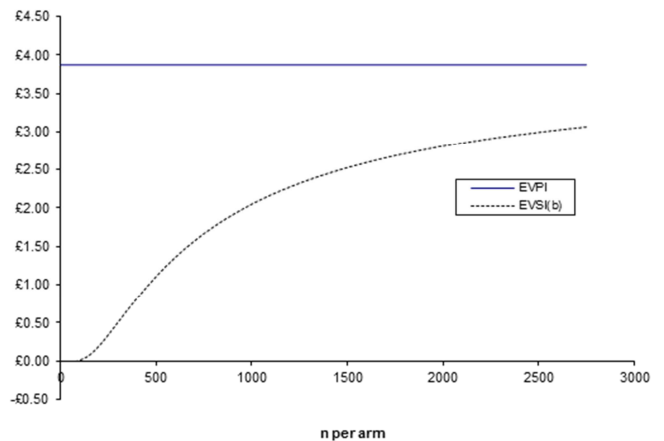


Figure 3-16: ELEVATE Optimal sample size for a new trial

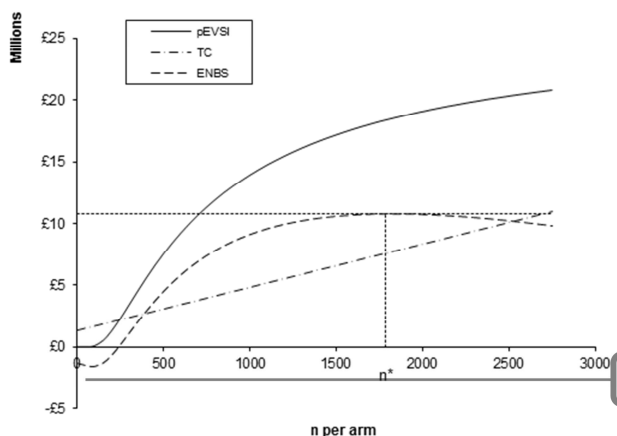


Figure 3-17: ELEVATE Optimal sample size (n^*) of a study measuring INB, and ENBS at n^* as a function of ρ

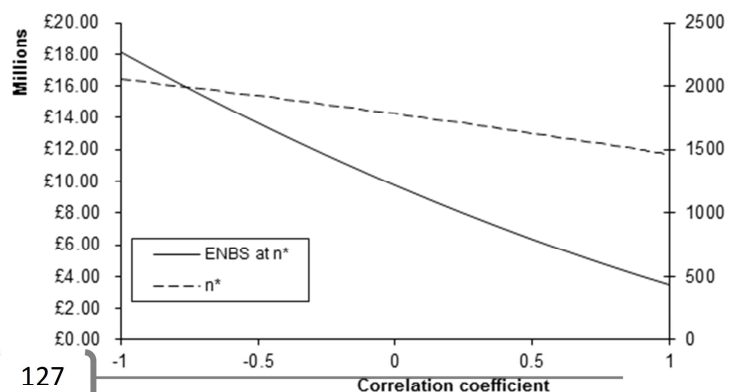


Table 3 7: ELEVATE Summary Statistics

	Intervention	Control	Increment
N	160	164	
\hat{E}_j	1.618	1.710	$\hat{\Delta E} = -0.091$
\hat{C}_j	£564.71	£355.22	$\hat{\Delta C} = £209.48$
NB_j	£47,986.80	£50,939.09	$\hat{\Delta NB} = \hat{b} = -£2,952.30$
$S^2(E_j)$	0.158	0.120	
$S^2(C_j)$	£384,427.24	£125,503.32	
$S^2(NB_j)$	£144,833,880.17	£108,776,891.69	
$V(\hat{E}_j)$	0.001	0.001	$V(\hat{\Delta E}) = 0.002$
$V(\hat{C}_j)$	£2,402.67	£765.26	$V(\hat{\Delta C}) = £3,167.93$
$\rho(\hat{E}_j, \hat{C}_j)$	-0.142	-0.125	$\rho(\hat{\Delta E}, \hat{\Delta C}) = -0.134$
$V(NB_j)$	£905,211.25	£663,274.07	$V(\hat{\Delta NB}) = V(\hat{b}) = £1,568,485$

Table 3 8: ELEVATE Vol Statistics

Parameter	Per patient	Population*
$\hat{\Delta NB}(= \hat{b})$	-£2,952	-£2,952
$V(b)$	£1,568,485	£1,568,485
EVPI	£3.87	£26,276,000
EVSI _b (n=100 in each arm)	£0.01	£84,521
Trial of:	n=100 per arm	n=n* per arm
n	100	1822
EVSI	£84,521	£18,485,000
Fixed cost	£1,305,470	£1,305,470
Variable cost	£57,716	£1,052,000
Total cost	£1,658,000	£7,736,000
ENBS	-£1,574,000	£10,749,000

Table 3 9: ELEVATE Vol Statistics, rho=0

Parameter	Per patient	Population*
$\hat{\Delta NB}(= \hat{b})$	-£2,952	-£2,952
V(b)	£1,549,758	£1,549,758
EVPI	£3.69	£25,036,000
EVSI _b (n=100 in each arm)	£0.01	£71,366
Trial of:	n=100 per arm	n=n* per arm
n	100	1783
EVSI	£71,366	£17,347,000
Fixed cost	£1,305,470	£1,305,470
Variable cost	£57,716	£1,029,000
Total cost	£1,658,000	£7,598,000
ENBS	-£1,587,000	£9,749,000

CESAR

Summary statistics from the CESAR trial data are in Table 3-10, with net benefit calculated at a threshold of £30,000 per QALY. The point estimate ICER is £13,995 and the incremental net benefit £55,073.51 (variance £1.592bn). On this basis the decision would be to adopt the intervention. The 95% confidence ellipse of incremental cost and QALYs is shown in Figure 3-18, and the CEAC in Figure 3-19. The distribution of incremental net benefit is shown in Figure 3-20.

Expected value of perfect information, of sample information and expected net benefit of sampling

The per-patient EVPI at a £30,000 threshold is £1,526.74 (Table 3-11). The per-patient EVSI of a trial of sample size $n=100$ in each arm gathering data on b is £361.49. As the sample size approaches infinity, the EVSI of b approaches the EVPI of b (Figure 3-21).

I make the following assumptions regarding the patient population and cost of sampling: I assume a beneficial population of 504,028 (10 year time horizon, 3.5% discount rate). Based on a 2011 start date, I estimate a new trial would have fixed costs of £1,827,720 and variable costs of £65,102 per patient (see Appendix A Tables A-7 to A-9 for details).

Under these assumptions, the population EVPI is £769.5m, and the EVSI of a trial of $n=100$ patients in each arm would be £182.2m (Table 3-9). A trial this size would cost £20.4m, thus the expected net benefit of sampling would be £161.8m. The ENBS-maximising sample size is 722 patients per arm (Figure 3 22), at an ENBS of £480.5m.

In summary, current evidence is in favour of accepting the intervention, but there is sufficient decision uncertainty to warrant a further trial with 722 patients enrolled into each arm.

Impact of ignoring correlation

In this example, $\rho_{\Delta C, \Delta E}$ is positive (0.353; Table 3-10), so ignoring it will overestimate the variance of INB, and thus overestimate the value of additional information (Table 3-12). Assuming $\rho_{\Delta C, \Delta E}$ is zero leads to an efficient sample size estimate of 767 patients per arm, rather than 722, or a 6% overestimate. Plotting for $\rho_{\Delta C, \Delta E}$ between -1 and +1, the optimal sample size varies between a maximum of 826 for $\rho_{\Delta C, \Delta E} = -1$, falling to 555 at $\rho_{\Delta C, \Delta E}$ (Figure 3-23).

In conclusion, in this dataset, the point estimate of ρ is $+0.353$. Ignoring this correlation leads to a 6% overestimate in the optimal sample size of a future study estimating b .

Figure 3-18: CESAR Cost-effectiveness plane and 95% confidence ellipse

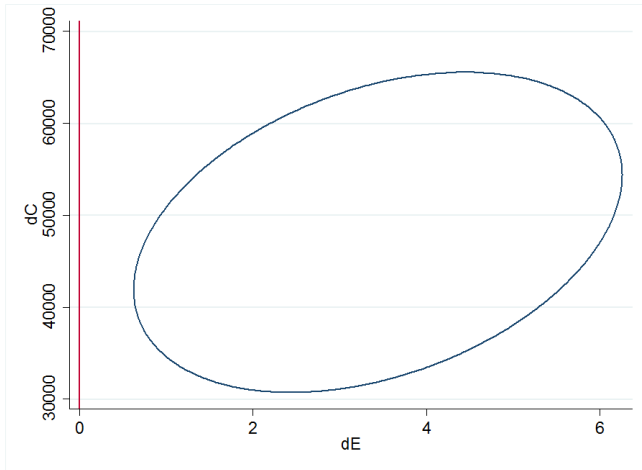


Figure 3-19: CESAR Cost-effectiveness Acceptability Curve

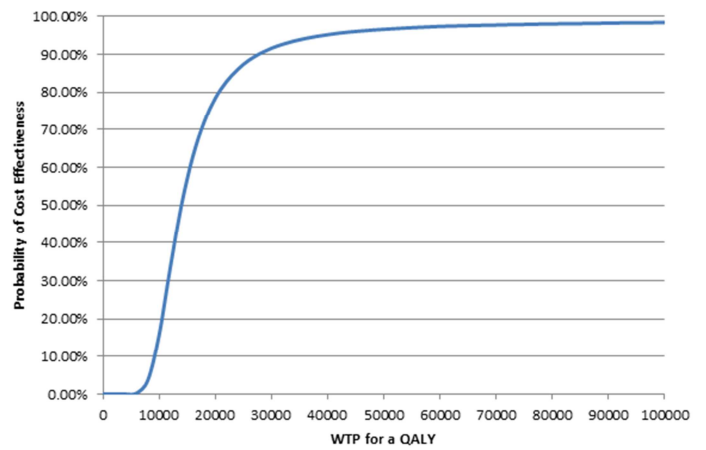


Figure 3-20: CESAR plot of incremental net benefit

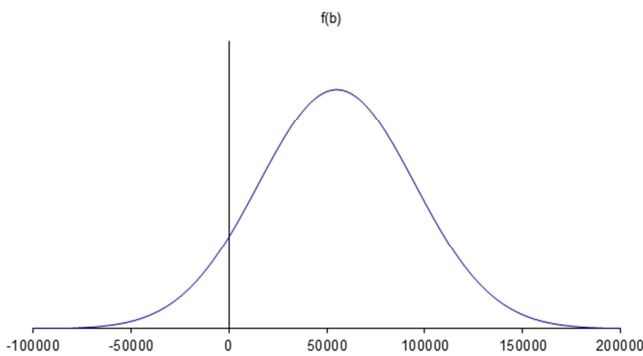


Figure 3-21: CESAR per patient EVSI and EVPI

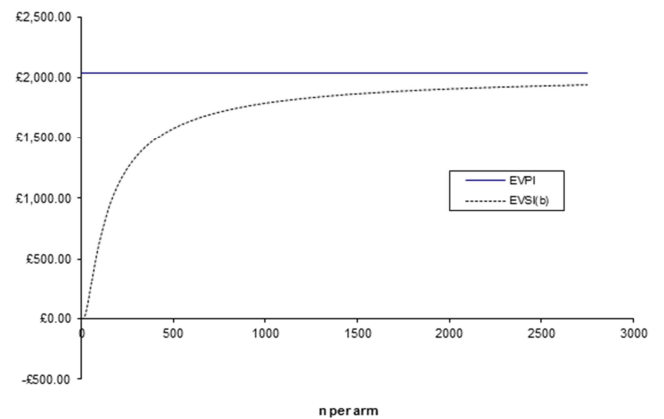


Figure 3-22: CESAR Optimal sample size for a new trial

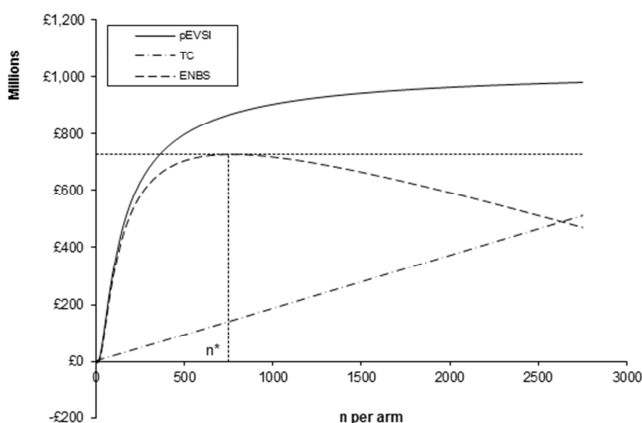


Figure 3-23: CESAR Optimal Sample size (n^*) of a study measuring INB, and ENBS at n^* as a function of ρ

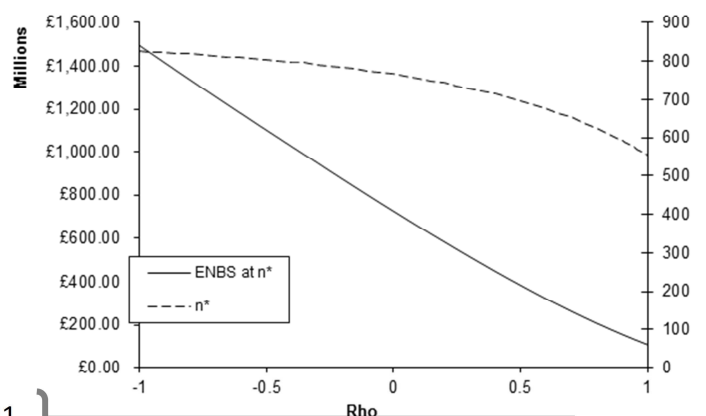


Table 3 10: CESAR Summary Statistics

	Intervention	Control	Increment
N	84	79	
\hat{E}_j	10.751	7.310	$\hat{\Delta E} = 3.441$
\hat{C}_j	£87,463.54	£39,305.70	$\hat{\Delta C} = £48,157.84$
NB_j	£235,067	£179,994	$\hat{\Delta NB} = \hat{b} = £55,073.51$
$S^2(E_j)$	80.717	79.810	
$S^2(C_j)$	£4,112,868,974	£2,087,523,142	
$S^2(NB_j)$	£69,895,319,898	£59,992,880,652	
$V(\hat{E}_j)$	0.961	1.010	$V(\hat{\Delta E}) = 1.9712$
$V(\hat{C}_j)$	£48,962,726	£26,424,344	$V(\hat{\Delta C}) = £75,387,069$
$\rho(\hat{E}_i, \hat{C}_i)$	0.199	0.569	$\rho(\hat{\Delta E}, \hat{\Delta C}) = 0.353$
$V(NB_j)$	£832,087,142	£759,403,553	$V(\hat{\Delta NB}) = V(\hat{b}) =$ £1,591,490,694

Table 3 11: CESAR Vol Statistics

Parameter	Per patient	Population*
$\hat{\Delta NB} (= \hat{b})$	£55,074	£55,074
$V(b)$	£1,591,491,000	£1,591,491,000
EVPI	£1,527	£769,522,000
EVSI _b (n=100 in each arm)	£361.49	£182,202,000
Trial of:	n=100 per arm	n=n* per arm
n	100	722
EVSI	£182,202,000	£616,137,006
Fixed cost	£1,827,720	£1,827,720
Variable cost	£13,020,161	£94,007,726
Total cost	£20,355,532	£135,598,522
ENBS	£161,846,210	£480,538,484

Table 3 12: CESAR Vol Statistics, rho=0

Parameter	Per patient	Population*
$\hat{\Delta NB}(\hat{b})$	£55,074	£55,074
V(b)	£1,849,400,000	£1,849,400,000
EVPI	£2,039.98	£1,028,200,000
EVSI _b (n=100 in each arm)	£645.58	£325,391,000
Trial of:	n=100 per arm	n=n* per arm
n	100	767
EVSI	£325,391,000	£870,532,987
Fixed cost	£1,827,720	£1,827,720
Variable cost	£13,020,161	£99,866,934
Total cost	£20,355,532	£143,936,037
ENBS	£305,035,464	£726,596,950

Acute Cholecystitis

Summary statistics for the acute cholecystitis model with uncorrelated inputs are in Table 3-13, and with the structural correlations in Table 3-14. Incorporation of the correlation does not, in this case alter the adoption decision: incremental net benefit is positive in both cases. However, the point estimate is lower where the inputs are correlated (£718 per patient vs £649). More importantly, the variance of mean incremental net benefit is substantially lower (£7.385m vs £6.452m). The scatterplot of incremental cost and QALYs for both the correlated and uncorrelated models are shown in Figure 3-24, both generating very similar distributions of increments. The resulting CEACs and distributions of incremental net benefit are in Figures 3-25 and 3-26. Whilst the distributions of incremental net benefit are almost identical (Figure 3-26), the CEAC tails off more steeply when inputs are correlated such that at a threshold of £30,000 per QALY, there is an approximately 60% probability that early LC is cost-effective in the base case (uncorrelated) model, but 50% in the correlated model (Figure 3-25).

Expected value of perfect information, of sample information and expected net benefit of sampling

Overall estimates of the EVPI and EVSI appear little changed between the two models, with observed differences possibly attributable to random noise from the simulation approach. The per-patient EVPI at a £30,000 threshold is £463 in the original, uncorrelated model, and £465 in the version with correlations (Tables 3-15 and 3-16). In both cases, the EVPPI is highest around the utility values attached to health states. The per patient EVPPI for utility inputs (EVPPI_{QoL}) is £322 in the base model, but £343 in the correlated model and the EVSI of a health state utility elicitation study of 100 observations is £283 in the base case but £265 in the correlated model.

To calculate the ENBS, I assumed a fixed cost of £50,000 and variable costs of £500 per patient. The potentially beneficial population was estimated at 111,900 (10 year time horizon, 3.5% discount rate).¹⁵

Under these assumptions, the population EVPI is £51.8m in the base model compared with £52.0m in the correlated model. The ENBS of a study of 100 patients is estimated at £32.1m vs £29.1m. Again these differences are most likely due to simulation noise.

However, the optimal sample size for such studies appears to be highly sensitive: I estimate an optimal n of 436 when model inputs are uncorrelated, but only 206 when correlations are

incorporated. The explanation for this could simply be due to simulation noise, however the variable cost per patient also has a large effect on the optimal sample size: If the gradient of the TC curve is very flat (as is the case here), the optimal sample size occurs where the EVSI is similarly flat. Therefore very small changes in the EVSI curve can have dramatic effects on n^* as is the case here (Figure 3-27).

In conclusion, ignoring correlation between parameters may be having a large impact on the optimal sample size. However, it is difficult to disentangle this from noise due to the simulation process. Performing enough simulations to minimise simulation noise would require an unfeasible amount of computer processing time. In this example, the optimal sample size is particularly sensitive to the EVSI because the variable cost per patient is ‘small’ relative to the EVSI.

Figure 3-24: LC Cost-effectiveness plane: Scatterplot and 95% Confidence ellipse

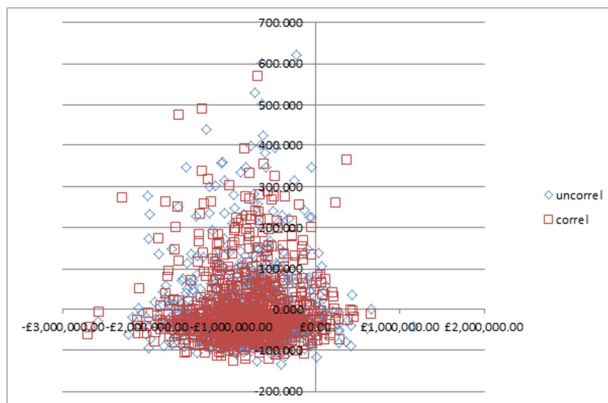


Figure 3-25: LC Cost-effectiveness acceptability curve



Figure 3-26: LC Incremental Net Benefit at $\lambda=£30,000$

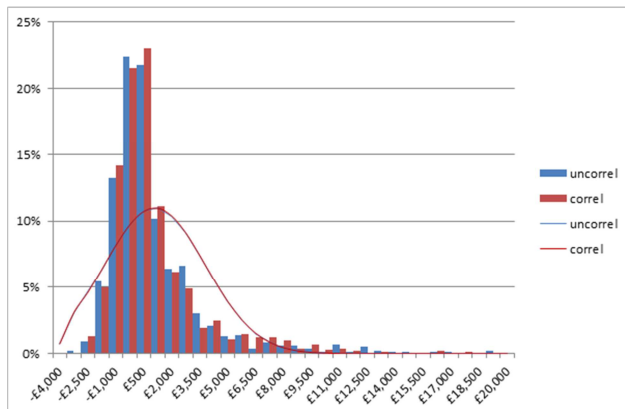
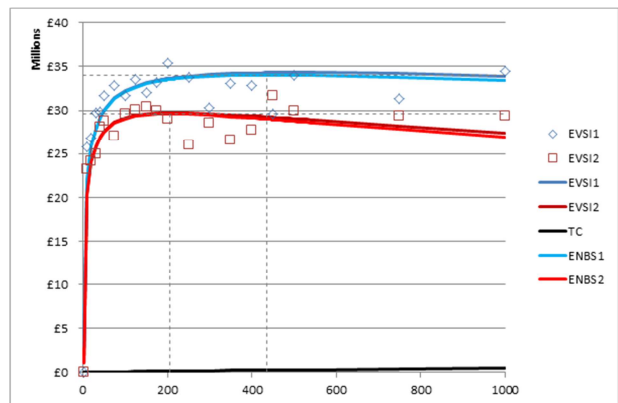


Figure 3-27: LC EVPI & EVSI(b) under i) independent parameters ii) structurally induced correlation*



* Note the EVSI for trials with varying sample sizes was calculated via simulation. This generated a ‘noisy’ estimate of the EVSI curve as shown by the scatterplot (red squares for

the un-correlated and blue diamonds for the correlated model). Two lines were fitted by OLS, $EVSI = a + b \cdot \ln(n)$ and $EVSI = a + b_1 \cdot \ln(n) + b_2 \cdot \ln(n)^2$. The latter incorporating the quadratic term yielded a much better fit of the sampled estimates in both the correlated and uncorrelated models (r^2 of approximately 0.9 versus 0.45 in each case)

Table 3-13: Cholecystectomy summary statistics: uncorrelated

	Early LC	Delayed LC	Increment
\hat{E}_j	0.879	0.882	$\hat{\Delta E} = -0.003$
\hat{C}_j	£2,563	£3,365	$\hat{\Delta C} = -£801$
NB_j	£23,812	£23,093	$\hat{\Delta NB} = \hat{b} = £718$
			$V(\hat{\Delta E}) = 0.008$
			$V(\hat{\Delta C}) = £218,132$
			$\rho(\hat{\Delta E}, \hat{\Delta C}) = 0.003$
$V(NB_j)$	£646,442	£8,197,915	$V(\hat{\Delta NB}) = V(\hat{b}) = £7,384,892$

Table 3-14: Cholecystectomy summary statistics: correlated

	Early LC	Delayed LC	Increment
\hat{E}_j	0.878	0.883	$\hat{\Delta E} = -0.005$
\hat{C}_j	£2,578	£3,384	$\hat{\Delta C} = -£806$
NB_j	£23,758	£23,108	$\hat{\Delta NB} = \hat{b} = £649$
			$V(\hat{\Delta E}) = 0.007$
			$V(\hat{\Delta C}) = £207,676$
			$\rho(\hat{\Delta E}, \hat{\Delta C}) = 0.005$
$V(NB_j)$	£744,651	£7,386,643	$V(\hat{\Delta NB}) = V(\hat{b}) = £6,452,016$

Table 3-15: Cholecystectomy Vol statistics: uncorrelated

Parameter	Per patient	Population*
$\hat{\Delta NB} (= \hat{b})$	£718	£718
V(b)	£7,384,892	£7,384,892
EVPI	£463	£51,840,271
EVPPi _{QoL}	£322	£36,117,227
EVSI _{QoL} (n=100 in each arm)	£283	£32,225,532
Study of:	n=100	n=n*
n	100	436
EVSI _{QoL}	£32,225,532	£34,304,608
Fixed cost	£50,000	£50,000
Variable cost	£50,000	£218,000
Total cost	£100,000	£268,000
ENBS _{QoL}	£32,125,532	£34,036,608

Table 3-16: Cholecystectomy Vol statistics: correlated

Parameter	Per patient	Population*
$\hat{\Delta NB} (= \hat{b})$	£649	£649
V(b)	£6,452,016	£6,452,016
EVPI	£465	£52,031,746
EVPPi _{QoL}	£343	£38,407,816
EVSI _{QoL} (n=100 in each arm)	£265	£29,155,540
Trial of:	n=100	n=n*
n	100	206
EVSI	£29,155,540	£29,748,089
Fixed cost	£50,000	£50,000
Variable cost	£50,000	£103,000
Total cost	£100,000	£153,000
ENBS	£29,055,540	£29,595,089

3.3.4. Discussion

The purpose of this section was to illustrate the impact of ignoring correlation between input variables in a decision model. In order to focus solely on this issue, and to prevent confusion with other issues such as model structural uncertainty and uncertainty due to simulation 'noise', I devised the simplest possible model defining incremental net benefit as a function of two variables, ΔE and ΔC (as well as the constant λ). The source was assumed to be a single randomised controlled trial with piggy-backed economic evaluation, with simultaneous collection of both ΔE and ΔC . The correlation coefficient between ΔE and ΔC is thus observable. Setting this to zero is analogous to developing a decision model drawing on data from separate sources where the covariance structure between the parameters is unknown (for example, ΔE being drawn from one trial, and ΔC from some other source). In order to assess the impact in a 'real life' model, I also imposed a number of correlations between parameters in a previously published model.

In this discussion, I first restate the results of the analyses, finding that in the observed trial-based studies the impact of excluding correlation is minimal, but that there could be a substantial difference in optimal sample size in the decision model based analysis, although it is unclear whether this could simply be due to random noise from the simulation. I then consider two specific issues: determinants of the shape of the n^* curve elicited for the three trial-based studies and a potential solution to obtaining information on the correlation between parameters when the impact is of concern.

Restatement and comparison of results

The BECCA data yielded a point estimate incremental net benefit close to zero with a large variance (-£1,490 and £6.1m respectively). Therefore a large proportion of the distribution (27.3%) was to the right of the y-axis. (Note this is the same as the CEAC showing a 72.7% probability of cost-effectiveness at a £30,000 threshold). This decision uncertainty is reflected in the per-patient EVPI and EVSI. Given the relatively large population would could benefit from the trial results (crudely estimated at 770,000 carers), and assumptions over the cost of sampling, I estimate a new trial of 2,280 patients per arm would be optimal.

In comparison, there was very little decision uncertainty in the ELEVATE results: it is highly likely that the incremental net benefit is negative (99.08% probability). Therefore the per-patient value of information (EVPI and EVSI) is small. However, the disease is very common, thus the total potential health gain is large and when aggregated across the UK population a new trial would be efficient ($n=1,641$ per arm).

The point estimate of incremental net benefit in the CESAR data was positive, suggesting the decision should be to adopt the intervention, but again, there was sufficient uncertainty to warrant an additional trial of 722 patients in each arm.^{xlviii}

The base case decision model-based analysis yielded a positive incremental net benefit suggesting that on average early LC^{xlix} is cost-effective compared with delayed LC, however there is substantial decision uncertainty (point estimate INB £718, SE £2718). Uncertainty was concentrated in estimates of health state utilities, with the EVPPI for such estimated at £322 per patient. Given assumptions over study costs and the beneficial population, a study collecting health state utility data enrolling 436 respondents would be the most efficient sample size.

In each of the three trial-based cases, the optimal sample size for new trials was much larger than the original trial. The estimated budget of each trial was correspondingly high (£13.6m, £10.7m and £135.6m respectively). It is highly unlikely that funders would be willing or able to dedicate such large sums to a single trial. Nevertheless, the results state the optimum sample sizes for future trials, conditional on the value placed on the expected health gain as a result of the additional information yielded. If funders are not willing to spend such large sums on single trials then a possible explanation is that funders have a diminishing marginal valuation on the reduction in uncertainty from a given trial. Alternatively they may be risk averse to allocating an entire budget to one project and place value on funding a variety of projects. This is an area that may be worthy of investigation but is beyond the scope of this thesis.

Setting the correlation coefficient between ΔC and ΔE to zero does appear to change the optimal sample size in the trial-based analyses, but this was by only approximately 2%-6% in each case. Given the uncertainties inherent in patient recruitment and retention to trials, such a small difference is unlikely to be of consequence. However, the model based analysis appeared much more sensitive to the correlations, suggesting an optimal study size of only 206 participants (versus 436). Assuming this is not due to random noise in the simulation approach, an explanation for this is the low variable cost of the study relative to the EVSI (the variable cost is the gradient of the TC curve): The point at which the difference between the EVSI and TC is maximised is where the gradients of the two curves are equal. As this is at

^{xlviii} Of note is that such a trial would take over six years to recruit in the UK if 100% of eligible cases were randomised. My analysis does not adjust for this but could have a large impact on the value of additional information.

^{xlix} Laparoscopic cholecystectomy

a point where the EVSI curve is relatively flat, the optimal sample size can change dramatically for only a very small change in the EVSI curve. Brennan and colleagues explored the impact of correlation between model parameters on estimates of the EVPPI (rather than n^*).²⁰⁷ Their study focused on comparing two methods to calculate the EVPPI for model parameters in the presence of correlation, rather than comparing the impact of ignoring correlation per se. Using a nested two-stage Monte Carlo algorithm (as performed in my acute cholecystitis example above), Brennan and colleagues' results show that ignoring correlations has the potential to underestimate the EVPPI of some parameters.

Whilst the observed analyses in the trial-based studies showed limited impact from ignoring correlation, plotting the optimal sample size for all values of rho between its logical limits of -1 to +1 shows that there is the potential for the optimal sample size to be grossly overestimated in the BECCA data (Figure 3-11): there is a dramatic fall in the optimal sample size where the correlation between ΔC and ΔE is very strongly positive (greater than approximately 0.8). In cases where such a strong positive correlation exists and is ignored, the optimal sample size could be overestimated. Examination of the correlation coefficients between components of the BECCA dataset suggests that generally there are no such strong correlations between parameters, with no values greater than 0.5 observed (Appendix B). If this is typical of the correlations between input parameters, then the impact of ignoring correlation on value of information analyses may be minimal.

The shape of the n^ curve.*

Intuitively, one would expect that the n^* curve would be at minima at perfect negative and positive correlation. This is because a correlation coefficient of -1 or +1 means that there is a deterministic relationship between two parameters (in this case ΔE and ΔC). It would follow therefore that information about one yields information about the other, therefore there should be less to gain from additional information at these two extremes, and hence a lower optimal sample size.

However, the shape of the n^* curve does not follow this expectation. In all three examples, whilst the curve does exhibit a degree of concavity in each case, the optimal sample size at $\rho=-1$ is positive. In the case of BECCA (Figure 3-11), n^* then rises steadily until there is a strong positive correlation ($\rho \approx 0.8$), before declining rapidly. At extremely high values of ρ there is no size of trial which yields a positive ENBS, therefore no trial should be undertaken. It is interesting to note that the equivalent curves for the ELEVATE (Figure 3-17) and CESAR (Figure 3-23) data do not exhibit the same behaviour, showing only a steady decline in n^* as

ρ increases: the rate of change in the rate of change in n^* is less (second derivative of n^* , illustrated by the lesser degree of concavity).

The logic behind this observation is as follows. Consider the BECCA data. The optimal sample size at any ρ is the sample size associated with the maximum net benefit of the trial. That is, where the marginal cost of enrolling the last patient is equal to the marginal gain. This occurs where the gradient of the TC curve is equal to the gradient of the pEVSI curve (Figure 3-9). The gradient of the TC curve is fixed. The gradient of the pEVSI curve is a function of the expected reduction in decision uncertainty from each marginal enrollee (i.e. the per patient EVSI, Figure 3-8). Where there is a great deal of decision uncertainty, the information gained from one additional patient enrolled is proportionally greater than where there is little decision uncertainty (there are diminishing marginal returns to information). Therefore in the case of the BECCA data, the per-patient EVSI curve is steep for small sample sizes.

As the correlation coefficient increases from -1 to +1, the variance of incremental net benefit declines (see Figure 3-28, column 2). This shifts the EVPI and EVSI curves downwards (Figure 3-28, column 3). This has the effect of first increasing n^* , before n^* falls sharply (Figure 3-28, column 4). At a ρ of approximately 0.9576, the ENBS associated with n^* is just zero. At higher values of ρ , there is no n leading to a positive ENBS. This is because at $\rho=0.9576$, two conditions are satisfied:

$$\frac{\delta EVSI}{\delta n} = \frac{\delta TC}{\delta n}$$

AND

$$EVSI(n^*) = TC(n^*)$$

At $\rho=0.9576$, $\frac{\delta EVSI}{\delta n} = \frac{\delta TC}{\delta n} = \text{£}2131$, which is equal to the per patient variable cost, and

$$EVSI(n^*) = TC(n^*) = \text{£}7,929,177.$$

This point is not reached in either the ELEVATE or CESAR examples (Figures 3-29 and 3-30). The reason for this is that no matter how high ρ , the EVSI curve is always greater than the TC curve at some value of n . In other words, the EVSI curve does not decline so dramatically as ρ rises. This is because there is less decision uncertainty in these two examples and hence less to gain from additional information.

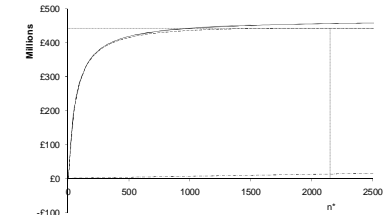
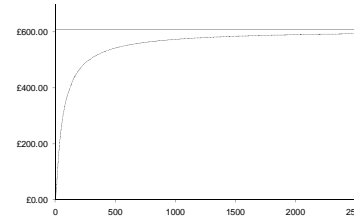
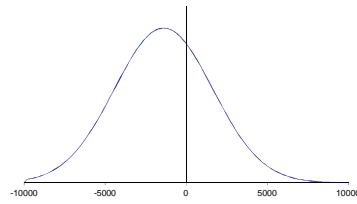
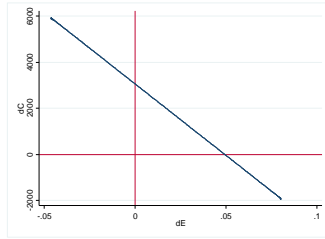
Figure 3-28: Vol statistics for BECCA data as rho is increased from -1 to +1

CE Plane showing 95% Confidence Incremental net benefit Ellipse

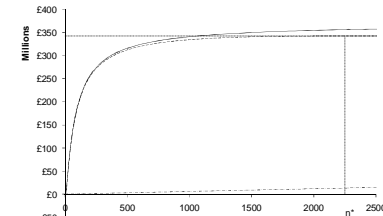
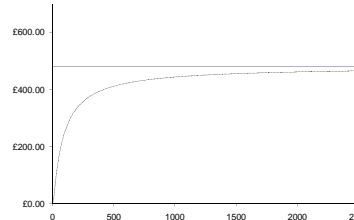
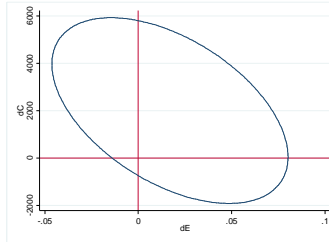
Per px EVSI & EVPI

n*

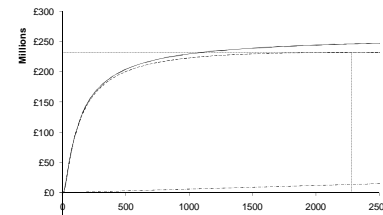
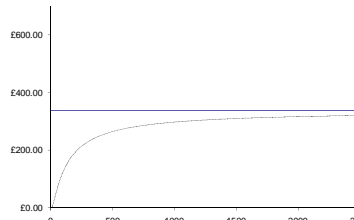
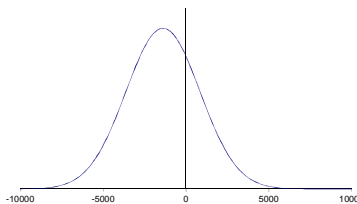
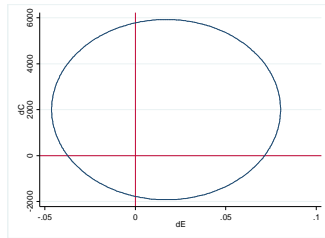
$\rho=-1$



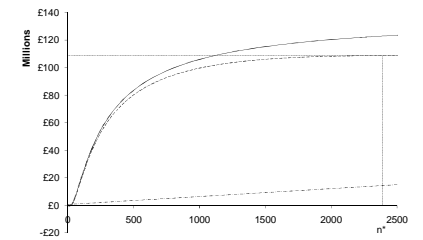
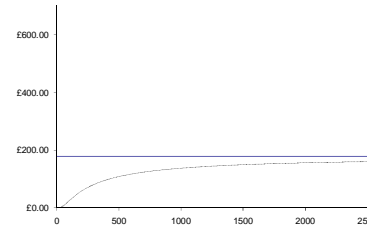
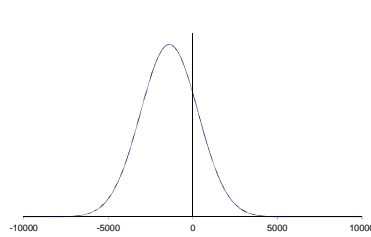
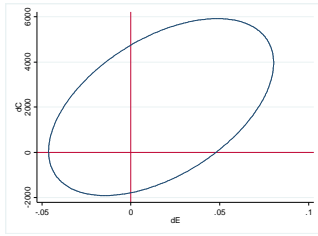
$\rho=-0.5$



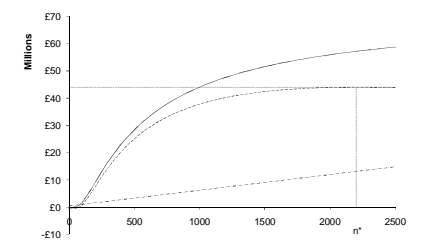
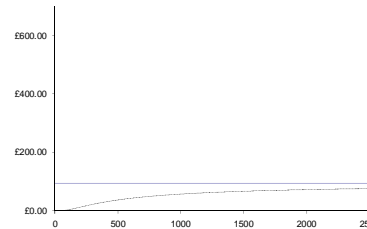
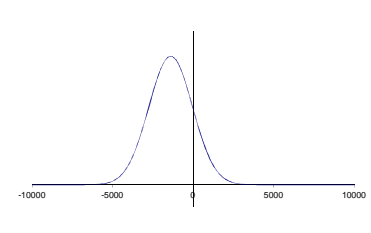
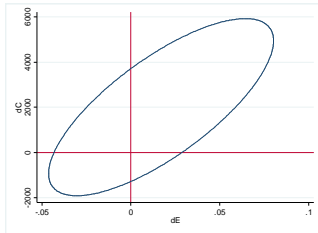
$\rho=0.00$



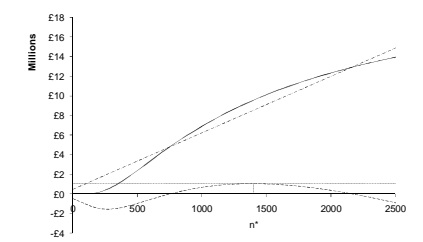
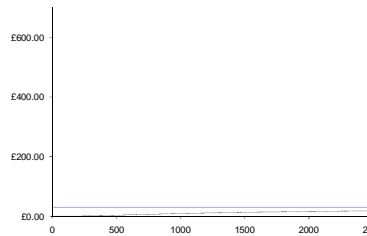
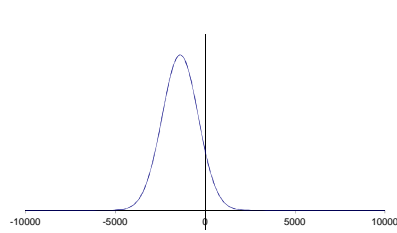
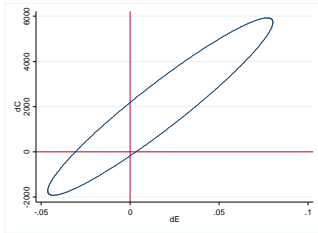
$\rho=0.50$



$\rho=0.75$



$\rho=0.95$



$\rho=1.00$

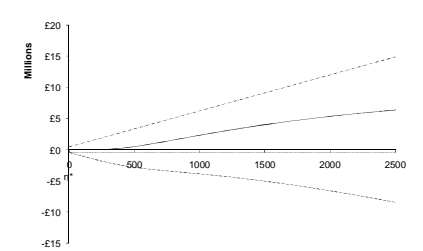
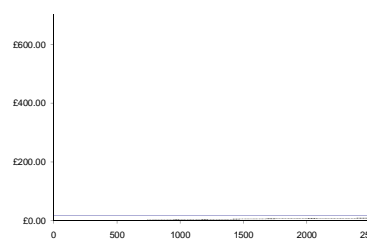
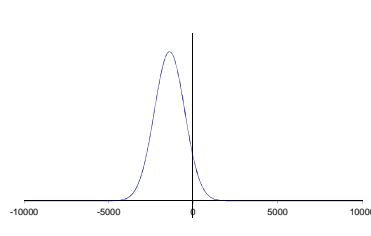
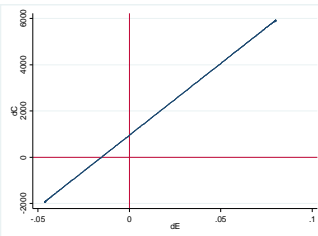
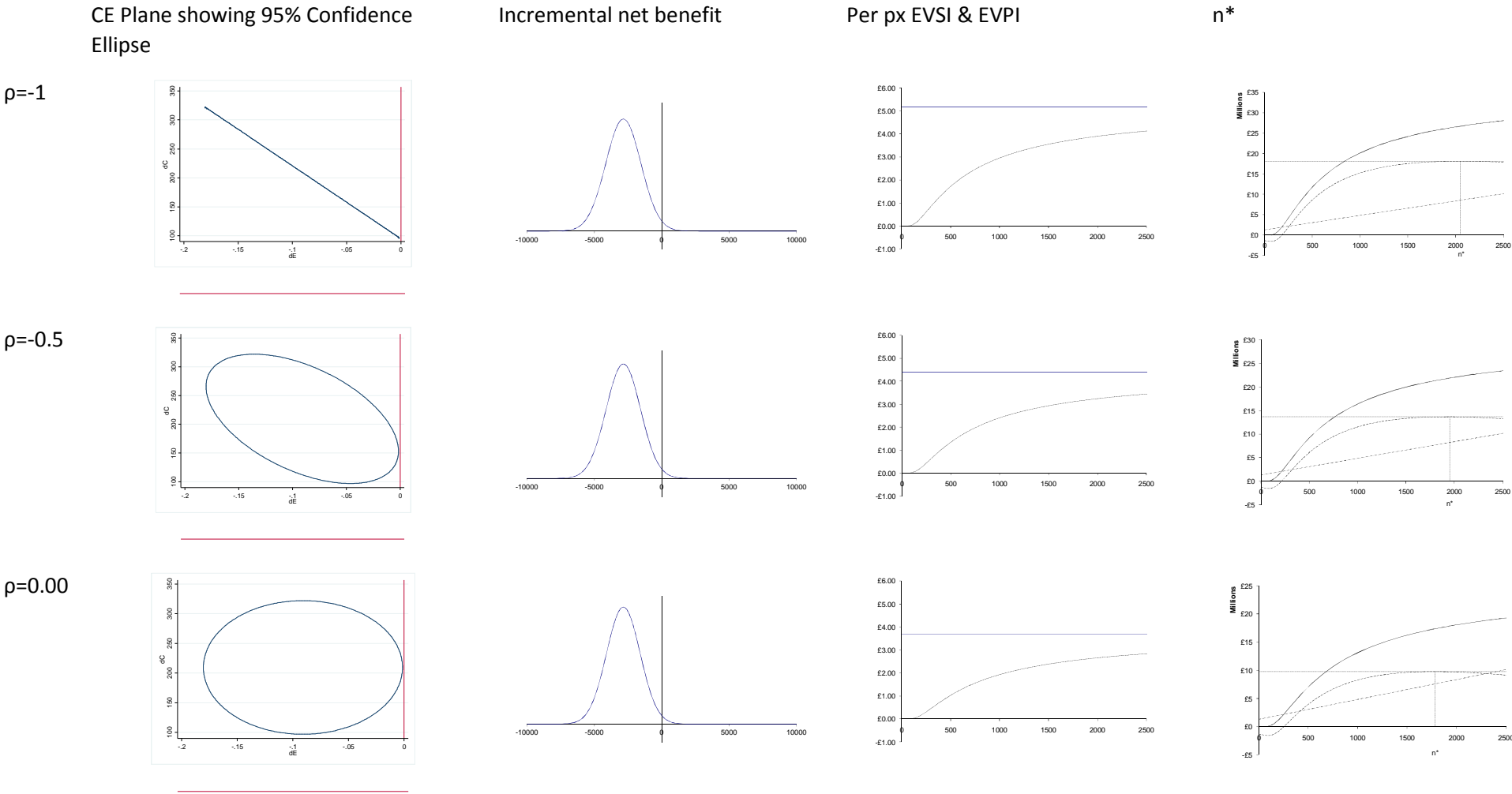
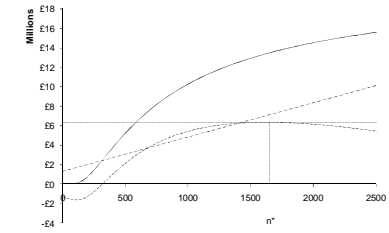
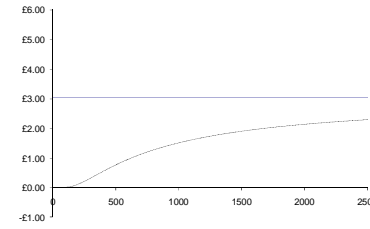
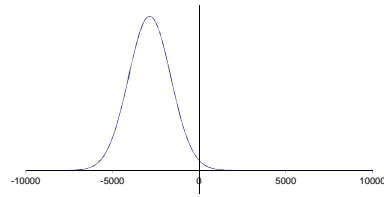
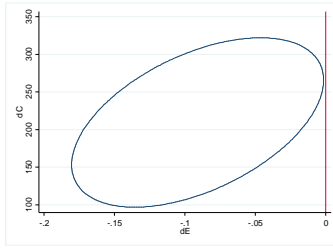


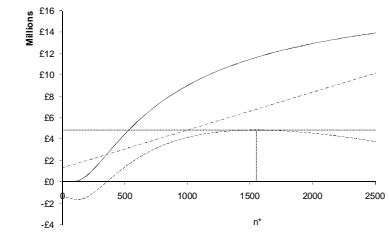
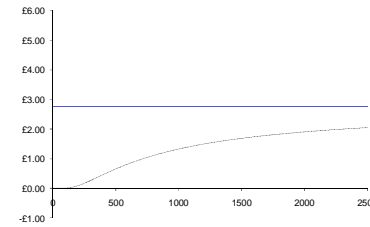
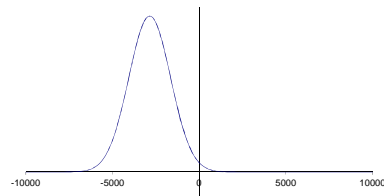
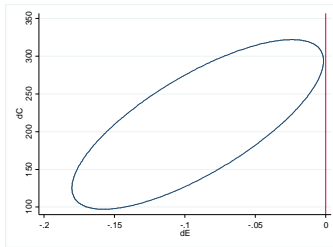
Figure 3-29: Vol statistics for ELEVATE data as rho is increased from -1 to +1



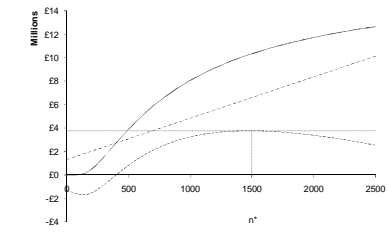
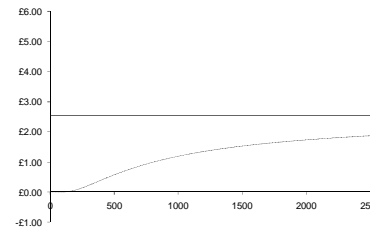
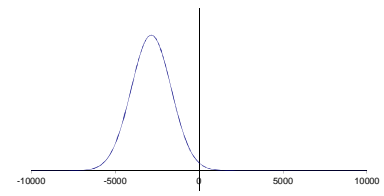
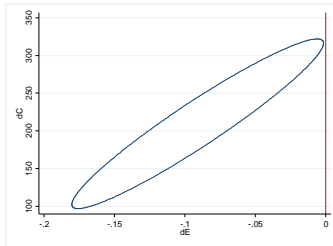
$\rho=0.50$



$\rho=0.75$



$\rho=0.95$



$\rho=1.00$

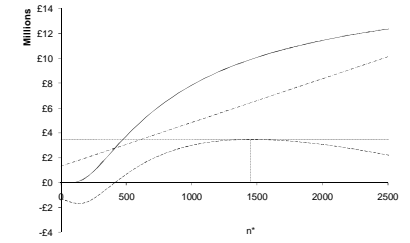
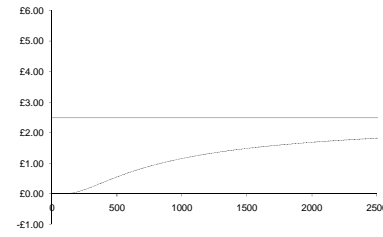
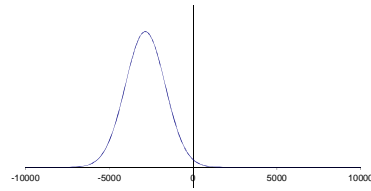
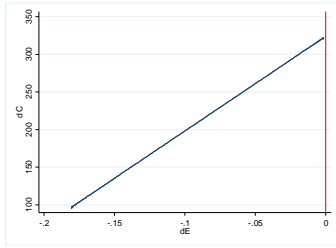


Figure 3-30: Vol statistics for CESAR data as rho is increased from -1 to +1

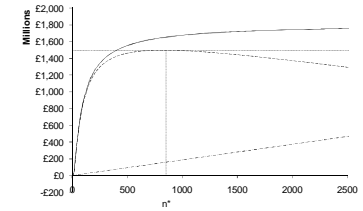
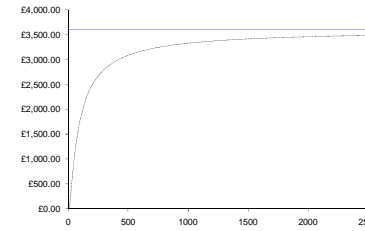
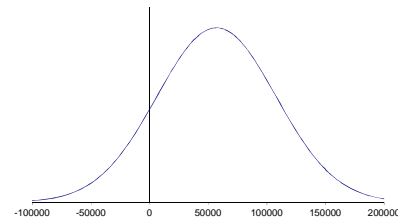
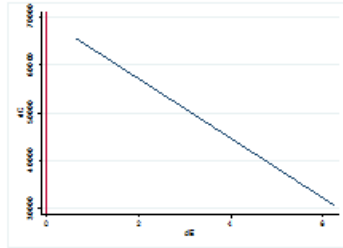
CE Plane showing 95% Confidence
Ellipse

Incremental net benefit

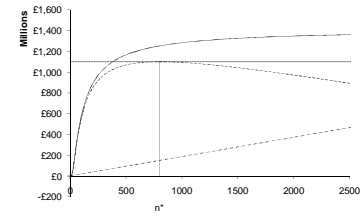
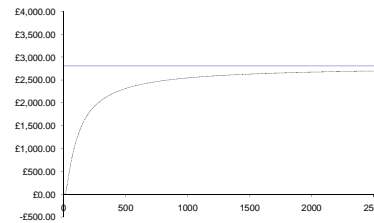
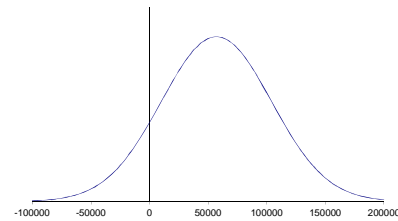
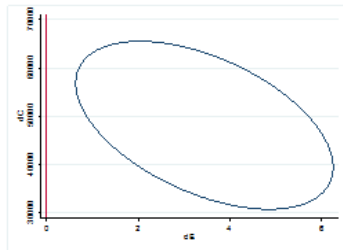
Per px EVSI & EVPI

n*

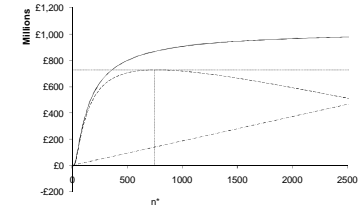
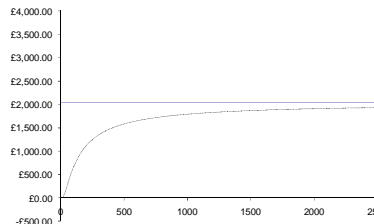
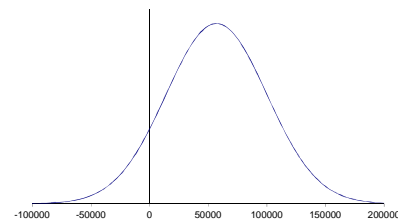
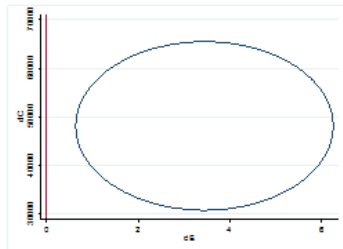
$\rho=-1$



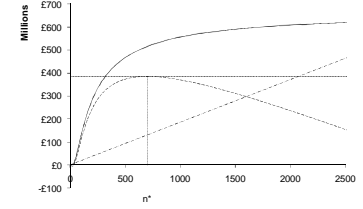
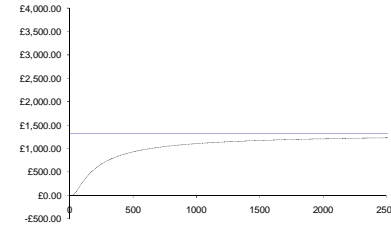
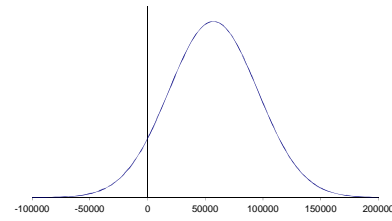
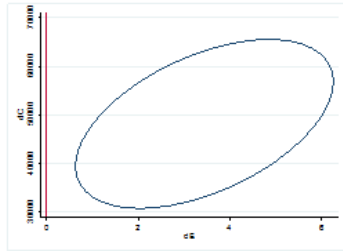
$\rho=-0.5$



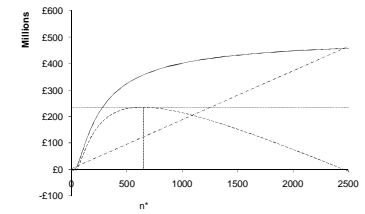
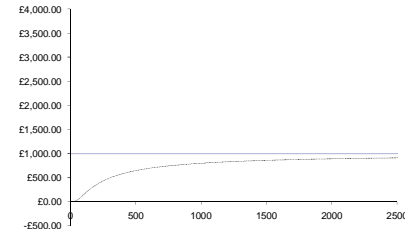
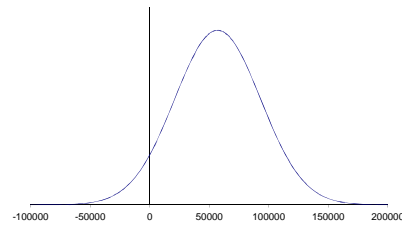
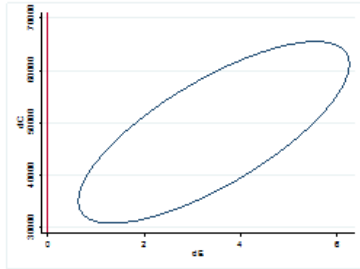
$\rho=0.00$



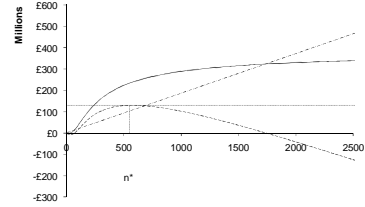
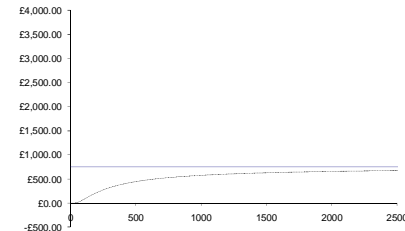
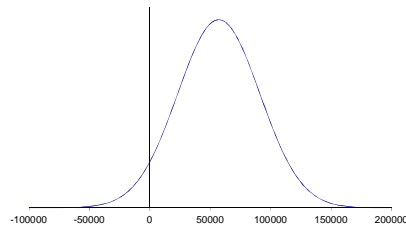
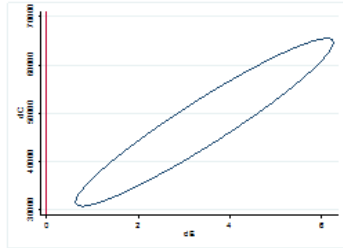
$\rho=0.50$



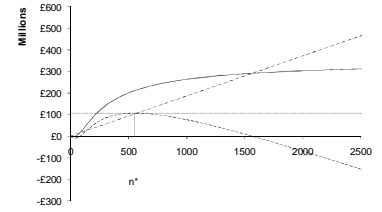
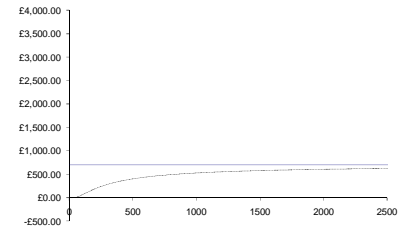
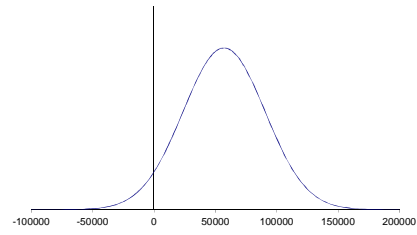
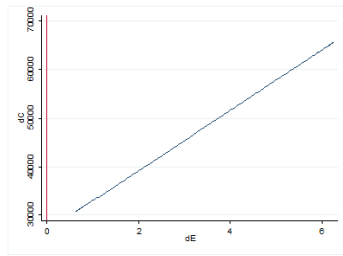
$\rho=0.75$



$\rho=0.95$



$\rho=1.00$



Potential solution

If it is desired to incorporate correlation between parameters in Vol statistics, the only solution is to estimate or elicit some measure. This could be estimated from the literature, analysis of databases or via a formal elicitation process from experts. A tool to assist with this is the SHEffield ELicitation Framework ('SHELF').^{98 218} This is a software tool written in the R language²⁴² and designed to be used in a workshop setting with 5-6 experts from whom an opinion is sought.

The exact format of the elicitation process is fairly flexible, and several formats have been proposed^{243 244} but it typically includes an introductory session to train experts in the techniques, a full presentation of existing information and data, individual elicitation of the distribution of the unknown parameter, and finally combination of the individual distributions into a single distribution representing some aggregate of the group's opinion.

The SHELF software provides a number of methods to elicit individual preferences, based on quartiles (where the expert is first asked for a median value of a parameter followed by the lower and upper quartile), or tertiles (where just the lower and upper tertiles are elicited). Alternatively, a 'roulette' mode allows respondents to place tiles in bins representing a range of plausible values from lowest to highest. The software calculates the best fit of a number of distributions from either technique and presents the results. The distribution is then verified with the expert by considering extreme values and questions such as "do you really believe there is only a 1% / 5% / 25% probability of X being below x?". The distribution is modified until the expert is happy that it closely resembles her beliefs.

It is possible to pool the individual distributions to obtain a composite representing the average either linearly (as the arithmetic mean of the individual distributions) or multiplicatively (the geometric mean). However, O'Hagan does not recommend these as either involves loss of potentially desirable features of the distribution,²⁴⁵ favouring a discursive approach instead. It is important to consider that the summary distribution will not necessarily be the opinion of any of the individual experts, but should be seen as the facilitator's (or commissioner of the panel's) distribution of the parameter of interest, derived from discussion with the experts.

After eliciting univariate distributions, it is natural to consider whether multivariate distributions can be similarly elicited. However, this is not straightforward. O'Hagan mentions copulas as a means of eliciting multivariate distributions, but does not recommend this as it requires direct elicitation of correlation between the marginal distributions. This is conceptually difficult.²⁴⁶ The alternative is to restructure the elicitation process. For example,²⁴⁶ suppose we wish to elicit two parameters A and B, the mean effectiveness of standard treatment and a new drug. It is reasonable to believe that A

and B are not independent: if A is higher than thought, it is reasonable to suppose B will also be higher than thought. To elicit the correlation between A and B, the problem can be restructured by defining $C = A/B$, i.e. the relative treatment effect. It may be reasonable to suppose that the relative effect (C) is independent from the standard treatment effect (A). Therefore A and C could be elicited independently, and the multivariate distribution of (A,B) indirectly estimated.

This approach thus provides a prior estimate of the correlation coefficient, which is itself then a parameter about which further information could be sought. However, it should be noted that information on the correlation coefficient cannot be collected without also collecting information on the parameters themselves. Whilst I do consider this issue in chapter 4, this remains an area in need of further exploration.

3.4. Summary & Conclusion

Value of information analysis is a viable method to estimate optimal sample sizes in trials, although there are a number of limitations. In particular, this chapter considered the issue of input parameter correlation in decision analytic models, and the tendency for analysts to ignore such correlations during the development of models. In some cases the implications can be minor, resulting in an approximately 2%-6% error in optimum sample size estimation, however in others ignoring correlation has the potential to grossly overestimate the efficient sample size. Overestimation is more likely where two parameters are very strongly positively correlated (>0.8), or the variable cost of sampling is low relative to the EVSI. Where correlation is considered to be an issue, the only approach in the absence of data is to undertake a formal elicitation process from 'relevant experts'. Direct elicitation of correlations is not recommended, but it may be possible to restructure the problem in terms of independent parameters and compute the resulting multivariate distribution.

Despite the limitations explored in this chapter, value of information analysis provides a useful approach to inform the design of clinical trials, providing a quantitative method to estimate the returns to reducing uncertainty in different inputs to a decision question. In chapter 4, I explore an approach to make use of the principles to address the question of this thesis, that is to estimate when it is worth investing in a more detailed costing method, or when a simpler approach will suffice.

4. Using Vol to assess the level of detail required in resource data collection in an economic evaluation alongside a clinical trial

4.1. Introduction

In previous chapters I outlined the iterative approach to analysing both the adoption and research questions using the techniques of economic evaluation and value of information analysis respectively. I proposed value of information analysis as an appropriate technique potentially adaptable to the question posed in this thesis, considered the strengths and weaknesses of the approach, and examined the issue of correlation between input parameters in particular detail.

In this chapter I revisit the study question and explore the means to address it. I extend the principles of value of information analysis to compare two approaches to measuring the same parameter. The theory is based on the work by Raiffa and Schlaiffer,⁴ the original developers of statistical decision theory. As described in Chapter 2, elements of this theory have been used to inform macro-scale decisions (i.e. comparing RCT with retrospective cohort designs²⁰), but I am not aware of any attempts to use the theory to inform more detailed elements, for example to choose between alternative methods to collect the same data within an RCT.

Consider a parameter of interest, θ . This could, for example, be the increment¹ of a particular resource data item between treatment and control such as rehabilitation attendances post MI, long term care costs post stroke or cost of drugs. Thus θ is an input parameter into the calculation of incremental net benefit. When designing a clinical trial with an objective of measuring incremental net benefit, there are three questions a trialist may ask about θ :

1. Is it necessary to include θ in the estimate of incremental net benefit at all?

¹ Note that θ is therefore the incremental cost or quantity between the two arms, for example incremental drug costs

2. If θ is to be included, then is additional information on θ required or is current evidence sufficient?
3. If additional information on θ is required, then how should the data be collected?

Briefly, (1) can be answered by considering the proportion of incremental net benefit for which θ accounts.ⁱⁱ (2) can be answered by a standard application of value of information analysis as described in Chapter 1, whilst (3), the focus of this thesis, can be addressed with an extension of the principles of value of information analysis to consider two alternative data collection processes, one detailed and hence costly, and another more approximate, but less expensive.

Note I use the term 'data collection process' to cover both the method of measurement and valuation of a particular parameter. Thus processes can vary in measurement, such as obtaining resource use data from a questionnaire versus patient records, or in valuation method, such as costing drugs by applying an average daily cost by drug class versus costing each chemical or brand by mg consumed.

The approximate (and hence inferior) process is assumed to be more likely to give a biased estimate of the incremental net benefit, leading to a higher probability of making the 'wrong' decision. Phrasing this in Bayesian terms, the trialist has prior beliefs about the relative quality of the two data processes, which may be informed by prior data, 'gut feeling' or expert consensus. Thus the question is whether the expected gain from the 'better' data collection process is worth the extra cost.

Decisions are required as to how to collect data on specific parameters. At issue is whether using an inferior data collection process leads to sufficient bias in the estimate of incremental net benefit for it to affect the decision,ⁱⁱⁱ and whether it is worth investing in the superior (and assumed more expensiveⁱⁱⁱⁱ) study design.

This issue (i.e. question (3) above) is the focus of the method explored in this chapter, although in the example analysis I present results which answer both questions (2) and (3).

ⁱⁱ There is very little guidance as to what cost items should be included beyond a 'rule of thumb' comprising subjective judgement as to whether a particular item will have a sufficient impact on incremental net benefit or whether it is appropriate to the study question.

ⁱⁱⁱ In other words, if the inferior process leads to a negative estimated mean INB when the 'true' mean is positive and vice versa.

ⁱⁱⁱⁱ Where the less biased process is also less expensive, that process dominates the biased process and the decision to use it is unambiguous.

As stated earlier, Pratt, Raiffa and Schlaiffer define an approach to calculating the expected value of information from one data collection process compared with another.⁴ Briefly, the (prior) relationship between the two data processes is defined as a bivariate normal distribution. Critically the covariance between the two is specified. Such data could be obtained from a pilot study where both processes are observed in the same patient group, or obtained from experts using a suitable elicitation process.⁹⁸ Specifying this prior relationship allows calculation of the predicted posterior mean and variance of both processes after data are observed on either process or some mix of the two.

A priori, one process is considered superior to the other but is also more expensive to collect.^{liv} As stated above, an example of this is estimation of drug costs by quantifying exact consumption of every drug by every patient compared with an approximation based on total number of prescriptions of a particular drug or drug class. Another example is the use of time and motion studies combined with exact drug use, consumables and allocation of overhead costs to estimate exact resource consumption for an inpatient hospital admission compared with an approximation based on length of stay or mean cost per admission.

Call the superior process A and the inferior process B. The estimate of (incremental) cost yielded from process A should be used in the calculation of incremental net benefit (as it is the 'better quality' estimate). Specifying the prior bivariate distribution allows one to calculate the posterior mean and variance of process A with information on process B alone, or a mix of information on A and B (or indeed from A alone, although in this case the analysis reduces to the method described in Chapter 1).^{lv}

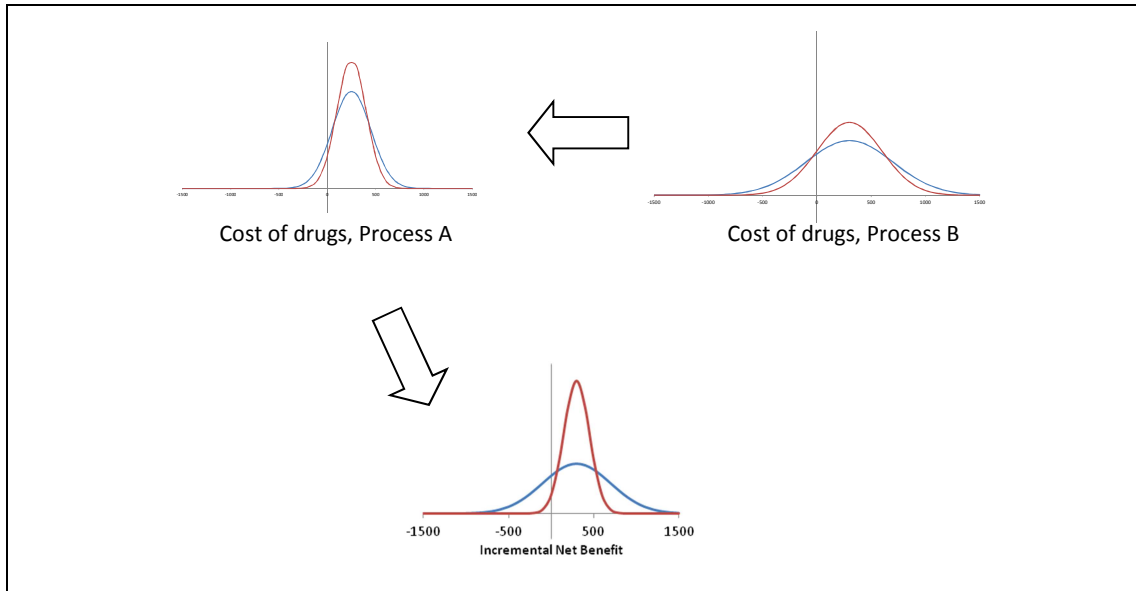
Predicting the expected reduction in variance in process B from a study of given sample size (i.e. the preposterior distribution of B) then allows prediction of the expected reduction in variance in process A, which is then followed through to a predicted reduction in decision uncertainty (i.e. predicted reduction in variance of incremental net benefit). The EVSI and the ENGS of the study (when combined with information on cost) can then be calculated. This approach can be repeated with various combinations of observations on A and B. The combination that maximises the ENGS is the optimal combination. This is shown conceptually in Figure 4-1, where processes A and B are alternative methods for calculating drug costs in a clinical trial: the predicted reduction in variance of drug costs following a data

^{liv} For a discussion of comparisons of different methods in the literature, please see Chapter 2

^{lv} An alternative approach to articulating the problem is to define the costs function using processes A and B, where process B has an extra term to account for the bias. The correlation between the residuals using processes A & B is then the matrix of interest. This may provide a more flexible approach and is discussed in Chapter 5.

collection exercise of size n using process B is used to predict the reduction in variance of drug costs using process A, which in turn is used to predict the reduction in variance of incremental net benefit.

Figure 4-1: Calculation of ENGS of a trial using a combination of processes A & B



The remainder of this chapter is structured as follows. Section 4.2 outlines the algebra required to solve the optimal mix of observations on processes A and B. In this section I also describe the source data for the applied example, where incremental net benefit is defined as a function of three components: incremental QALYs, non-drug costs and drug costs. I then apply these methods to the subject dataset in Section 4.3, conducting a complete analysis to predict the optimal sample size for a new trial reporting incremental net benefit as its primary outcome (4.3.1), then disaggregated in to trials reporting incremental QALYs and cost, and of trials collecting drug and non-drug cost alone (4.3.2). Section 4.3.3 is the key analysis where I compare two alternative approaches to calculating drug costs (processes A and B) and calculate the optimal mix of observations on each in a trial collecting drug cost data alone. Finally I bring 4.3.2 and 4.3.3 together into one analysis reporting the optimal number of observations of QALYs, non-drug cost and drug costs measured using processes A and B (4.3.4). This section thus shows how a decision analytic approach can be used to determine the optimal design of a proposed trial in terms of how many observations are required on each component of incremental net benefit. The results are discussed in Section 4.4.

4.2. Methods

In this section I outline the algebra required to calculate the ENGS of collecting data using one or both of two processes, A and B.

I make the following assumptions:

1. Decision makers are risk neutral and so will adopt the course of action with the highest expected net benefit.
2. There are only two possible courses of action (i.e. two comparator treatments), thus the decision problem simplifies to choosing the new treatment if the incremental net benefit is greater than zero.^{lvi}
3. There are two possible approaches to collecting the target parameter (itself a component of (incremental) net benefit), one imperfect (i.e. yields a biased estimate of incremental net benefit, referred to as process 'B') and the other 'perfect' (unbiased, process 'A'), but more expensive.
4. Data can only be collected from each patient enrolled in a study using either process A or process B, but not both.

The fourth assumption is a simplifying assumption which to a certain extent limits the generalizability of this analysis. For example, a reasonable approach to data collection would be to collect data using process B on all participants, and to verify the quality of process B by collecting data using process A on a subset. This analysis currently does not allow this situation: Pratt, Raiffa and Schlaifer⁴ suggest that 'alternative methods' are required to accommodate this, but they do not elaborate further. I explore why this is the case in Appendix D.

I define net benefit in arm j , b_j , as the value of the health gain (defined as QALYs gained, E multiplied by the value attached to a QALY, λ , less the cost (equation [4-1]; arm T=treatment, C=control). For this purpose of this example, cost comprises two components: cost of drugs, denoted C_d^A , and all other (non-drug) costs, C_n (equation [4-2]).^{lvii} If the current estimates of mean costs and outcomes are from a clinical trial where individual patient data are available, these are calculated as per equation [4-3]. Alternatively they may

^{lvi} This is to maintain simplicity in the example. Extensions of the analysis to multiple comparators are discussed in Chapter 5.

^{lvii} Note the superscript 'A' refers to drug costs collected using process A. This is explained below.

be based on a meta-analysis of existing data or expert beliefs elicited using an appropriate method (e.g. ²⁴⁷).

Incremental net benefit, denoted Δb , can be defined as the difference in net benefit with each course of action (equation [4-4]). The variance of incremental net benefit, $v(\Delta b)$ is therefore the sum of the variances of net benefit in each arm (equation [4-5]). As b_j is the difference between E_j and C_j , the variance of b_j is the sum of the variances between the components E_j and C_j , less twice the covariance (equation [4-6]). Likewise, as C_j is the sum of the two components C_d^A and C_n , the variance of C_j is the sum of the variances of the components plus twice the covariance (equation [4-7]). As before, if the current estimates of the parameters are extracted from trial data, the variances and covariances are calculated as per Equations [4-8] and [4-9]. As before, alternatively these parameters may be based on meta-analyses of other trials and/or expert opinion.

Inserting equation [4-6] into [4-5] provides an alternative expression for the variance of incremental net benefit as the sum of the variances of incremental cost and outcomes less twice the respective covariances (outcomes rescaled to cost units using λ , equation [4-10]). I then express the covariance as the product of the correlation coefficient and the standard errors (equation [4-11]) and add the subscript '0' to denote the prior parameter estimates to derive equations for the variance of prior incremental net benefit (equation [4-12]) and incremental cost (equation [4-13]).

$$b_j = \lambda E_j - C_j \quad j = T, C \quad [4-1]$$

$$C_j = C_{n,j} + C_{d,j}^A \quad j = T, C \quad [4-2]$$

$$X_j = \frac{\sum_{i=1}^{n_j} x_i}{n_j} \quad j = T, C; X = E, C_n, C_d^A \quad [4-3]$$

$$\Delta b = b_T - b_C \quad [4-4]$$

$$v(\Delta b) = v(b_T) + v(b_C) \quad [4-5]$$

$$v(b_j) = \lambda^2 v(E_j) + v(C_j) - 2\lambda \text{Cov}(E_j, C_j) \quad j = T, C \quad [4-6]$$

$$v(C_j) = v(C_{d,j}^A) + v(C_{n,j}) - 2\lambda \text{Cov}(C_{d,j}^A, C_{n,j}) \quad j = T, C \quad [4-7]$$

$$v(X_j) = \frac{\sum_{i=1}^{n_j} (x_{i,j} - X_j)^2}{n_j(n_j - 1)} \quad j = T, C; X = E, C_n, C_d^A \quad [4-8]$$

$$Cov(X_j, Y_j) = \frac{\sum_{i=1}^{n_j} (x_{i,j} - X_j)(y_{i,j} - Y_j)}{n_j(n_j - 1)} \quad j = T, C; \quad \{X, Y\} = \{C_n, C_d^A\}, \{E, C\} \quad [4-9]$$

$$\begin{aligned} v(\Delta b) &= \lambda^2 v(E_T) + v(C_T) - 2\lambda Cov(E_T, C_T) - \lambda^2 v(E_C) + v(C_C) - 2\lambda Cov(E_C, C_C) \\ &= \lambda^2 (v(E_T) - v(E_C)) + v(C_T) + v(C_C) \\ &\quad - 2\lambda (Cov(E_T, C_T) + Cov(E_C, C_C)) \\ &= \lambda^2 v(\Delta E) + v(\Delta C) - 2\lambda Cov(\Delta E, \Delta C) \end{aligned} \quad [4-10]$$

$$Cov(X, Y) = \rho_{X,Y} \sqrt{v(X)} \sqrt{v(Y)} \quad [4-11]$$

$$v(\Delta b)_0 = \lambda^2 v(\Delta E)_0 + v(\Delta C)_0 - 2\lambda \rho_{\Delta E, \Delta C, 0} \sqrt{v(\Delta E)_0} \sqrt{v(\Delta C)_0} \quad [4-12]$$

$$v(\Delta C)_0 = \lambda^2 v(\Delta C_n)_0 + v(\Delta C_d)_0 + 2\lambda \rho_{\Delta C_n, \Delta C_d, 0} \sqrt{v(\Delta C_n)_0} \sqrt{v(\Delta C_d)_0} \quad [4-13]$$

Note that there are five parameters on which information could be obtained after which the distribution of b can be revised. These are not only ΔE , ΔC_n , and ΔC_d^A , but also $\rho_{\Delta E, \Delta C}$ and $\rho_{\Delta C_n, \Delta C_d^A}$.

Now assume that a second data process (process B) is available which can be used to estimate the incremental cost of drugs, ΔC_d^B . I assume that process A is the superior process, that is, using process A will lead to an unbiased estimate of b_0 , and that process B will lead to a biased estimate of b_0 .^{lviii} Given prior belief that A is a 'superior' data process, the results from process A will be used to calculate incremental net benefit.

The prior expectations and variance/covariance matrix are in Equation [4-14] Similarly, a sample of observations using process A or B, denoted $\Delta C_{d,s}^A$ and $\Delta C_{d,s}^B$ with sample sizes n_A and n_B respectively has a mean and variance/covariance as shown in Equation [4-15].

Thus v_i/n_i is the variance of the mean estimated from sampled data using process i alone (without the prior information).

^{lviii} Or alternatively, process B could still lead to an unbiased estimate of b (incremental net benefit), but be less efficient, i.e. yielding a greater standard error for the same number of observations.

$$E \begin{bmatrix} \Delta C_d^A \\ \Delta C_d^B \end{bmatrix} = \begin{bmatrix} \Delta C_{d,0}^A \\ \Delta C_{d,0}^B \end{bmatrix}, \quad V \begin{bmatrix} \Delta C_d^A \\ \Delta C_d^B \end{bmatrix} = \begin{bmatrix} V(\Delta C_d^A)_0 & Cov(\Delta C_d^A, \Delta C_d^B)_0 \\ Cov(\Delta C_d^A, \Delta C_d^B)_0 & V(\Delta C_d^B)_0 \end{bmatrix} \quad [4-14]$$

$$E \left(\begin{bmatrix} \Delta C_{d,s}^A \\ \Delta C_{d,s}^B \end{bmatrix} \middle| \begin{bmatrix} \Delta C_{d,0}^A \\ \Delta C_{d,0}^B \end{bmatrix} \right) = \begin{bmatrix} \Delta C_{d,0}^A \\ \Delta C_{d,0}^B \end{bmatrix}, \quad [4-15]$$

$$V \left(\begin{bmatrix} \Delta C_{d,s}^A \\ \Delta C_{d,s}^B \end{bmatrix} \middle| \begin{bmatrix} \Delta C_{d,0}^A \\ \Delta C_{d,0}^B \end{bmatrix} \right) = \begin{bmatrix} v(\Delta C_d^A)_s/n_A & 0 \\ 0 & v(\Delta C_d^B)_s/n_B \end{bmatrix}$$

The objective is to combine [4-15] (the likelihood) with [4-14] (the prior) to estimate the posterior distributions. This is done as follows⁴:

1. Define \mathbf{H}' as the inverse of the prior var/covar matrix (Equation [4-16])
2. Define \mathbf{H} as one over the var/covar matrix of the sample data (i.e. the precision matrix, Equation [4-17])
3. Define \mathbf{H}'' as the sum of \mathbf{H}' and \mathbf{H} (Equation [4-18])

The joint posterior variance/covariance matrix is the inverse of \mathbf{H}'' thus the posterior distribution is summarised in Equations [4-19] and [4-20], where \mathbf{m} is the mean from each data process (Equation [4-21]).

$$\mathbf{H}' = \begin{bmatrix} H'_{11} & H'_{12} \\ H'_{21} & H'_{22} \end{bmatrix} = \begin{bmatrix} v(\Delta C_d^A)_0 & Cov(\Delta C_d^A, \Delta C_d^B)_0 \\ Cov(\Delta C_d^A, \Delta C_d^B)_0 & v(\Delta C_d^B)_0 \end{bmatrix}^{-1} \quad [4-16]$$

$$= \frac{1}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2} \begin{bmatrix} v(\Delta C_d^B)_0 & -Cov(\Delta C_d^A, \Delta C_d^B)_0 \\ -Cov(\Delta C_d^A, \Delta C_d^B)_0 & v(\Delta C_d^A)_0 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} n_A/v_A & 0 \\ 0 & n_B/v_B \end{bmatrix} \quad [4-17]$$

$$\mathbf{H}'' = \mathbf{H}' + \mathbf{H} = \begin{bmatrix} H'_{11} + n_A/v_A & H'_{12} \\ H'_{21} & H'_{22} + n_B/v_B \end{bmatrix} \quad [4-18]$$

$$\begin{bmatrix} \Delta C_{d,1}^A \\ \Delta C_{d,1}^B \end{bmatrix} = \mathbf{H}''^{-1} \left(\mathbf{H}' \begin{bmatrix} \Delta C_{d,0}^A \\ \Delta C_{d,0}^B \end{bmatrix} + \mathbf{H}\mathbf{m} \right) \quad [4-19]$$

$$\mathbf{V}'' = \begin{bmatrix} v(\Delta C_d^A)_1 & Cov(\Delta C_d^A, \Delta C_d^B)_1 \\ Cov(\Delta C_d^A, \Delta C_d^B)_1 & v(\Delta C_d^B)_1 \end{bmatrix} = \mathbf{H}''^{-1} \quad [4-20]$$

$$\mathbf{m} = \begin{bmatrix} m_A \\ m_B \end{bmatrix} \quad [4-21]$$

The algebra above thus show how the posterior mean and variance of both parameters are calculated after collection of data on either of the processes, or in the case of preposterior

analysis, the predicted posterior means and variances after a proposed data collection exercise of sample size (n_A, n_B) on each process. The predicted posterior mean and variance of process A are then used in the calculation of the predicted posterior mean and variance of the objective function (incremental net benefit), and thence the ENGS, defined as the EVSI less the cost of sampling. The EVSI is as defined in Chapter 1, which can be summarised as the unit normal loss multiplied by the predicted reduction in standard error of incremental net benefit, multiplied by the beneficial population. The cost of sampling is divided into a fixed cost which is incurred if either n_A or n_B are greater than zero, and a variable cost per patient of ks_A and ks_B for processes A and B respectively. Algebraically then the problem is to choose (n_A, n_B) that maximises the ENGS (Equation [4-22]).

$$ENGS_{n_A, n_B} = (N - 2(n_A + n_B)) \cdot \sigma^* L_{N^*}(D^*, b_0) - [k_{sA}n_A + k_{sB}n_B + K_s I\{n_A > 0 \cup n_B > 0\} + (n_A + n_B)b_0] \quad [4-22]$$

where:

$b_0 =$ prior mean incremental net benefit

$\sigma^* = \sqrt{v_0 - v_1} =$ expected reduction in standard error of incremental net benefit

$D^* = \frac{|b_0|}{\sigma^*}$

$L_{N^*}(\cdot) =$ unit normal linear loss integral

The posterior estimate of the variance of ΔC_d^A (cell 1,1 of matrix \mathbf{V}'') can be obtained by predicting the results of n_A and n_B observations using process A or B respectively as described in equations [4-16] - [4-20] (Box 4-1). The resulting reduction in the variance of incremental net benefit can then be used to estimate the ENGS as per Equation [4-22] (Box 4-2 shows calculation of each input element to Equation [4-22]). By calculating for a large range of values of n_A or n_B , the combination yielding the highest ENBS can be identified.

Box 4-1: Preposterior variance of mean of Process A

$$v(\Delta C_d^A)_1 = V''_{11} = H''_{11}^{-1} = \frac{H'_{22} + \frac{n_B}{v(\Delta C_d^B)_s}}{\left(H'_{11} + \frac{n_A}{v(\Delta C_d^A)_s}\right)\left(H'_{22} + \frac{n_B}{v(\Delta C_d^B)_s}\right) - (H'_{12})(H'_{21})}$$

Noting that $H'_{12} = H'_{21} = \frac{-Cov(\Delta C_d^A, \Delta C_d^B)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2}$

thus:

$$\begin{aligned} v(\Delta C_d^A)_1 &= \frac{H'_{22} + \frac{n_B}{v(\Delta C_d^B)_s}}{\left(H'_{11} + \frac{n_A}{v(\Delta C_d^A)_s}\right)\left(H'_{22} + \frac{n_B}{v(\Delta C_d^B)_s}\right) - \left(\frac{-Cov(\Delta C_d^A, \Delta C_d^B)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2}\right)^2} \end{aligned}$$

Substituting in equations for H'_{11} and H'_{22} :

$$\begin{aligned} v(\Delta C_d^A)_1 &= \frac{\frac{v(\Delta C_d^A)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2} + \frac{n_B}{v(\Delta C_d^B)_s}}{\left(\frac{v(\Delta C_d^B)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2} + \frac{n_A}{v(\Delta C_d^A)_s}\right)\left(\frac{v(\Delta C_d^A)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2} + \frac{n_B}{v(\Delta C_d^B)_s}\right) - \left(\frac{-Cov(\Delta C_d^A, \Delta C_d^B)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2}\right)^2} \end{aligned}$$

This is more clearly written by noting that:

$$|\mathbf{H}'| = v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2$$

Thus:

$$v(\Delta C_d^A)_1 = \frac{\frac{v(\Delta C_d^A)_0}{|\mathbf{H}'|} + \frac{n_B}{v(\Delta C_d^B)_s}}{\left(\frac{v(\Delta C_d^B)_0}{|\mathbf{H}'|} + \frac{n_A}{v(\Delta C_d^A)_s}\right)\left(\frac{v(\Delta C_d^A)_0}{|\mathbf{H}'|} + \frac{n_B}{v(\Delta C_d^B)_s}\right) - \left(\frac{-Cov(\Delta C_d^A, \Delta C_d^B)_0}{|\mathbf{H}'|}\right)^2}$$

Box 4-2: Expected Reduction in Variance of Mean Incremental Net Benefit

$$\sigma^* = \sqrt{v_0 - v_1}$$

$$= \sqrt{\begin{aligned} & \left(\lambda^2 v(\Delta E)_0 + v(\Delta C)_0 - 2\lambda \rho_{\Delta E, \Delta C, 0} \sqrt{v(\Delta E)_0} \sqrt{v(\Delta C)_0} \right) \\ & - \left(\lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1} \right) \end{aligned}}$$

Noting that $v(\Delta E)_0 = v(\Delta E)_1$ and $\rho_{\Delta E, \Delta C, 0} = \rho_{\Delta E, \Delta C, 1}$ also noting that:

$$v(\Delta C)_1 = v(\Delta C_n)_0 + v(\Delta C_d^A)_1 + 2\rho_{\Delta C_n, \Delta C_d^A, 0} \sqrt{v(\Delta C_n)_0} \sqrt{v(\Delta C_d^A)_1}$$

$$\therefore \sigma^* = \sqrt{(v(\Delta C)_0 - v(\Delta C)_1) - 2\lambda \rho_{\Delta E, \Delta C, 0} \sqrt{v(\Delta E)_0} (\sqrt{v(\Delta C)_0} - \sqrt{v(\Delta C)_1})}$$

$$D^* = \frac{|b_0|}{\sigma^*}$$

$$L_{N^*}(D^*, b_0) = (\phi(D^*) - D^*[\Phi(-D^*) - I\{b_0 < 0\}])$$

Where $\phi(x)$ = standard normal PDF and $\Phi(x)$ = standard normal CDF.

4.2.1. Data

The data used in the example are taken from the ELEVATE study, a study of leukotriene receptor antagonists compared with conventional treatment in asthma patients.²³⁶ The study comprised two separate trials, on 'step 2' and 'step 3' patients. The data used here are those relating to the more severe 'step 3' patients,²⁴⁰ reanalysed as detailed below from the perspective of the NHS with costs comprising drug and NHS-non drug items (i.e. primary, secondary and tertiary resource use) at two years, and outcomes as QALYs gained at two years (costs and QALYs incurred in year two were discounted at 3.5%).^{lix} Incremental net benefit was calculated at a threshold of £5000 per QALY.^{lx}

Drug cost in the original trial analysis was calculated based on individual items with the unit cost per item extracted from BNF 2005²⁴⁸ using unique BNF code at the individual preparation level. There were 27,028 items of data in the raw dataset extracted from the study database, representing individual prescription items dispensed to 683 patients over two years enrolled in the two trials comprising the ELEVATE study. The cost for each datum was recalculated at the BNF chapter section level, using aggregate cost per prescription as reported in the Prescription Cost Analysis 2005.²⁴⁹ For eight observations, no sub-paragraph or chapter section data were available. Four of these were costs for specific wound dressings so the original unit cost included was applied to both summary cost estimates. The other four were blank entries that were subsequently excluded from all analyses.^{lxi} Therefore every patient had two estimates of drug costs over the two year study period; one based on actual prescribed doses of drugs and the other an approximation aggregated at the BNF section level. Complete drug data were available on all patients. I define process A as the drug costs estimated using actual prescribed doses, and process B as the approximation aggregated at BNF section level.

The other data items were NHS resource use and QALYs gained at two years. 47 (6.9%) and 283 (41.4%) of 683 observations on NHS cost and QALYs were missing. Multiple imputation was performed on the missing data including step, group, sex, age, education and employment status as coefficients. Five iterations were calculated and the results combined using Rubin's rules.²³⁸ Data on the step 2 patients was discarded.

^{lix} The reanalysis was required to structure the data into the three inputs of QALYs, non-drug and drug costs.

^{lx} The threshold was chosen in order to illustrate the point of this chapter. At higher thresholds, due to the nature of the data, there is no value to be gained from research into drug costs.

^{lxi} Given the small number of data (8 out of 27,028 data entries), I did not consider further adjustments for missing data on drug cost necessary.

The resulting summary statistics for the step 3 patients (n=359) are in Table 4-1 and Table 4-2 below. Point estimates suggest an increase in QALYs of 0.034, but also increased NHS non-drug costs of £13.18. Incremental drug cost using process A is estimated at £102.54, leading to an overall incremental cost of £115.72. At a willingness to pay for a QALY of £5000, this leads to an incremental net benefit of £56.41. The adoption decision would therefore be in favour of intervention.

As process 'A' is considered superior to 'B', estimates of mean INB should be based on data from process A. Nevertheless, for the purpose of illustration, recalculating the results using process B yields an incremental cost of drugs of £289.83, leading to an overall incremental cost of £303.00. At the £5000 threshold, the value of the incremental QALYs (0.034*£5000) is insufficient to offset this additional cost, thus the incremental net benefit is negative (-£130.86), and the adoption decision would therefore be in favour of control (Table 4-1).

Estimates of sample variance/covariance and uncertainty in estimates of means are in Table 4-2. I present the estimates of standard deviations and covariance of the different parameters calculated from the study data. The standard errors of the means are calculated by dividing by \sqrt{n} , and form the prior estimates of uncertainty in the mean values (final seven rows of Table 4-2).

The beneficial population and cost of research are as defined in Appendix 1. Briefly, the present and future population is estimated at 6,786,978. The fixed cost of sampling is estimated at £1,305,470, with a variable cost of £288.58 per patient to collect all data components (QALYs, non-drug and drug costs). I assume that the variable costs for a trial collecting solely QALY or cost data are two thirds those of one collecting all data components (£192.39), and that the variable cost of a trial collecting data on non-drug or drug data alone is one third the full cost (£96.19 per patient). Finally, I assume the cost per patient of collecting drug data using process B is one tenth that of process A (£9.62 per patient). When conducting a trial including all cost elements, I assume the variable cost per patient of collecting QALY, non-drug and drug cost data using process A is £96.19 for each (the variable cost of drug cost data using process B is £9.62). Cost estimates are based on trial budget records (University of East Anglia finance office) and personal communication with a trial analyst (Stan Musgrave, University of East Anglia).

Table 4-1: Summary statistics - means

Statistic	Description	Intervention	Control	Increment
n	Sample size	175	184	
E_j	QALYs at two years	1.612	1.578	0.034
$C_{n,j}$	NHS cost (excl. drugs)	£190.53	£177.35	£13.18
$C_{d,j}^A$	Drug cost, process A	£665.58	£563.04	£102.54
C_j^A	Total cost, process A	£856.11	£740.39	£115.72
b_j^A	Net Benefit, process A	£7203.89	£7149.61	£56.41
$C_{d,j}^B$	Drug cost, process B	£801.38	£511.56	£289.82
C_j^B	Total cost, process B	£991.91	£688.91	£303.00
b_j^B	Net Benefit, process B	£7069.70	£7200.57	-£130.86

Figures subject to rounding, net benefit calculated at a value of £5000 per QALY, NHS cost perspective

Table 4.2: Summary statistics - variance and covariance

Statistic	Description	Equation	Intervention	Control	Increment
n	Sample size	-	175	184	
$s(E_j)$	standard deviation, QALYs	Footnote i	0.371	0.386	0.536
$s(C_{Nj})$	standard deviation, non-drug costs	Footnote i	£395.46	£536.81	£666.75
$s(C_{Dj}^A)$	standard deviation, drugs, process A	Footnote i	£416.11	£443.34	£608.03
$s(C_{Dj}^B)$	standard deviation, drugs, process B	Footnote i	£516.65	£384.42	£643.97
$Cov(C_{d,j}^A, C_{d,j}^B)_s$	Sample covariance, drug costs processes A and B	Footnote ii	£183,804.36	£137,185.50	
$s(C_j^A)$	Standard deviation, total cost, drugs estimated using process A	Footnote i	£619.84	£846.73	£1,049.35
$s(b_j^A)$	Standard deviation, net benefit, process A	Footnote i	£2,010.64	£2,356.20	£3,097.47
$v(E_j)^{0.5}$	Standard error, QALYs	Footnote iii	0.028	0.028	0.040
$v(C_{n,j})^{0.5}$	Standard error, non-drug costs	Footnote iii	£29.89	£39.57	£49.60
$v(C_{d,j}^A)^{0.5}$	Standard error, drug costs, process A	Footnote iii	£31.46	£32.68	£45.36
$v(C_{d,j}^B)^{0.5}$	Standard error, drug costs, process B	Footnote iii	£39.05	£28.34	£48.25
$Cov(C_{Dj}^A, C_{Dj}^B)$	Covariance between mean drug costs processes A and B	Eq. [4-9]	£1,056.35	£749.65	£1,805.99
$v(C_j^A)^{0.5}$	Standard error, total costs, process A for drugs cost	Footnote iii	£46.88	£62.48	£78.11

$v(\Delta b_j^A)^{0.5}$	Standard error of incremental net benefit, process A for drug costs	Footnote iv	£217.15
$\rho_{\Delta C_n, \Delta C_d^A}$	Correlation coefficient between ΔC_n and ΔC_d^A	Footnote v	0.352
$\rho_{\Delta C, \Delta E}$	Correlation coefficient between ΔC and ΔE	Footnote v	-0.036

- i) $s(x_j) = \sqrt{\frac{\sum_{i=1}^{n_j} (x_{i,j} - \bar{x}_j)^2}{(n_j - 1)}}$
- ii) $Cov(x_j, y_j) = \sqrt{\frac{\sum_{i=1}^{n_j} (x_{i,j} - \bar{x}_j)(y_{i,j} - \bar{y}_j)}{(n_j - 1)}}$
- iii) square root of equation [4-8]
- iv) square root of equation [4-12]
- v) re-arrangement of equation [4-11]

4.3. Results

The ultimate objective of the analysis is to identify the optimal mix of observations on each data process to collect drug costs (thus answering question (3) posed in Section 4.1). However I also present standard value of information analyses on other components (therefore answering question (2) posed in Section 4.1). This is then broadened to identify the overall optimal number of observations on each drug cost process, non-drug costs and QALYs simultaneously, thus providing a decision analytic approach to overall trial design. Therefore, in the results section, I present the following:

1. Value of information analysis for a repeat of the ELEVATE step 3 trial.

Analysis of uncertainty in incremental net benefit and standard value of information analysis (reporting the expected value of perfect and sample information, expected net benefit of sampling and optimal sample size of a trial measuring incremental net benefit as the outcome collecting all data on all patients). As stated previously, the estimate of drug costs from process A is employed in this analysis.

2. Value of information analysis for studies collecting one component of data alone.

I report results pertaining to studies collecting (a) incremental QALYs (b) incremental cost alone. I then sub-divide cost into two individual studies collecting (c) incremental non-drug costs and (d) incremental cost of drugs (using process A) alone.

3. Comparison of value of alternative data collection processes on drug costs.

Here I report the efficient mix of observations between processes A and B to measure drug costs. Of interest is a comparison of these results with that in 2d above to observe any increase in expected return on investment (ENBS).

4. Overall efficient trial design

The efficient numbers of observations on drug cost process A, drug cost process B, non-drug costs and QALYs are determined simultaneously in this analysis, providing a guide to efficient trial design. The optimal mix is identified using a search algorithm.^{lxii} Working for all analyses are in Appendix C, and summarised in Table 4-3.

4.3.1. Value of information analysis for a repeated ELEVATE trial

Mean incremental net benefit is £56.41 with a standard error of £217.15 (Figure 4-2). The population EVPI is £416.3m (Table 4-3, Appendix 3, Box A3.1). Given a fixed and variable cost of sampling of £1.3m and £288.58 per patient respectively, and beneficial population of 6.8m, the ENBS-maximising sample size is a trial enrolling 8,589 patients per arm (Table 4-3, Figure 4-3, Appendix C, Box C-2), yielding an ENBS of approximately £401.9m. Thus the efficient sample size of a trial reporting incremental net benefit as its outcome is 8,589 per arm.

4.3.2. Value of information analysis of four separate studies reporting incremental QALYs, incremental cost, incremental non-drug cost and incremental drug cost alone

The expected value of eliminating uncertainty in outcomes (QALYs) alone (i.e. $EVPI_{QALYs}$) is £378.3m (Table 4-3, Appendix 3 Box A3.3). Assuming the fixed costs of a study collecting outcomes alone are the same as for a full trial (£1.3m), and the variable costs are 2/3rds that

^{lxii} I implemented this by firstly defining an arbitrary maximum and minimum number of observations for each of the four parameters, e.g. 0-1000. The ENGS was calculated at five values between the min and max (i.e. 0, 200, 400, 600, 800 and 1000 so the 'step' is 200) for each parameter = 5^4 calculations. If the highest ENGS was with one of the parameters at its maximum sample size (i.e. 1000), then the range for that parameter was doubled (i.e. to 2000). The 5^4 calculations were repeated. The algorithm thus identified a first approximate solution. The max and min ranges for each parameter were then set to halfway between the approximate optimum and the previous maximum and minimum to narrow the range (rounded to the nearest whole number), and the calculations repeated. This continued until the step for each parameter was equal to 1, and the optimum solution could therefore be identified. This was written as a Microsoft Excel macro, and took approximately 10 seconds to identify the optimal solution.

of a trial collecting both cost and outcomes data (£192.39 per patient enrolled), the ENGS-maximising sample size is 9,458 per arm, at which point the ENGS is approximately £366.0m (Table 4-3, Figure 4-4, Appendix 3, Box A3.4).

The expected value of eliminating uncertainty in costs alone (i.e. $EVPI_{cost}$) is £87,476,000 (Table 4-3, Appendix 3 Box A3.5). Given the same costs of sampling as for an outcomes study, the ENGS-maximising sample size is 6,735 per arm for this study, at which point the ENGS is approximately £78.7m (Table 4-3, Figure 4-5, Appendix 3 Box A3.6). Further dividing costs into non-drug and drug costs, the EVPI is £51.0m and £44.4m respectively (Table 4-3 and Appendix 3 Box A3.7 and Appendix 3 Box A3.9 respectively). Assuming a variable cost of one third that of a trial measuring incremental net benefit (£96.19 per patient enrolled), the optimal sample sizes of studies collecting data on those components alone are 9,456 and 9,197 per arm respectively, and the ENGS of each are £42.9m and £36.5m (Table 4-3, Figure 4-7 and Appendix 3, Box A3.8 & Figure 4-6 and Appendix 3, Box A3.10 respectively).

Figure 4-8 summarises the EVPI and EVPI on QALYs, non-drug costs and drug costs.

4.3.3. Comparison of value of alternative data collection processes on drug costs

As stated in Section 4.2.1, in the following I assume the variable cost of collecting data using process A is £96.19 per observation, and for process B, £9.62 (one tenth of process A). The optimal sample size of a study collecting drug cost data alone is estimated at 9081 observations using process A plus 240 observations using process B per arm (Table 4-3, Figure 4-9 and Appendix 3, Box A3.11). Figure 4-9 shows a three dimensional plot of the ENGS as a function of the sample size of each component. This peaks at an expected net gain of sampling of £36.5m.

4.3.4. Overall efficient trial design

Calculating for different combinations of $n_{\Delta E}$, $n_{\Delta Cn}$, $n_{\Delta CA d}$ and $n_{\Delta CB d}$, (that is, the number of observations per arm collecting QALYs, non-drug costs, drug costs using process A and drug costs using process B respectively), the ENGS maximising combination can be identified. The combination is (10,787, 4,264, 3,693, 904) for $(n_{\Delta E}, n_{\Delta Cn}, n_{\Delta CA d}, n_{\Delta CB d})$, yielding an ENBS of £403.7m (Table 4-3, Appendix 3, Box 3.12). This compares with the maximum ENGS of a trial reporting INB alone of £401.9m (Table 4-3).

Table 4-3: Summary results

	EVP(P)I	£fixed	£var	n*	EVSI	TC	OC	ENBS
INB	£416.3m	£1.3m	£288.58	8,589	£406.5m	£6.3m	£484,500	£401.9m
QALYs	£378.3m	£1.3m	£192.39	9,458	£366.2m	£4.9m	£533,500	£366.0m
Cost	£87.5m	£1.3m	£192.39	6,735	£ 82.57m	£3.9m	£379,900	£ 78.673m
Non-drug Cost	£51.0m	£1.3m	£ 96.19	9,456	£ 46.06m	£3.1m	£533,400	£ 42.934m
Drug cost (process A)	£44.4m	£1.3m	£ 96.19	9,197	£ 41.49m	£3.1m	£518,800	£ 36.485m
Drug cost (processes A, B)	£44.4m	£1.3m	(£96.19, £9.62)	(9081, 240)	£ 41.52m	£2.5m	£525,800	£ 36.494m
INB (QALYs, non-drug cost, drug cost process A, drug cost process B)	£416.3m	£1.3m	(£96.19, £96.19, £96.19, £9.62)	(10787, 4264, 3693, 904)	£409.3m	£4.9m	£608,500	£403.7m

Figure 4-2: Distribution of Incremental Net Benefit

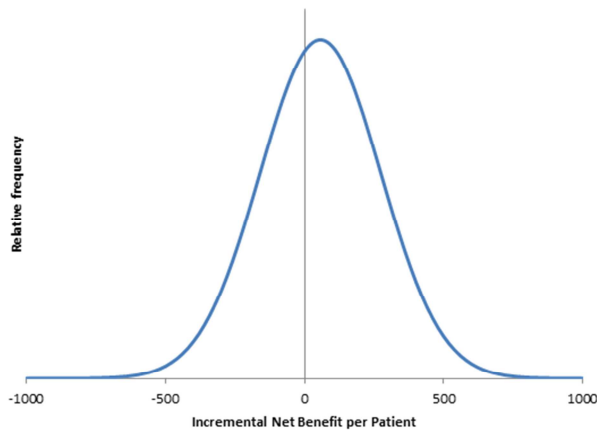


Figure 4-4: ENGS QALYs

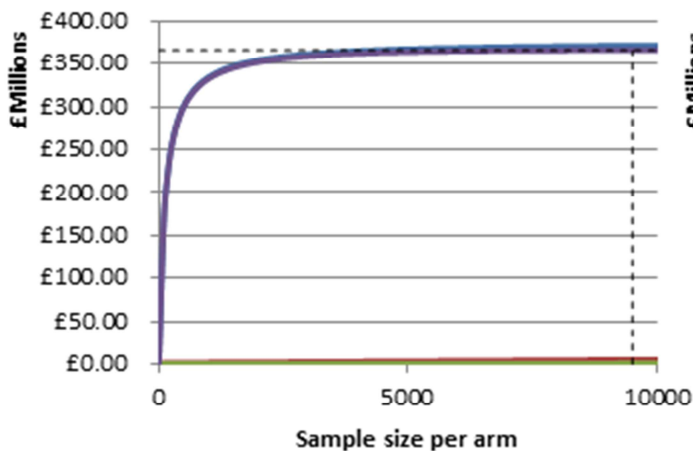


Figure 4-3: EVSI, total cost and opportunity loss, and ENGS

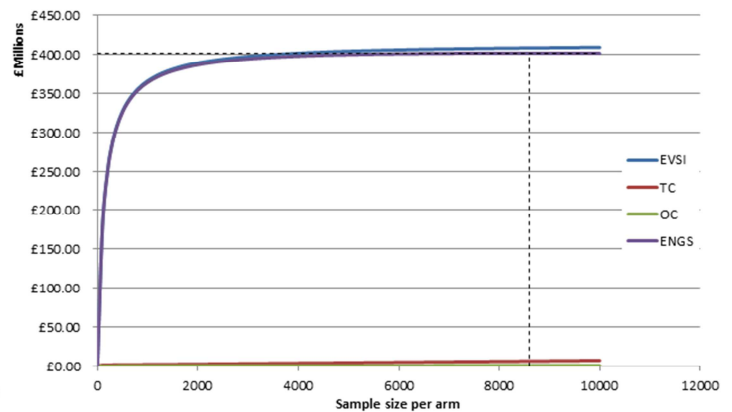


Figure 4-5: ENGS Cost

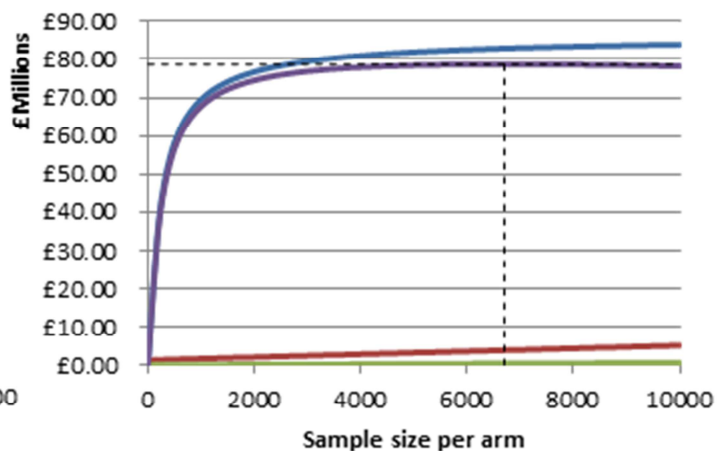


Figure 4-6: ENGS Non-drug Cost

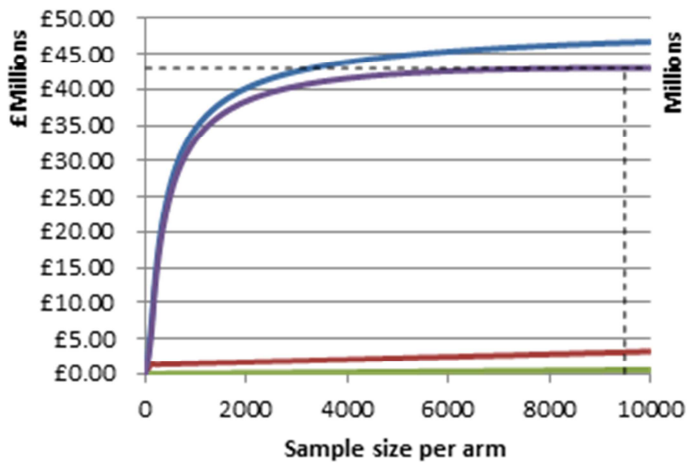


Figure 4-7: ENGS Drug Cost

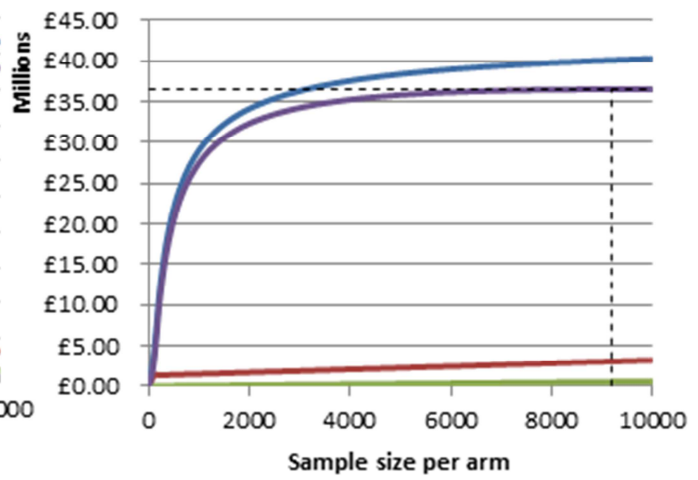
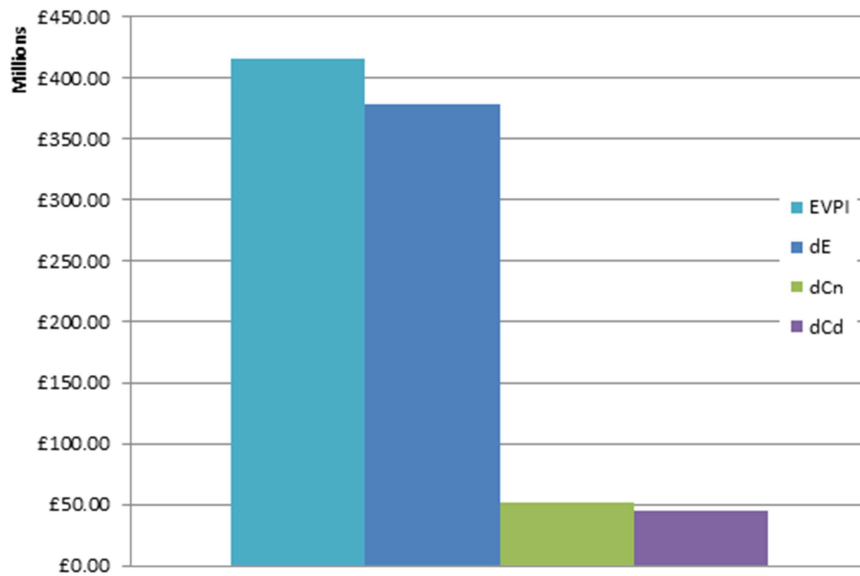
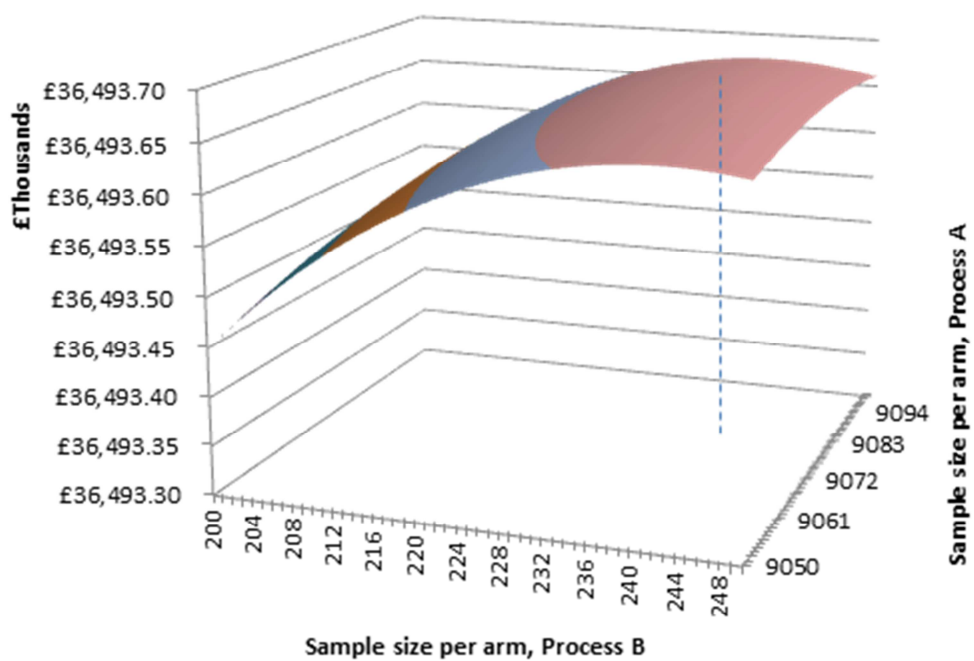


Figure 4-8: EVPI and EVPPI @ $\lambda = \text{£}5,000$



EVPI = Expected value of perfect information; dE = EVPPI, QALYS; dCn = EVPPI, non-drug costs; dCd = EVPPI, drug costs

Figure 4-9: Optimal mix of observations from each data process



4.4. Discussion

4.4.1. Implications of results

The objective of this chapter was to explore a method to answer the question of this thesis, that is, how detailed does a resource data collection process need to be in order to maximise the return on investment in a clinical trial with piggybacked economic evaluation (question (3) posed in Section 4.1). In Sections 4.2 and 4.3 I have shown how value of information analysis can be used to compare the expected return from two different data processes to estimate drug costs, one more ‘detailed’ than the other and thus how it can be used to tackle the study question. I have then combined this with standard Vol analysis to present an overall decision analytic approach to clinical trial design.

The results in the example analysis demonstrate a very high value from eliminating all decision uncertainty (Table 4-3; £416.3m).^{lxiii} This is due to both the high per-patient decision uncertainty (a mean incremental net benefit of £56.41 with standard error £217.15) and the large population who could benefit from this information: asthma is a common disease (N = approximately 6.8m^{lxiv}). There are very few other Vol studies in respiratory disease, but this compares with the EVPI of pharmacogenomic approaches to diagnosing non-small-cell lung cancer of to the US of \$31.4m.²⁵⁰

If a trial were proposed with the objective of estimating incremental net benefit, the optimal sample size would be 8,589 patients per arm. Such a trial is expected to cost £6.3m (plus an opportunity loss of £485,000), but would yield an expected net benefit of £401.9m. This would be a very large trial, and possibly beyond the budget of all but the most major funders. Nevertheless it is the predicted optimal sample size taking into account the cost of acquiring the data and the expected value of the information to the population.^{lxv} Of interest is a comparison with the predicted sample size using a conventional power calculation. Under the assumption that £100 is the minimally important difference in net benefit, I estimate a sample size of 3,700 per arm to be sufficient to exclude chance with 95% confidence if the true difference between the arms is at least £100 (Appendix D).

The EVPPI for each component shows there is more potential value in reducing uncertainty in incremental QALYs than in either of the two cost components (£378.3m for QALYs vs £51.0m for non-drug and £44.4m for drug costs; Figure 4-8 and Table 4-3). Given

^{lxiii} Strictly speaking, this is the value of eliminating decision uncertainty attributable to uncertainty in input parameters, and does not include any model structural uncertainty: the implication is that the model structure is ‘correct’ and certain. This is a weakness in Vol analyses based on a single model.

^{lxiv} Calculated over a 10 year time horizon, discounted at 3.5% per annum.

^{lxv} Although note that this is only true at $\lambda = £5000$.

assumptions about the cost of sampling, I estimate optimal sample sizes of studies collecting only QALYs, non-drug and drug costs at 9,458, 9,456 and 9,197 respectively, yielding ENGS of £366.0m, £42.934m and £36.485m respectively. Interestingly the optimal sample size for a study reporting each component alone is greater than that for one reporting incremental net benefit (8,589). However this is explained by the lower (variable) cost of sampling for the individual studies, thus the ENBS-maximising point (which is where the marginal cost of sampling equals the marginal gain) is at a higher sample size.

Following this, I reached the key analysis for this thesis estimating the expected return on a study of incremental drug costs alone, comparing two alternative approaches to collecting the data. I estimate an optimal mix of 9081 observations using process A and 240 observations using process B. The cost of such a study would be £3.1m, yielding an ENGS of £36.494m. This represents the expected return from a trial collecting data on incremental drug cost alone. By using a mix of both processes, a small increase (of £9000) in the return on investment can be expected compared with using process A alone (rows 5 and 6 of Table 4-3).

As stated, the above analyses represent the expected return on trials measuring QALYs, non-drug costs and drug costs alone. A more realistic scenario would be where all data could be collected within the context of a single trial. Section 4.3.4 brings all the components analysed separately in sections 4.3.1-4.3.3 together to estimate the optimal sample size for each component within one study, including the optimal mix between the two data processes for collecting drug cost data. From this, I estimate that the optimal sample size of a trial would be 10,787 patients per arm collecting QALY data, 4,262 observations on non-drug cost, 3,693 observations on drug cost using process A and 904 observations using process B. The total cost of this trial would be £4.9m with an opportunity loss of £608,000, totalling £5.5m and yielding an ENGS of £403.7m. This is compared with £6.3m for the trial reported in section 4.3.1, measuring all parameters on 8,589 observations, and yielding an ENGS of £401.9m, thus selectivity in which data are collected on which observations in this case reduces the cost of the research and leads to a higher expected net return (of approximately £1.9m).

4.4.2. Practical applications

The analyses above provide an optimal solution to the sample size unconstrained by any research budget, defining the optimum at the point where the marginal gain from an observation is equal to the marginal cost (analogous to the profit maximising condition in

the theory of the firm). At the point of designing a clinical trial, this technique can be used to provide a rational approach to determining what data to collect on which patients. The prior distributions of each of the inputs and research cost estimates are required to do this, information ideally provided by a pilot or feasibility study conducted in preparation for a full trial. Alternatively uncertainty in parameters can be captured via a formal elicitation process.²¹⁸

However, in reality the trialist may be faced with a number of constraints. For example budgets for a research project are usually subject to an upper limit. In order to adapt the analysis to such a constraint, it is simply a question of defining a feasible set of observations on each component such that the cost of sampling is less or equal to the budget, B (Equation [4-23]).

$$k_{SE}n_E + k_{SC_n}n_{C_n} + k_{SC_d^A}n_{C_d^A} + k_{SC_d^B}n_{C_d^B} + K_S I \{n_E > 0 \cup n_{C_n} > 0 \cup n_{C_d^A} > 0 \cup n_{C_d^B} > 0\} \leq B \quad [4-23]$$

The optimal solution is the combination of observations that maximises ENGS subject to the budget constraint and is identified using a simple search algorithm.^{lxvi} Suppose the maximum budget was £2m. In this case the combination of observations on ($n_{\Delta E}$, $n_{\Delta C_n}$, $n_{\Delta C_d^A}$, $n_{\Delta C_d^B}$) is at (2562, 627, 331, 900). That is, 2562 observations on QALYs, 627 on non-drug costs, 331 on drug costs with process A and 900 on drug costs with process B per arm. This trial would cost £1,999,981 and yield an ENBS of £388,168,000. Likewise, it is straightforward to choose the optimal mix from a feasible set where the sample size has already been determined (e.g. via a conventional power calculation based on a clinically important difference in the primary outcome).

4.4.3. Determinants of results of 4.3.3

The optimal mix of observations on two data processes is a function of the relationship between the two (as expressed in the covariance, or equivalently the correlation coefficient) and the relative cost of sampling. Where the data processes are very closely related, i.e. with a correlation coefficient close to 1, then one would expect the inferior process (process B) to be the optimal choice due to the lower cost of sampling: observations on B can be used

^{lxvi} I achieved this simply by running the search algorithm employed to identify the optimal mix of observations as before, except with a flag setting the ENBS to zero if the cost of a particular combination of sample sizes exceeded the budget.

to revise the mean of A simply by adjusting for the prior estimate of bias.^{lxvii} However, where the processes are less closely related there is a trade-off between the extra cost of A and the extra information it yields. Where the processes are completely independent (correlation coefficient of zero), gathering information using B provides no information on A, therefore it would never be efficient to use process B.

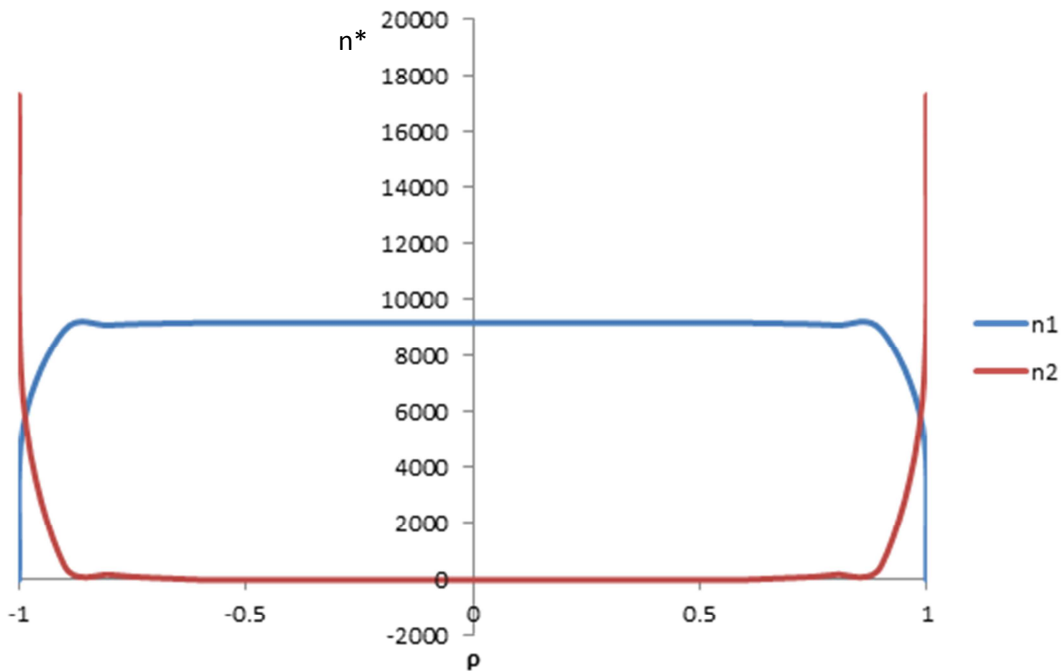
In the dataset used in this chapter, the correlation coefficient between ΔC_d^A and ΔC_d^B is 0.83. Even though the poorer process is only one tenth of the cost of the better process, the information it provides is such that only about 2.6% (240/9321) of the observations should be on this process.

Given this, it is worth investigating how the optimal mix changes with different values of the correlation coefficient. This is shown in Figure 4-10. As predicted, at almost perfect positive or negative correlation, data process B provides equivalent information to process A. As process B is cheaper than process A, it is always preferable to draw observations on that process. As the correlation between the two processes falls, process B provides less information on ΔC_d^A , until the value of the information falls below the marginal cost of sampling at which point it is only worth collecting data using process A.

It should be noted that this analysis treats the correlation coefficient as constant and known. In reality it is a random variable about which additional information could itself be sought. This issue is discussed in more detail in Section 4.4.5 below.

^{lxvii} For example, suppose processes A&B are perfectly correlated and process B always estimated A plus £50. If B is cheaper to measure than A, it would make sense to collect data using process B and subtract £50 from the estimate before plugging this into the equation for incremental net benefit.

Figure 4-10: Optimal mix of observations on each process as a function of rho



4.4.4. Comparison with other studies

The origins of this analysis lie in statistical decision theory, originally developed in the 1960s at Harvard Business School.⁴ However I am aware of only one previous application of value of information principles to help choose study designs (see Chapter 2 for full details of the literature review). Shavit and colleagues²⁰ presented a method to compare the 'net information benefit' of an RCT with an observational study. They define this as a function of the current evidence and estimates of the magnitude of five discrete sources of bias associated with the two designs: representation, selection, time frame, real-life reflection and accuracy of records. Measures of the degrees of each bias were assumed and expressed as percentage deviation from the true mean. Unfortunately the authors did not provide a full numeric example of their method, and thus comparison with our approach is hampered somewhat. However a key difference between the approaches is the question being asked: Shavit and colleagues²⁰ are concerned with the choice between a prospective randomised design and retrospective cohort study to answer a particular decision question. The analysis in this thesis is useful where the decision to pursue a randomised study design has already been made, but the approach to collecting various components of the data is undecided.

4.4.5. Strengths and Weaknesses

This analysis presents a decision analytic approach to clinical trial design, providing not only an answer to the efficient number of observations in the trial, but an efficient number of

observations for each component. Furthermore, it also provides an efficient number of observations on alternative data processes to measure one data component (and thus answers the question posed in this thesis).

However, there are a number of weaknesses and assumptions in the analysis which must be considered.

Firstly, I have referred to the activity of measuring and valuing a particular resource item as a 'data collection process' without explicitly differentiating between measurement and valuation. Failure to distinguish these two distinct processes may have led to some unclarity, however the principles demonstrated are insensitive to this: two data collection processes may vary in how resource use data are collected (e.g. use of medical records vs patient self-report) or by valuation technique (costing individual items vs applying common unit costs to classes of items). The example of this chapter simulates a hybrid of the two: the simpler data process collects data at a more aggregate level (requiring knowledge only of drug class rather than precise brand name of individual chemical), and applies an average unit cost for that drug class based on a representative daily dose.

Secondly, throughout the analyses, I expressed the covariance between parameters as the product of the correlation coefficient and standard errors. This allows the correlation coefficient to be treated as independent from the variance, and potentially, as a further parameter about which information could be sought. However, for simplicity, I assumed the preposterior estimates of the correlation coefficient to be equal to the priors. Addressing this issue would require specifying a distribution for the correlation coefficient. A proposed data collection process collecting information on the correlation coefficient could then be combined with the prior to estimate the preposterior along the same lines as with other parameters. However, this is not so straightforward as the correlation coefficient is unlikely to be normally distributed, thus complicating the analytic solution. Furthermore, it is, of course, not possible to collect data on the correlation coefficient without collecting data on the mean and variances of processes A and B! A simulation approach would be an appropriate method to relax the normality assumptions, and extend the generalizability of the method.

This would be achieved by specifying the prior distributions, then simulating the results of a trial of size n_A and n_B on each parameter, (and where $n_A \neq n_B$, specifying the number of observations where both n_A and n_B are collected in the same patient) by sampling from the joint distribution of the two parameters. This would provide a 'sample' estimate of the

mean and variance of each parameter and covariance between them. This hypothetical information is then used to update the priors to estimate the preposteriors. Repeating many times would build up an empirical distribution of the preposteriors. This would then require repeating over various combinations of sample sizes n_A and n_B .

Related to this point is the assumption regarding sample sizes in the 'combined' analysis (Section 4.3.4). I assumed that the overall sample size of the trial would be the maximum of each individual component, namely 10,787 patients in each arm (on which QALY data would be obtained). Of those 10,787, 4,264 would be chosen at random from which non-drug cost data would be obtained, then a further 4,597 (3,693+904) would be chosen from which drug cost data would be obtained, 3693 of which using process A and 904 using process B. The trial was costed on this basis. However, where patients provide data on more than one component the covariance and hence correlation between those components can be observed and used to revise the prior estimates of the relevant correlation coefficients. My analysis currently ignores this additional information and thus may be overestimating the optimal trial sizes. A simulation approach described above would address this issue.

A key limitation of the analytic approach is the assumption of normality: costs are known to be right skewed, whilst QALYs tend to be left skewed (depending on the patient population). The major advantage of the analytic solution is speed of computation, but at the risk of misleading conclusions should a normal distribution be a poor representation of the data.

A further limitation of this work is exclusion of a number of issues and developments designed to better represent the true value of the information from a particular research project. When originally developed, the scenarios envisaged in value of information analysis related to an industrial setting: the objective function was profit for a firm, and optimal sampling related to quality assurance procedures for an industrial process or market research to inform production decisions.²⁵¹ Thompson first introduced value of information analysis to the health care field,²⁵² whilst Claxton was the first to explicitly apply the technique to address the research (as distinct from the adoption) question,³ however this led to a number of challenges: the information yielded from a particular research project is of value to both present and future population of patients with the disease in question. As discussed in Chapter 3, defining the relevant patient population is not straightforward, with issues such as whether patients enrolled in the trial can benefit from the information or not, cost of reversal of decisions and the opportunity loss of the delay whilst a trial recruits, is

analysed, reported, disseminated and acted upon affecting the expected value of that information.^{199 200}

The final weakness of this analysis is the assumption of a constant marginal cost of recruitment. This may be an oversimplification of the cost function as the first patients are likely to be easier to recruit than the last ones, as stocks of 'willing volunteers' get exhausted, and further effort is required to identify new patients.

4.5. Conclusion

In this chapter I have demonstrated how value of information analysis can be used to guide trial design issues to choose between two data collection processes and thus provide an answer to the question posed in this thesis as to how much detail is required in a resource data collection exercise in an economic evaluation conducted alongside a clinical trial. Rather than directly answering the question as posed, the method estimates the 'optimal mix' of observations on each process, taking into account both the relationship between the two processes (i.e. how good an observation using process B is at predicting the value of an observation using process A), and the relative cost of each. Furthermore I have extended that to show how the principles can be used to determine the efficient number of observations of different outcome measures reported in a trial (namely QALYs, non-drug costs and drug costs). I have shown that such a trial would be less expensive and yield a higher expected net benefit of sampling than one measuring all data on all patients.

There are a number of avenues for extending this research. Firstly I have not considered the possibility of unequal allocation between trial arms.²⁵³ Secondly, whilst I have taken account of the opportunity cost to patients enrolled into the 'wrong' arm of the trial, other aspects such as the opportunity loss to all patients whilst a trial is underway, and the cost of reversing decisions have been excluded.^{144 199} This would be relatively simply addressed through adjustments to the beneficial population and adding in the cost of reversal.

Other enhancements include adopting a more complicated cost function with increasing marginal cost representing increasing difficulty in recruitment as the supply of patients is exhausted, and application of a simulation approach to calculation of the Vol statistics, relaxing the assumptions of normality intrinsic in the analytic approach.

A key area for future work, as previously stated, is in estimating the return on reducing uncertainty in the correlation coefficients between the input parameters. Ignoring the impact on this at present may lead to overestimation of the efficient sample sizes, but nevertheless the analysis of this chapter provides a rational approach to designing a clinical trial, firmly rooted in the decision analytic principles, leading to maximising expected health gain subject to the budget available.

Finally, the analysis presented in this chapter pertains to one example of data collection processes only. Repeating this analysis comparing data processes that vary in terms of method of collection (e.g. questionnaires vs medical records) as well as valuation (costing by individual resource item versus some aggregate approach) would increase familiarity with

the method in the literature, as well as generating new methodological issues requiring further investigation, for example the impact of different response rates from more detailed and laborious questionnaire based tools versus simpler designs, and the impact of more restrictive questionnaires leading to spuriously precise estimates (i.e. underestimating the standard error).

5. Discussion & Further Work

5.1. Summary of Findings

Trialists are faced with many issues when designing data collection methods in clinical trials. This is particularly the case when collecting resource use data due to the multidimensional nature of costs: whereas a clinical outcome may be a measurement on a single scale, a cost is always a function of a number of components such as hospitalisations, prescribed medications, and lost productivity. Furthermore these can be collected at different levels of aggregation.

The precise list of data items that should be collected is traditionally driven by the study question (for example the study perspective will determine whether patient out of pocket costs or lost productivity should be estimated). However there are numerous approaches to costing elements such as hospitalisation: a 'gross' or 'top-down' approach would assign a unit cost per admission or per bed-day. Whereas a 'micro' or 'bottom-up' costing approach would, at the extreme, attempt to measure every individual component of resource use such as minutes of medic and nursing time, exact doses of drugs prescribed and the time required to administer them, exact number and types of diagnostic tests ordered etc.

Again, the appropriate 'level' of costing may be determined by the study question: a comparison of two surgical techniques would *a priori* require a micro-costing approach to the surgery itself, with perhaps a more top-down approach to costing other elements such as length of stay and readmissions.

In this thesis I apply a more systematic method: given the desire to collect data on a particular cost element, what is the added benefit of a more detailed approach compared with its added cost?

My literature review (Chapter 2) attempted to identify whether any previous attempts had been made to examine the cost-effectiveness of one data collection technique compared with another. It also attempted to identify any studies comparing two or more approaches to measuring the same data and any studies exploring the cost of conducting research. My search identified over 100 previous studies comparing different approaches to collecting the same data, suggesting this is an area of importance to researchers (and by implication, to funders of research), as well as a number of studies reporting attempts to minimise the cost of particular research projects. Of note is the success of trialists in more than halving the

cost of a very large prospective study from \$200m to \$95m by rigorous scrutiny of expenditure and modification of the trial design (the Women's Health Trial¹⁶⁴). However, the impact of this on the results of the trial (which I would define here as whether this led to any bias in the results) is unknown.

Another issue of concern evident in the literature was whether enrolling patients in clinical trials led to increased costs for health insurers as a result of protocol driven activity. These appeared to be US based concerns, nevertheless is of issue in the UK too: 'excess treatment costs' must be considered explicitly in clinical trial grant applications.²⁵⁴

Thus comparisons of different data collection techniques and the costs of research are important considerations, according to the literature. It is somewhat surprising therefore that there have been so few attempts to explicitly measure the cost-effectiveness of 'better quality' data techniques compared with 'quick and dirty' approaches. My review identified only nine studies where some concept of economics or cost-effectiveness had been used to assist trial design. Two of these considered the cost-effectiveness of prospective versus retrospective study designs,^{20 140} but only one appeared rooted in the principles of value of information analysis.¹⁹⁵

In Chapter 3 I explored the suitability of value of information analysis as an analytic approach to answer the question of this thesis. I identified a number of limitations of the technique, centring around the application of the general techniques as developed by Raiffa and Schlaiffer²⁵¹ to the specific area of health care decision making.

Particular issues are around aggregating the EVPI and EVSI from an individual patient level to a population level. Specifically, the value of information to society as a whole should be a function of the current and future population who can benefit from the information. Depending on the disease and intervention(s) in question, this may include the prevalent population plus the incident population over an 'appropriate' time horizon (which, by convention is then discounted at the prevailing recommended rate). The definition of an appropriate time horizon, however, is somewhat arbitrary.¹⁹⁸ Furthermore, those patients enrolled in the study may not be able to benefit from the information,³ those enrolled in the 'wrong' arm of a study experience an opportunity cost in terms of foregone health benefit, and there will also be a delay between initiating a new clinical trial and dissemination of results leading ultimately to changes in practice.¹⁹⁹ The population EVSI should therefore be adjusted for these aspects.

Other issues include the costs of reversing decisions, challenging the assumption of independence of the adoption and research decisions (if a decision is not reversible then the value of additional information following that decision is zero; alternatively the cost of reversal should be incorporated into the initial estimates of the value of information),¹⁴⁴ the risk of free-riding leading to a sub-optimal Nash equilibrium (it is cheaper to free-ride and let someone else pay for the research then adapt the results to a local setting, thus no-one is willing to pay for it and the research does not get done), and the optimal allocation of research across multiple jurisdictions.²⁰¹

There are limitations specific to the two main implementations of value of information analysis. The analytic solution can be calculated (almost) instantaneously on any modern computer whereas numeric solutions can require weeks or more of processing time to adequately propagate parameter uncertainty into decision uncertainty.^{lxviii} However the analytic solution requires an assumption of normality in input parameters, whereas the numeric (simulation) approach relaxes these assumptions. It is also more amenable to incorporation of 'all relevant evidence',⁹¹ as it is typically associated with decision modelling (whereas the former is more often conducted alongside single clinical trials). However, due to decision models incorporating evidence from diverse sources, the correlation structure between inputs is unknown and frequently ignored.⁷⁴ This has consequences for value of information analysis, as *a priori* one may expect an assumption of independence in input parameters to overestimate the variance of the output (i.e. incremental net benefit), although such an assumption can actually bias the estimate of variance in either direction, depending both on whether the correlation is positive or negative, and on the structure of the model.^{lxix}

I explored the impact of this assumption on four diverse case studies: three based on clinical trial data and a fourth based on a decision model. My findings were that ignoring correlation between input parameters did not materially affect the expected optimal sample size for future studies in the clinical trial examples, except when input parameters were very highly positively correlated. In the examples considered, such high positive correlation appeared rare. However, in the model based example, the optimal sample size for a study to elicit health state valuations did appear to be sensitive to whether or not parameters

^{lxviii} Parallelisation of the analysis across large arrays of processors will speed the actual time taken, but the processor-hours of computation required remains the same.

^{lxix} For example, where two parameters are summed, the variance of the sum is the sum of the variance of each plus twice the covariance. Where two parameters are subtracted, the variance is the sum of the variance of each minus twice the covariance. Assuming zero covariance will therefore under- (over-)estimate the variance of the sum (difference).

were correlated, although it is not possible to be certain that this is not simply due to random noise from the simulation: in this particular example, the variable cost of sampling was very low. In this situation, the gradient of the EVSI (and by definition, that of the total cost curve) at n^* is very close to zero and so a very small change in the EVSI can have a dramatic effect on n^* .

Having explored the strengths and despite the weaknesses of value of information analysis, I conclude that may well be a viable technique to answer the question posed in this thesis. In Chapter 4 therefore I adapted the techniques, based on the methods of Pratt, Raiffa and Schlaifer⁴ to compare both the added value and added cost of one data collection process with another. This can be considered a special case of a more general question as to the optimal mix of observations using either data process.

The concept is as follows:

- Given two approaches to measuring the same component of incremental net benefit (for example (incremental) drug costs or (incremental) hospitalisation costs), assume one is superior to the other (that is, one is a less biased estimator of the target parameter).
- Define the prior bivariate distribution of the two data processes. This can be estimated from previous data (e.g. a pilot trial where both data processes were observed) or elicited from experts (see Chapter 3.3.4, 'Potential Solution' for a discussion of one approach to elicitation). Either way it represents the best estimate or 'current knowledge' about the relationship between the two data processes.
- The estimate of the target parameter using the 'better quality' process will always be used to estimate incremental net benefit.
- Gathering information using either data process can be used to update the estimate of the mean and variance of the better quality process via the prior bivariate distribution.
- Thus the predicted posterior mean and variance of incremental net benefit from (n_1, n_2) observations drawn using processes 1 and 2 respectively can be calculated and hence the expected value of the information estimated. Comparing this with the cost provides an estimate of the ENGS. The optimal mix is (n_1, n_2) that maximises the ENGS.

Applying this to an example dataset, I find it is possible to observe a small increase in expected efficiency (as measured by the ENGS) of a trial by employing a mix of observations using both data processes, rather than just the 'superior' dataset.^{lxx}

The approach successfully extends the principles of value of information from identifying what parameters on which it is efficient to gather more information (and the efficient sample size thereof), to the efficient method of collecting data on those parameters. In other words, my thesis extends value of information analysis from *what* data to collect in a clinical trial designed to inform economic analysis to *how* to collect it.

^{lxx} Although it should be noted that due to the nature of the data, this is only true at the artificially low valuation of £5000 per QALY.

5.2. Discussion Points

In this section I bring together a number of points arising during development of, and as a consequence of, this work. I firstly reiterate the strengths and weaknesses of the method described in Chapter 4, before reflecting on some broader issues.

5.2.1. Strengths and weaknesses of the method

Key limitations/simplifications of the method as described in Chapter 4 are the assumption of a constant correlation coefficient, assumptions regarding the overall sample size of a future trial, assumptions that all parameters are normally distributed, exclusion of other developments in value of information analysis that have occurred alongside this work, and assumption of a constant marginal cost of recruitment.

Perhaps the key limitation is the assumption of a constant correlation coefficient between estimates of cost derived from the two alternative data processes. Conceptually, there is no reason that this cannot be treated as a random variable about which additional information could be sought. As stated in Chapter 4, a complicating factor is that gathering information on the correlation coefficient requires collection of data using both processes, which can itself then be used to revise the marginal distributions of each.

Related to this point is the assumption I made with regard to the overall optimal design of a trial, where the total number of recruits to each arm was assumed to be the maximum of observations on each individual component (10,787 in the example), with other parameters collected on a subset of the 10,787. Where more than one parameter is observed in the same patient, information on the correlation between them becomes available.

I have explored whether it is possible to incorporate the added information gained from observing more than one parameter in the same patient in the analysis. The base case analysis comparing 'process A' with 'process B' for drug costs assumes that n_A observations are collected using process A and n_B with process B, and that there are no observations where both processes are collected. In some circumstances this may be plausible (it could be argued that we are ultimately only interested in C_d^A , therefore why collect C_d^B on some patients for which we are already collecting C_d^A ?). However, a more realistic scenario would be where data are collected using the inferior process (process B) on all patients, with process A being collected on a subset of those in order to verify the quality of the data with process B. For example consider the case where a study proposes to collect primary care contacts by questionnaire thus relying on patient recall, but with the medical records of a sample, say 10%, accessed at random to estimate the degree of any recall bias.

I attempted to incorporate this by specifying both n_A , n_B as well as the ‘overlap’, $n_{A \cap B}$. The overlap reduced the prior covariance by a proportionate amount as per Box D-1 of Appendix D. However, this led to an unanticipated outcome: the predicted efficient scenario was to collect data on all patients using process A, with process B being collected on a subset of those (see Appendix D). This is counter-intuitive as it would be illogical to collect the ‘ideal’ data on all patients, and then some inferior data on a subsample of those: the optimal solution would have to be the opposite way around.

Investigating this further showed that the simple method of incorporating $n_{A \cap B}$ overestimates the reduction in variance associated with a given sample of observations n_A , n_B . In some cases this leads to a negative preposterior variance, and thus an error when computing the EVSI (see Appendix D). A more thorough investigation of this is required and would include exploration of why and how the approach was overestimating the reduction in preposterior variance, and development of the mathematics to adjust for this and correctly estimate the Vol statistics.

This aspect should be incorporated into the analysis, but as the paragraph above shows, doing so appropriately is not necessarily straightforward. Accommodation of non-normal distributions (such as that of the correlation coefficient) should also be considered, for which a numeric (simulation) solution would be appropriate.

The other parallel developments in Vol to which I allude above focus primarily on the appropriate definition of the beneficial patient population. For example, *inter alia*, taking into account the delay before the results of a proposed trial become disseminated and lead to a change of practice.¹⁹⁹ A related aspect I have not considered in this thesis is the ‘value of implementation’, where consideration is given to the value of implementing current best practice as distinct from conducting new research to reduce decision uncertainty.¹²⁵

Finally, the assumption of a constant marginal cost of recruitment may be an oversimplification. Often, a principal investigator in a trial will have a ready pool of ‘willing volunteers’ amongst his or her own patients. The cost of recruiting these is therefore low. However, as this pool gets exhausted, additional search efforts are required to secure the marginal enrollee. This may be simply a case of publicising the trial more widely. Alternatively, consideration may be given to adding additional recruitment centres. The latter would suggest a step function shape to the cost function.

5.2.2. Broader issues

Acceptability of the framework to decision makers

The acceptability to decision makers of the analysis demonstrated in this thesis is contingent on acceptance of the iterative approach to decision making described in Chapter 1,^{193 95} and the assumptions underpinning economic evaluation. Specific issues are the requirement for risk neutrality, and the application of standard techniques of discounting future costs and outcomes and the determination of an ‘appropriate’ time horizon over which to calculate value of information statistics.

Arrow and Lind argued that public sector decision makers should be risk neutral,⁹⁶ but the reality is that decision makers are extremely risk averse: licensing bodies such as the FDA and EMA require a reasonable probability of excluding chance as an explanation for observed benefits or harms before granting marketing authorisations for new chemical entities,^{lxxi} and anecdotal evidence based on my personal experience with local decision makers in the NHS suggests that they too are risk averse. This may thus be a barrier to acceptance of the methods. However, it may be possible to adjust the analyses to incorporate risk aversion (see ‘further work’ below).

Notwithstanding the above, perhaps the major limitation of value of information analysis is the definition of the appropriate population of patients who can benefit from a given research project.¹⁹⁸ I discussed this in detail in Chapter 1: the *de facto* standard appears to be ten years but there is no logic behind this. This is of concern as the population Vol statistics are extremely sensitive to the assumed time horizon. Further research into the appropriate time horizon for a value of information analysis would therefore be of use.

Further challenges to the framework are more fundamental: economic evaluation is based within the extra- or non-welfarist paradigm, where health is the only maximand and thus information is only of value in terms of its ability to increase health.

In their analysis of the relative value of alternative study designs, Shavit et al.²⁰ included two thresholds: one representing a willingness to pay for health and the other for information. Correspondence following publication of the manuscript challenged this.¹⁹⁴ The correspondent (Grandjour) suggested that the same threshold be used for both research and adoption decisions. This is logical where the purpose of information is solely to reduce decision uncertainty and thus increase the expected health gain of the decision. However in

^{lxxi} That is, findings must be statistically significant, usually defined as a p-value of less than or equal to 0.05.

response Shavit and colleagues disagreed,¹⁹⁵ arguing that the societal willingness to pay for information may well differ from that it is willing to pay for health gain, the reason being that information for its own sake may have consumption qualities (desire to feel ‘in control’) as well as production qualities (use in reducing decision uncertainty in health care).

These ideas are shared by Cohen and colleagues¹⁴¹ where they state that “research outputs are multidimensional”. In a strict extra- or non-welfarist approach then this argument would be rejected: the output, or rather value, of a research project is only in terms of whether it increases expected health gain. However as discussed in Chapter 1, the use of the extra- / non-welfarist approach is partly for pragmatic reasons: for the sake of simplicity it may be better to conduct economic evaluations and value of information analyses within the extra-welfarist framework and allow the other elements of value to enter into subsequent discussion and interpretation of the results. This is consistent with the definition of economic evaluation being used to *inform* decisions rather than *make* those decisions.⁴⁸

Trials vs decision models for decision making

A general issue raised by the iterative approach to decision making, and specifically the use of decision models drawing on many different sources, is whether conducting ‘piggybacked’ economic evaluations alongside clinical trials is ever an appropriate means to estimate the cost-effectiveness of interventions. Furthermore, value of information analysis simply generates a list of parameters about which it is worthwhile eliciting more information, and the optimal sample size for each: it does not dictate the nature of the study.

This suggests that a full RCT with piggybacked economic evaluation may not be the most efficient study design, and that a ‘piecemeal’ approach, gathering information on one parameter at a time may be preferable. The limitations of the piecemeal approach are both methodological and ethical. Methodologically, the piecemeal approach means that correlation between the input parameters cannot be observed. Furthermore there may be incompatibilities between data sources (e.g. different patient populations), leading to risk of bias. From an ethical perspective it may not be considered appropriate to (for example) conduct a randomised study measuring only cost where the benefits of a treatment are already established.

In 1974, Williams¹⁹⁶ suggested that economic evaluations^{lxxii} only be conducted under a number of specific situations, for example in decisions involving a large movement of

^{lxxii} Williams referred only to cost-benefit analysis as this work predated much of the development of other types of economic evaluation.

resources, where responsibility is fragmented, and the alternative courses of action are radically different. The justification for this is the research expense of such studies. Since then, there is greater acceptance that knowledge of not only the benefits of a treatment are necessary to inform decision making, but also the opportunity cost to other patients in the system (most clearly expressed in the incremental net (health) benefit, derived from the ICER.^{lxxiii}

Other potential approaches to tackling the study question

My starting point in this thesis was that value of information analysis, a method for prioritising future research, has the potential to provide an appropriate solution to the question. However other approaches to research prioritisation exist, a number of which were identified and reviewed by Fleurence and Torgerson.⁹⁹ These are:

- Informal ‘gut feeling’
- Burden of disease^{255 256}
- Welfare loss from clinical practice variations²⁵⁷
- Trial sequential analysis²⁵⁸
- Payback²⁵⁹⁻²⁶¹

Fleurence and Torgerson⁹⁹ argued that a research prioritisation mechanism should be consistent with the objective of the health system within which it operates; assumed to be the maximisation of health gain, subject to the budget and equity considerations. The approaches identified can be broadly divided into ‘non-economic’ and ‘economic’ approaches.^{lxxiv}

Non-economic approaches

‘Gut feeling’ is perhaps the most commonly used means to prioritise research projects. However, due to its subjectivity, lack of replicability, and lack of explicit quantification of the value of research, it is unknown whether it will be consistent with the health system objective stated above.

Burden of disease approaches^{255 256} assume a direct link between the size of a problem and the value of research into it. However, this need not necessarily be the case: there may be very little uncertainty regarding the effectiveness of a treatment for a highly burdensome

^{lxxiii} The existence of bodies such as NICE provides evidence to support this claim.

^{lxxiv} The division between economic and non-economic approaches to research prioritisation is a concept introduced by Mitton & Donaldson in a different context.²⁶²

disease. Research into that treatment would have very little chance of changing policy and hence very little impact on health gain: resources may be spent to greater effect through dissemination of current knowledge.

Priority setting according to the estimated welfare loss from clinical variations in practice argues that the reason for variation in practice (after adjustment for casemix, demographics etc) is disagreement as to the effectiveness of interventions. However, this approach does not distinguish between variation due to uncertainty in the effectiveness of a treatment and that due to lack of dissemination of existing knowledge. Under this approach, a further trial could be recommended even where there was little uncertainty regarding the effectiveness of a treatment. Under these circumstances there are probably more efficient means of changing practice than undertaking a new RCT.⁹⁹

Finally, trial sequential analysis has been proposed as a means of establishing when sufficient evidence has been gathered.²⁵⁸ This approach is an extension of group sequential analysis,^{263 264} a means of adjusting significance levels for repeated interim tests. As this does not take into account the opportunity cost of conducting research, it cannot assess the relative value of investments in alternative research projects.

In summary, the limitations of the 'non-economic' approaches are that they are unlikely to be consistent with the objectives of the health system to which they contribute, and that they do not consider the issue of the opportunity cost of research: namely, that resources invested in research may have generated more health for the population had they been spent either in alternative research projects, or in direct care provision.

Economics-based approaches

The 'payback' approach involves the use of scenario analysis to estimate the likely cost-effectiveness of a proposed trial. As with the 'clinical variations' approach, this is a function not only of the results of the trial, but also any change in policy resulting from that trial. That is, following publication and dissemination of a trial result, clinical practice may change, resulting in a change in costs and health outcomes in the population. The expected change in cost (including the cost of conducting the trial itself) divided by the expected change in QALYs gives the expected cost-effectiveness of the trial, expressed in terms of incremental cost per QALY gained. Expression of the results in terms of this commonly used metric allows direct comparison with the cost-effectiveness of other health care interventions, and/or the expected cost-effectiveness of further trials of other health care interventions (and application of the same willingness-to-pay thresholds).

Townsend and Buxton²⁶⁰ used this approach to determine the likely cost-effectiveness of a trial of the long term effects of Hormone Replacement Therapy (HRT). They hypothesised three possible outcomes from the trial:

- Positive, meaning long-term cardio-vascular (CV) and fracture benefits and breast cancer risk were in accordance with the currently available data (1996);
- Negative, meaning long-term CV and fracture benefits were 25% and 50% of those currently thought, with breast cancer risk twice that of the positive outcome; or
- Inconclusive, meaning breast cancer risk was as per the positive outcome, and CV and fracture risks as per negative outcome.

They then assumed likely policy changes of +50% uptake of HRT in the target population given a positive result, limitation of the use of HRT to short-term relief only given a negative result, and no change with the inconclusive result. The total change in cost and QALYs gained in the population and incremental cost-effectiveness ratio (ICER) from each of the three scenarios compared with current practice was then calculated (including the cost of the trial itself).

The final step was to calculate a weighted average of the three ICERs, based on an assumed likelihood of each (the authors stated a weighting of 0.5, 0.25 and 0.25 for each scenario respectively as the 'most plausible', but presented results for alternative weightings too). Based on this, they estimated an incremental cost of £1153 per QALY gained from conducting the trial. Given a typical UK threshold of £20 - £30,000,⁷² this would be considered a good value for money investment.

This 'payback' approach is useful in that it estimates the likely cost-effectiveness based on policy/practice change and across an entire population. It also explicitly recognises the opportunity cost of an action: resources allocated to research cannot be allocated to treatment.

The method has a lot in common with value of information analysis. However, it does have a number of limitations. The key limitation is that the outcome of the trial is divided into a discrete number of scenarios, which do not represent all possible outcomes (for example, suppose the results of the trial illustrated above had shown greater than expected CV benefits, but worse than expected fracture benefits?). To adequately take into account all possible outcomes and their likely policy responses could easily become burdensome. Thus

whilst payback considers a number of discrete outcome scenarios, value of information analysis can be thought of as an extension of this considering all possible outcomes on a continuous scale. Secondly, the incorporation of prior knowledge into the decision is undertaken informally in terms of a subjective estimate of the likelihood of each of the three outcomes, rather than as an explicit combination of prior knowledge and the new information into a posterior belief. Finally, the role of systematic review and meta-analysis within this framework is unclear as policy changes are determined based on the result of the one 'definitive' trial, rather than a synthesis of all available evidence.

On this basis, I decided that value of information analysis provided the most appropriate research prioritisation technique upon which to base this thesis.

5.3. Implications for practice

In this thesis, I bring an extension of value of information analysis into the healthcare sector for the first time, namely, extension of the principles to consideration of alternative approaches to gathering the same data in an economic evaluation alongside a clinical trial.

This work sits firmly within the iterative decision framework outlined in Chapter 1. If followed, the framework will lead to maximisation of expected health gain subject to the resources available. The novelty of this thesis is that it adds an additional dimension in calculating the optimal mix between two data collection processes.

Whilst ideally the entire framework should be used to inform adoption and research decisions, the analyses presented can be used to assist the design of clinical trials either in their entirety or, where some trial aspects have already been determined (for example, where the total sample size has been determined by a power calculation on the primary outcome), to assist subsequent decisions as to what resource use (and indeed outcomes) data it is worth collecting, and on how many patients.

Such a scenario where it can be useful is where there is a desire to, for example, validate self-reported resource use data against medical records. Whereas 'gut feeling' may suggest a figure of 10% is adequate for this purpose, the methods presented in Chapter 4 can be used to estimate an efficient mix between self-reported and medical record data.

This method would fit into the 'ideal' overall trial design process as follows:

1. Systematic review and economic evaluation based on current evidence, with value of information analysis determining whether future research into the research question is worthwhile.
2. Feasibility / pilot studies
3. Full scale RCT
4. Updating of systematic review, economic evaluation and value of information analysis
5. Reiteration of the entire process.

Step one would comprise a number of sub-steps: firstly the overall expected value of perfect information would be calculated. This would be the EVPI of eliminating all uncertainty in incremental net benefit. If this was 'sufficiently large' (e.g. greater than the

reasonable minimum cost of any useful evidence collection exercise^{lxxv}) then the EVPI of various components of the decision question would be estimated separately (e.g. treatment effect, costs, and other relevant components such as health state utilities and longer term prognosis). The EVSI and ENGS of these components should also be estimated to determine the optimal sample sizes for either a series of mini-studies or one study collecting all the relevant data (as per Section 4.3.4).

On this basis, a feasibility study may be proposed, the purpose of which is to answer the question, 'Can this study be done?'.⁵⁵ The study will assess measures such as the standard deviation of the primary outcome to inform traditional sample size calculations, willingness of patients and clinicians to take part in the study, and usefulness of the proposed measures including designing appropriate questionnaires, e.g. resource use data collection tools.^{265 266}

This feasibility stage would therefore be the opportunity to test alternative data collection measures, for example medical records versus patient questionnaires, or more detailed vs less detailed prescriptions data. In each case, both data collection tools would be applied to all patients, thus providing data on the bias of one tool compared with the other, the variance of each and crucially the correlation between them.

At this point the information yielded from the feasibility study can be used to inform the design of a full scale trial, the method proposed in Chapter 4 of this thesis informing the decision as to which data collection process to use on which proportion of patients enrolled.

The results of the trial would then be used to update the systematic review and the process is repeated.

The iterative approach described above is, of course, an ideal rather than reality, and the exact process may not be followed completely. The case for conducting a systematic review of current evidence prior to undertaking any primary data collection process (e.g. a trial) is extremely compelling in order to avoid unnecessary duplication of research.^{267 268,lxxvi} Incorporating the results of this into an economic evaluation and value of information analysis provides a stronger basis for deciding to pursue a particular piece of research.

However, from the perspective of a researcher interested in conducting a clinical trial with a concurrent economic evaluation, they are likely to begin with a review of the current

^{lxxv} A ball-park estimate may be around £50,000, representing the cost of a few months' researcher time analysing a database and writing up the results and including overheads (termed 'indirect costs' in management accounting).

^{lxxvi} Notwithstanding a fundamental tenet of science being replication and reproduction others' results.²⁶⁷

evidence, plan a feasibility and/or pilot study, before embarking on the full scale trial. Where a full decision model has not been developed prior to the feasibility study, costs and outcomes estimated from the feasibility study could be used as priors to estimate the value of information statistics, and the resulting value of using one data collection method compared with another.

My analysis is based within a UK context, where the economic evaluation is conducted from the perspective of the NHS, and the research funder is assumed to be the NHS (or rather, the National Institute for Health Research (NIHR), funding for which is top-sliced from NHS allocations). There is therefore a clear link between the opportunity costs of research and treatment: funds not spend by the NIHR can, in principle, be diverted directly to patient care (or, more likely, to other research projects). However this link is not essential: the approach is of most relevance to any funder interested in the efficient use of finite research funds, and thus the analytic techniques are very transferable to settings outside the NHS. The only adaptations required would be those necessary in any economic evaluation, for example, adjustment of the cost perspective allowing for patient out of pocket costs or not as appropriate.

These techniques also have potential outside of public research funding bodies, for example, a pharmaceutical firm undertaking an economic evaluation alongside a clinical trial to inform reimbursement decisions. From the firm's perspective, its objective is to maximise profit rather than QALYs. In order to extend the framework to incorporate this, the firm's profit is defined as a function of price, sales, and manufacturing and development costs. Sales are a function of the probability of acceptance by a reimbursement agency, where the probability of acceptance is a function of the ICER, itself a function of price. The same framework can then be employed to predict which research investments yield the greatest expected profit.

5.4. Further work

The limitations of the analysis described in Chapter 4 provide a framework for future research, relating both to methodological and applied aspects. These are summarised below.

5.4.1. Methodological issues

A key methodological area requiring deeper investigation is incorporation of uncertainty in the correlation coefficient (ρ) between input parameters. As suggested above this is not necessarily straightforward. The method described in Chapter 4 assumes ρ is known with certainty. However, this will not be the case: it is itself a random variable about which further information could be acquired. Incorporating uncertainty in ρ would require assigning a probability distribution between its logical limits of -1 to +1, to which a normal distribution may be a poor fit. Non-normality in the distribution of the correlation coefficient can best be accommodated by adopting a numeric (simulation) approach to calculating the value of information statistics. Comparisons of the analytic and numeric approaches would also be of interest, providing a measure of the sensitivity of the results to distributional assumptions.

An informative prior distribution of the correlation coefficient may be obtainable from pilot or feasibility study data. However, as discussed in Chapter 3, where no data are available, expert opinion may form a suitable alternative. Formal methods exist for eliciting informative priors from a panel of experts,²⁴⁷ but eliciting multivariate distributions and correlation coefficients is somewhat challenging. As discussed in Chapter 3, O'Hagan²⁴⁶ suggests restructuring the problem as a means to tackle this. For example, suppose the target distribution is a bivariate distribution between the response rate to two drugs and it is reasonable to propose that patients who perform well on one drug may also perform well on the other, thus the parameters are correlated. Eliciting the bivariate distribution directly would be cognitively challenging. However, restructuring the problem into two stages where the response rate to one drug is first elicited, followed by the relative risk of response with the other, may be more feasible as the two parameters elicited (baseline response and relative risk) can reasonably be assumed independent from each other.

Adapting this principle to eliciting the correlation coefficient would follow the same process. Suppose the aim was to estimate the relationship between two measures of the cost of hospitalisations. Restructuring would require eliciting the expected cost of hospitalisations using the 'ideal' data collection process (as well as a measure of the uncertainty around that

mean), followed by an elicitation of the proportionate bias of an inferior measurement approach (and a measure of uncertainty around that mean). For example it may be believed that an inferior measurement process will overestimate the 'true' cost of a hospitalisation by between 5% and 15%.

A key element in any elicitation process is determining who the experts are whose views are to be elicited. In this case, appropriate experts would be trialists with experience in collecting similar data in alternative ways, or financial managers of hospitals who may also have experience with alternative approaches to estimating the same data.

Exploration of the feasibility of this to inform prior correlation coefficients would be an interesting methodological extension to this thesis. Alternatively, in the absence of an informative prior, a sensitivity analysis could be conducted on the VoI statistics as a function of the correlation coefficient as illustrated in Chapter 3, Section 3.3.4 and Figures 3-14 to 3-16.

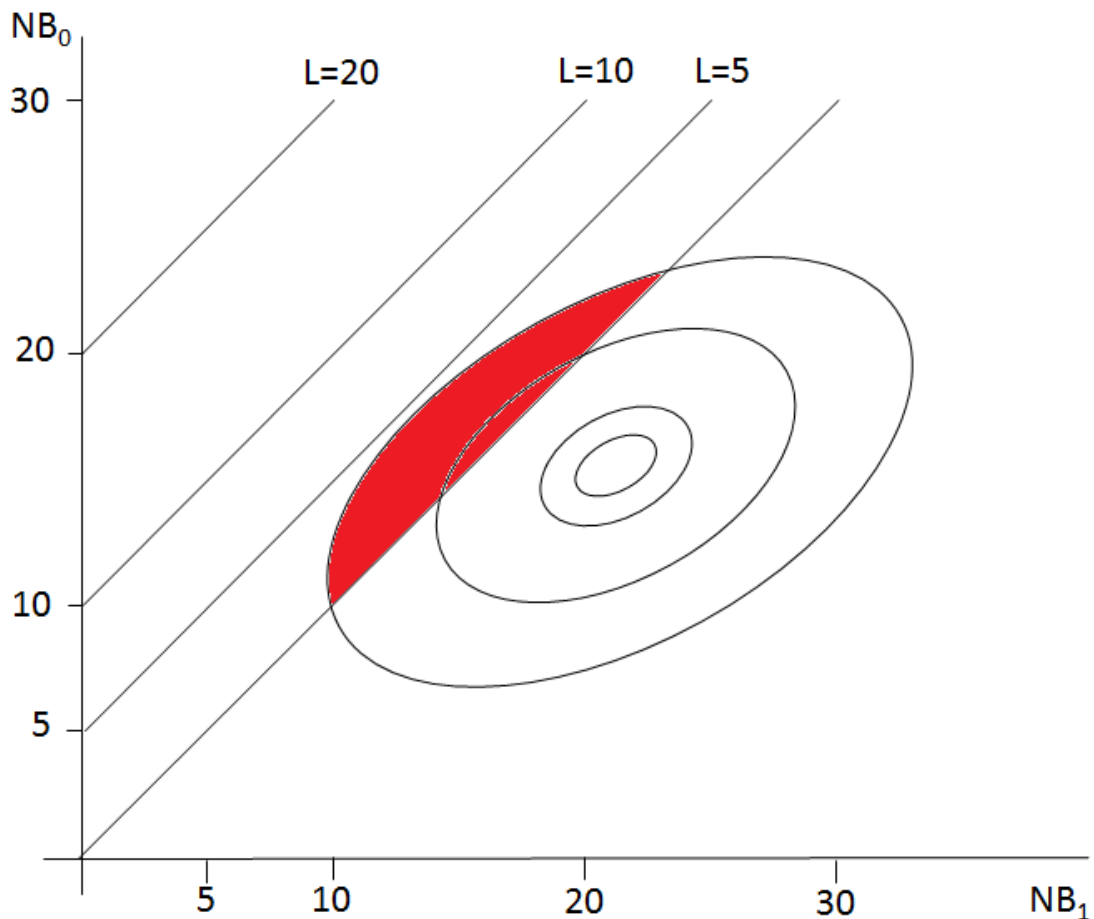
Another methodological consideration is the use of more sophisticated cost functions. Above, I suggest a step function may be an appropriate model where additional centres are required in order to boost recruitment. Whilst the exact cost function for a trial is an empirical matter, the sensitivity of the VoI results to alternative functional forms would provide additional insight into the robustness of the method.

For simplicity, I limited this analysis to a case where there were only two comparators. In order to fully inform decisions and avoid spurious comparisons and misleading conclusions, economic evaluations should consider all relevant comparators simultaneously. Therefore a further avenue for future work would be extension of these principles to multiple comparators. The analytic approach to calculating value of information statistics is somewhat complicated by the presence of multiple comparators as the objective function is now an n-dimension maximisation problem where n is the number of comparators, instead of a simple uni-dimensional problem of choosing the option which maximises expected incremental net benefit. Each comparator will have a mean and standard error of net benefit and these will not be independent of one another.

The easiest way of considering the multidimensional problem is to consider the bivariate case where instead of calculating the incremental net benefit and choosing the comparator that maximises this, the bivariate distribution is analysed directly. The decision rule is to choose the option with the maximum (expected) net benefit. In Figure 5-1, the peak of the

bivariate distribution is at a value of NB_0 of 15 and NB_1 of 22. Therefore the decision would be to adopt option 1. However the probability of being 'wrong' is equal to the probability that NB_0 is greater than NB_1 . This is the proportion of the probability mass to the north west of a 45° line. The loss function is now a plane: to the South East of the 45° line, this is a flat surface at a value of zero. To the North West of the 45° line it has a slope of 45° towards the North West: a series of 'iso-loss' lines plotted in the Figure illustrate this. The expected loss is then the value of the loss plane at every point multiplied by the probability of observing that point, i.e. the 'height' of the density function. This can be approximated numerically. However, generalising the unit normal linear loss to a multidimensional loss 'plane' would allow a convenient analytic solution to Vol along the lines described Chapter 1.

Figure 5-1: schematic of bivariate distribution of net benefit with two comparators



Another aspect I did not consider was the possibility of unequal allocation of observations between arms: *a priori*, where the standard error of net benefit (or some component of net benefit) in one arm is different from that in the other(s), it may be more efficient to allocate more patients to the arm with the greater uncertainty. Alternatively, where research costs differ between arms it may be more efficient to randomise more patients to the cheaper

arm.²⁶⁹ This has been considered in the context of standard value of information analysis,²⁵³ but this would be an interesting addition to the methods presented in this thesis.

An implicit assumption within the analytic framework is that information is only of value to the specific decision question posed. However, information is a public good and may give rise to positive externalities where that information can be used to inform other decisions. This may lead to a situation where a research project would not be efficient when two decision questions are considered separately, but when the value of the information to both questions is considered simultaneously, that research may be efficient. Future work should focus on whether this is theoretically possible, and if so, whether it can be demonstrated empirically. Demonstrating this theoretically would be relatively straightforward: two decision problems would be simulated where some evidence is common to both (for example the incidence or cost of a hospitalisation for a particular side effect from a drug used in more than one disease). By changing point estimates and standard errors of inputs and costs of research, the analysis would explore whether it is possible to achieve a situation where for each individual situation, the research is not worthwhile, but when the external value of the research in one disease area is incorporated (i.e. the benefit to the other decision question), the research does become worthwhile.

Empirical investigations could involve retrospective identification of areas where the same research is of value to multiple decision questions, calculation of optimal sample sizes and hence predicted efficient research plans, and observation of actual research conducted. A prospective approach would involve identifying parallel research programmes where there is the potential for 'mutual externalities', and conducting analyses to demonstrate efficient allocation of the research between the two programmes.

In their paper and subsequent correspondence, Shavit and colleagues^{20 194 195} raised the issue of whether there were separate thresholds for health gain and for research. Taking an extra-welfarist approach, it would be argued that the only maximand is health and therefore assigning different values to health gain from a research programme versus a health care intervention would be illogical. However, a pure welfarist approach would take the view that the information yielded from a research study may have consumption qualities in itself, and is therefore of intrinsic value, therefore there is no reason why the willingness to pay for a given quantity of information would be the same as the willingness to pay for the health gain it achieves alone.

Complicating this comparison is the public good nature of information, and the resulting externalities mentioned above: a research project may have (health) benefits to patients other than those with the disease and treatment(s) in question. This however is a separate issue to what I am concerned with here. A useful future empirical study would estimate whether there is indeed a different willingness to pay for information, over and above the value of the health gain it brings about. Care would be required to account for the externalities, that is, to ensure that the study is detecting any consumption value of information itself, and not just the value of health gain to other disease areas. Previous studies have attempted to estimate an appropriate threshold value for QALY based on willingness to pay⁵⁹ and the opportunity cost within the health sector.⁷³ The proposed study here would build on and complement these.

5.4.2. Applied issues

I provide only one example illustrating the method to decide the optimal mix between two data collection processes. To a certain extent this is quasi-hypothetical as it is based on previously collected data adapted to this purpose. As a result, rather than two separate data collection processes, I simulated the 'inferior' process B by applying a more approximate costing method to drug costs. I also did not specifically distinguish between different methods to both measure and value resource use, combining the two together in the catch-all term 'data collection process'.

Further empirical work should therefore focus on different approaches to measurement of resource use, for example comparing questionnaire data with medical records, interviewer-versus self-administered questionnaires, or alternative questionnaire designs. The latter would be of particular interest as a more complicated questionnaire may well yield 'better quality' information, but also be more onerous to complete thus affecting the response rate,¹⁷⁵ and hence estimates of standard error. Issues relating to bias may be relevant depending on whether or not data were missing at random. Evans and Crawford, in their review of the validity of different approaches to collecting prospective economic data,¹⁷¹ distinguished three areas for analysis: the effect of recall bias over different time periods, proxy completion of questionnaires and the mode of administration. This would provide a useful framework to explore the issue.

A recent systematic review looked at studies comparing the effect of questionnaire design on recall of pharmacological treatments.¹⁷⁵ Amongst the eight studies included in the review, the reported prevalence of drug use varied between 5% and 40% when prompts or

memory aids were provided as a part of the questionnaire (e.g. provision of drug names or pictures), suggesting this too could be a useful element to explore.

In Chapter 1, I stated that despite commissioning exploratory studies in VoI, neither NICE nor HTA formally adopted the principles in order to assist their prioritisation decisions. A major part of this was simple lack of comprehension of the technique.¹²⁶ Whilst education in the principles of economics would assist this, a qualitative research project to investigate the barriers to acceptance would be of interest.

It is possible that risk aversion may be a reason for failure to adopt these techniques (the outcomes of the qualitative research would hint at whether this was indeed a factor). Experiments to elicit the degree of risk aversion amongst decision makers (either at a local or national level) would be of value. Adjusting results of value of information analyses to take into account risk aversion may then lead to more acceptable conclusions and thus potential adoption of the techniques. Such a technique has been attempted to adjust QALY gains for risk aversion in a decision model,²⁷⁰ but I am not aware of any attempts to incorporate this into value of information analysis.

The three trial-based case studies explored in chapter 3 yielded very large and high cost trials, any one of which would likely consume the entire budget of a typical funding round. In section 3.3.4 I raised the possibility of diminishing marginal returns to investment as well the possibility of attaching an intrinsic value to funding a broad range of projects rather than just the most efficient one. Work disentangling these elements: risk aversion, diminishing marginal returns to investment and any intrinsic valuation of diversity would therefore be of value.

Risk aversion could be detected by exploring whether or not funding panels had any systematic preference between studies with identical ENGS, but one would lead to a much greater reduction in decision uncertainty. For example, a trial for a rare disease resulting in a large decrease in decision uncertainty and another for a large disease resulting in a small decrease in decision uncertainty.

Diminishing marginal returns to investment and identification of any intrinsic value for diversity could be estimated through discrete choice experiments, presenting panels with alternative baskets of research projects and observing their preferences as the variety and number of projects are varied (all totalling the same expenditure).

Finally, research exploring an appropriate time horizon for Vol studies would provide a firmer foundation for analyses. A logical time horizon would be one over which the decision question remains relevant. This could be interpreted as the time before the next therapeutic advance renders the current decision question redundant. Empirical research attempting to estimate or predict this time period would therefore be of value.

5.5. Conclusion

In this thesis, I sought an answer to the question of how much detail is required in an economic evaluation alongside a clinical trial to optimise evidence for decision making. I showed how the principles of value of information analysis could be extended to determine the optimal mix between two alternative data collection processes, one top-down and the other bottom-up. In the example considered, substituting the superior process for the inferior process in a few observations led to a small increase in expected return from the trial (as expressed in the expected net benefit of sampling).

The analysis was limited to a retrospective re-interpretation of previous data. Further work is therefore required to explore more sophisticated comparisons of data processes, for example comparing questionnaire with medical record data. A number of methodological issues also remain to be addressed, but the method shows promise as a tool to assist in the efficient design of expensive randomised controlled trials.

References

1. Wilson E, Abrams K. From Evidence Based Economics to Economics Based Evidence: Using Systematic Review to inform the design of future research. In: Shemilt I, Mugford M, Vale L, et al., eds. *Evidence Based Economics*. London: Blackwell Publishing, 2010.
2. Mugford M. How does the method of cost estimation affect the assessment of cost-effectiveness in health care? Exeter College and National Perinatal Epidemiology Unit, 1996.
3. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;**18**(3):341-64 doi: S0167-6296(98)00039-3 [pii].
4. Pratt J, Raiffa H, Schlaifer R. *Introduction to Statistical Decision Theory*. Cambridge, MA: Massachusetts Institute of Technology, 1995.
5. Briggs A, Sculpher M, Claxton K. Decision-making, uncertainty and the value of information. In: Gray A, Briggs A, eds. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press, 2006:165-200.
6. Claxton K, Cohen JT, Neumann PJ. When is evidence sufficient? *Health Aff (Millwood)* 2005;**24**(1):93-101
7. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. In: Senn S, Barnett V, eds. *Statistics in Practice*. Chichester, West Sussex: John Wiley & Sons Ltd, 2004:25.
8. Henriksson M, Lundgren F, Carlsson P. Informing the efficient use of health care and health care research resources - the case of screening for abdominal aortic aneurysm in Sweden. *Health Econ* 2006;**15**(12):1311-22
9. Garside R, Pitt M, Somerville M, et al. Surveillance of Barrett's oesophagus: exploring the uncertainty through systematic review, expert workshop and economic modelling. *Health Technol Assess* 2006;**10**(8):1-158
10. Robinson M, Palmer S, Sculpher M, et al. Cost-effectiveness of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling. *Health Technol Assess* 2005;**9**(27):iii-iv, ix-xi, 1-158
11. Tappenden P, Chilcott JB, Eggington S, et al. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis. *Health Technol Assess* 2004;**8**(27):iii, 1-78
12. Iglesias CP, Claxton K. Comprehensive decision-analytic model and Bayesian value-of-information analysis: pentoxifylline in the treatment of chronic venous leg ulcers. *Pharmacoeconomics* 2006;**24**(5):465-78
13. Castelnuovo E, Thompson-Coon J, Pitt M, et al. The cost-effectiveness of testing for hepatitis C in former injecting drug users. *Health Technol Assess* 2006;**10**(32):iii-iv, ix-xii, 1-93

14. Bartell SM, Ponce RA, Takaro TK, et al. Risk estimation and value-of-information analysis for three proposed genetic screening programs for chronic beryllium disease prevention. *Risk Anal* 2000;**20**(1):87-99
15. Wilson E, Gurusamy K, Gluud C, et al. Cost-utility and value-of-information analysis of early versus delayed laparoscopic cholecystectomy for acute cholecystitis. *Br J Surg* 2010;**97**(2):210-9 doi: 10.1002/bjs.6872.
16. Gurusamy K, Wilson E, Burroughs AK, et al. Intra-operative vs pre-operative endoscopic sphincterotomy in patients with gallbladder and common bile duct stones: cost-utility and value-of-information analysis. *Appl Health Econ Health Policy* 2012;**10**(1):15-29 doi: 10.2165/11594950-000000000-00000.
17. Wilson EC, Emery JD, Kinmonth AL, et al. The cost-effectiveness of a novel SIAscopic diagnostic aid for the management of pigmented skin lesions in primary care: a decision-analytic model. *Value Health* 2013;**16**(2):356-66 doi: 10.1016/j.jval.2012.12.008.
18. Colbourn TE, Asseburg C, Bojke L, et al. Preventive strategies for group B streptococcal and other bacterial infections in early infancy: cost effectiveness and value of information analyses. *BMJ* 2007;**335**(7621):655 doi: bmj.39325.681806.AD [pii] 10.1136/bmj.39325.681806.AD.
19. Bojke L, Hornby E, Sculpher M. A comparison of the cost effectiveness of pharmacotherapy or surgery (laparoscopic fundoplication) in the treatment of GORD. *Pharmacoeconomics* 2007;**25**(10):829-41 doi: 25103 [pii].
20. Shavit O, Leshno M, Goldberger A, et al. It's Time to Choose the Study Design!: Net Benefit Analysis of Alternative Study Designs to Acquire Information for Evaluation of Health Technologies. *Pharmacoeconomics* 2007;**25**(11):903-11 doi: 25112 [pii].
21. Drummond M, Sculpher M, Torrance G, et al. Basic Types of Economic Evaluation. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford: Oxford University Press, 2005:9.
22. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;**35**(11):1095-108
23. Briggs AH, O'Brien BJ. The death of cost-minimization analysis? *Health Econ* 2001;**10**(2):179-84 doi: 10.1002/hec.584.
24. Dakin H, Wordsworth S. Cost-minimisation analysis versus cost-effectiveness analysis, revisited. *Health Econ* 2013;**22**(1):22-34 doi: 10.1002/hec.1812.
25. Brazier J, Ratcliffe J, Salomon JA, et al. Foundations in Welfare Economics and Utility Theory: what should be valued? *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford University Press, 2007.
26. Pass C, Lowes B, L. D. *Collins Dictionary of Economics*. 2nd ed. Glasgow: Harper Collins, 1993.
27. Dr Foster. Dr Foster Health. Last Accessed 14/03/2013. <http://www.drfoosterhealth.co.uk/>.
28. Evans R. Supplier induced demand: some empirical evidence and implications. In: Perlman M, ed. *The Economics of Health and Medical Care*. London: Macmillan, 1974.

29. Culyer AJ. *Need and the National Health Service*. Oxford: Martin Robertson, 1976.
30. Morris S, Devlin N, Parkin D. Welfarist and Non-welfarist foundations of economic evaluation. *Economic Analysis in Health Care*. Chichester: Wiley and Sons Ltd, 2007:209-39.
31. Bentham J. *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press, 1780.
32. Donaldson C, Shackley P. Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Soc Sci Med* 1997;**44**(5):699-707
33. Brouwer WB, van Exel NJ, van den Berg B, et al. Process utility from providing informal care: the benefit of caring. *Health Policy* 2005;**74**(1):85-99 doi: S0168-8510(04)00291-X [pii] 10.1016/j.healthpol.2004.12.008.
34. Howard K, Salkeld G, McCaffery K, et al. HPV Triage Testing or Repeat Pap Smear for the Management of Atypical Squamous Cells (ASCUS) on Pap Smear: Is There Evidence of Process Utility? *Health Economics* 2008;**17**(5):593-605 doi: <http://www3.interscience.wiley.com/cgi-bin/jhome/5749>.
35. Pareto V. *Manuale di Economia Politica*: Lightning Source UK Ltd, 1906.
36. Elster J, Roemer J. *Interpersonal comparisons of well-being*. Cambridge: Cambridge University Press, 1991.
37. Kaldor N. Welfare Propositions of Economics and Interpersonal Comparisons of Utility. *The Economic Journal* 1939;**49**(195):549-52 doi: 10.2307/2224835.
38. Hicks JR. The Foundations of Welfare Economics. *The Economic Journal* 1939;**49**(196):696-712 doi: 10.2307/2225023.
39. Rawls J. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
40. Sen A. *Inequality Re-examined*. Oxford: Clarendon Press, 1992.
41. Sen A. *Commodities and Capabilities*. Amsterdam: Nord Holland, 1985.
42. Mooney G, Russell E. Equity in Health Care: The Need for a New Economics Paradigm? In: Scott A, Maynard A, Elliott R, eds. *Advances in Health Economics*. Chichester: John Wiley & Sons Ltd, 2003:205-21.
43. Culyer AJ. The normative economics of health care finance and provision. In: McGuire A, Fenn P, Mayhew K, eds. *Providing Health Care*. Oxford: Oxford University Press, 1991.
44. Culyer AJ. Commodities, characteristics of commodities, characteristics of people, utilities and the quality of life. In: Baldwin S, Godfrey C, Propper C, eds. *The Quality of Life: Perspectives and Policies*. London: Routledge, 1990.
45. Brouwer WB, Culyer AJ, van Exel NJ, et al. Welfarism vs. extra-welfarism. *J Health Econ* 2008;**27**(2):325-38 doi: S0167-6296(07)00090-2 [pii] 10.1016/j.jhealeco.2007.07.003.
46. Brouwer WB, Koopmanschap MA. On the economic foundations of CEA. Ladies and gentlemen, take your positions! *J Health Econ* 2000;**19**(4):439-59 doi: S0167-6296(99)00038-7 [pii].
47. Smith A. *The Wealth of Nations*. London: Penguin Group, 1776.

48. Drummond M, Sculpher M, Torrance G, et al. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford: Oxford University Press, 2005.
49. Byford S, Raftery J. Perspectives in economic evaluation. *BMJ* 1998;**316**(7143):1529-30
50. Wilson E, Thalanany M, Shepstone L, et al. Befriending carers of people with dementia: a cost utility analysis. *Int J Geriatr Psychiatry* 2009;**24**(6):610-23 doi: 10.1002/gps.2164.
51. Weisbrod BA, Test MA, Stein LI. Alternative to mental hospital treatment. II. Economic benefit-cost analysis. *Arch Gen Psychiatry* 1980;**37**(4):400-5
52. Brouwer W, Rutten FF, Koopmanschap M. Costing in Economic Evaluations. In: Drummond M, McGuire A, eds. *Economic evaluation in health care: merging theory with practice*. Oxford: Oxford University Press, 2001:68-93.
53. Wilson EC, Jayne DR, Dellow E, et al. The cost-effectiveness of mycophenolate mofetil as firstline therapy in active lupus nephritis. *Rheumatology (Oxford)* 2007;**46**(7):1096-101 doi: kem054 [pii] 10.1093/rheumatology/kem054.
54. Glick H, Doshi J, Sonnad S, et al. Designing economic evaluations in clinical trials. *Economic Evaluation in Clinical Trials*. Oxford: Oxford University Press, 2007.
55. Byford S, Leese M, Knapp M, et al. Comparison of alternative methods of collection of service use data for the economic evaluation of health care interventions. *Health Econ* 2007;**16**(5):531-6
56. Department of Health. NHS Payment by Results 2010-11 National Tariff Information. Last Accessed 29/03/2013. <http://data.gov.uk/dataset/payment-by-results-2010-11-national-tariff-information>.
57. Gerard K, Mooney G. QALY league tables: handle with care. *Health Econ* 1993;**2**(1):59-64
58. Koopmanschap MA, Rutten FF. A practical guide for calculating indirect costs of disease. *Pharmacoeconomics* 1996;**10**(5):460-6
59. Baker R, Bateman I, Donaldson C, et al. Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. *Health Technol Assess* 2010;**14**(27):1-162 doi: 10.3310/hta14270.
60. Dolan P, Shaw R, Tsuchiya A, et al. QALY maximisation and people's preferences: a methodological review of the literature. *Health Econ* 2005;**14**(2):197-208 doi: 10.1002/hec.924.
61. Sculpher M. The role and estimation of productivity costs in economic evaluation. In: Drummond M, McGuire A, eds. *Economic evaluation in health care: merging theory with practice*. Oxford: Oxford University Press, 2001:94-112.
62. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics* 1993;**4**(5):353-65
63. van Roijen L, Essink-Bot ML, Koopmanschap MA, et al. Labor and health status in economic evaluation of health care. The Health and Labor Questionnaire. *Int J Technol Assess Health Care* 1996;**12**(3):405-15

64. Glick H, Doshi J, Sonnad S, et al. Valuing Medical Service use. *Economic Evaluation in Clinical Trials*. Oxford: Oxford University Press, 2007.
65. Posnett J, Jan S. Indirect cost in economic evaluation: the opportunity cost of unpaid inputs. *Health Econ* 1996;**5**(1):13-23 doi: 10.1002/(SICI)1099-1050(199601)5:1<13::AID-HEC182>3.0.CO;2-J.
66. Gold MR, Siegel JE, Russell LB, et al. *Cost-effectiveness in Health and Medicine*. Oxford: Oxford University Press, 1996.
67. Brouwer WB, van Exel NJ, Koopmanschap MA, et al. The valuation of informal care in economic appraisal. A consideration of individual choice and societal costs of time. *Int J Technol Assess Health Care* 1999;**15**(1):147-60
68. Koopmanschap MA, Rutten FF, van Ineveld BM, et al. The friction cost method for measuring indirect costs of disease. *J Health Econ* 1995;**14**(2):171-89 doi: 0167629694000445 [pii].
69. Rice DP, Cooper BS. The economic value of human life. *American journal of public health and the nation's health* 1967;**57**(11):1954-66
70. Drummond M, Sculpher M, Torrance G, et al. Cost Analysis. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford: Oxford University Press, 2005:55-101.
71. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 2004;**13**(5):437-52 doi: 10.1002/hec.864.
72. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. Last Accessed <http://www.nice.org.uk/media/D45/1E/GuideToMethodsTechnologyAppraisal2013.pdf>.
73. Claxton K, Martin S, Soares M, et al. CHE Research Paper 81: Methods for the Estimation of the NICE Cost Effectiveness Threshold. York, UK: University of York, 2013.
74. Briggs A, Sculpher M, Claxton K. Chapter 4: Making decision models probabilistic. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press, 2006:77-120.
75. Glick H, Doshi J, Sonnad S, et al. Sampling uncertainty: calculation, sample size and power, and decision criteria. *Economic Evaluation in Clinical Trials*. Oxford: Oxford University Press, 2007.
76. Barber JA, Thompson SG. Analysis and interpretation of cost data in randomised controlled trials: review of published studies. *BMJ* 1998;**317**(7167):1195-200
77. Eschenbach TG. Technical Note: Constructing Tornado Diagrams with Spreadsheets. *The Engineering Economist* 2006;**51**(2):195-204 doi: 10.1080/00137910600695676.
78. Mullen PM. Quantifying priorities in healthcare: transparency or illusion? *Health Serv Manage Res* 2004;**17**(1):47-58 doi: 10.1258/095148404322772723.

79. Holm S. The second phase of priority setting. Goodbye to the simple solutions: the second phase of priority setting in health care. *BMJ* 1998;**317**(7164):1000-2
80. Raiffa H, Schlaifer R. *Applied Statistical Decision Theory*. Classics library edition, published 2000 (original 1961) ed. New York, NY: J Wiley & sons Inc., 2000.
81. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ* 1996;**5**(6):513-24
82. Willan AR, Pinto EM. The value of information and optimal clinical trial design. *Stat Med* 2005;**24**(12):1791-806
83. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. *Med Decis Making* 2004;**24**(2):207-27 doi: 10.1177/0272989X04263162.
84. Claxton K, Ginnelly L, Sculpher M, et al. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. *Health Technol Assess* 2004;**8**(31):1-103, iii
85. Claxton K, Eggington S, Ginnelly L, et al. A pilot study of value of information analysis to support research recommendations for NICE. Centre for Health Economics Research Paper 4. York, UK: Centre for Health Economics, 2005.
86. Willan A, Briggs A. Power and Sample Size Determination: The Value of Information Approach. *Statistical Analysis of Cost-effectiveness Data*. Chichester: John Wiley & Sons Ltd, 2006:108-16.
87. Drummond M, Davis L. Economic analysis alongside clinical trials: revisiting methodological issues. *Int J Technol Assess Health Care* 1991;**7**(4):561-73
88. Sculpher MJ, Claxton K, Drummond M, et al. Whither trial-based economic evaluation for health care decision making? *Health Econ* 2006;**15**(7):677-87
89. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. Last Accessed <http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf>.
90. Haute Autorité de Santé. Choices in Methods for Economic Evaluation: A methodological guide. Paris, FR, 2012.
91. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;**312**(7023):71-2
92. Buxton MJ, Drummond MF, Van Hout BA, et al. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997;**6**(3):217-27 doi: 10.1002/(SICI)1099-1050(199705)6:3<217::AID-HEC267>3.0.CO;2-W [pii].
93. Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. *J Health Serv Res Policy* 1997;**2**(1):26-30
94. Fenwick E, Claxton K, Sculpher M, et al. Improving the efficiency and relevance of health technology assessment: the role of decision analytic modelling. Centre for Health Economics Discussion paper 179. York: Centre for Health Economics, University of York, 2000.

95. Sculpher M, Claxton K. Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty--when is there sufficient evidence? *Value Health* 2005;**8**(4):433-46
96. Arrow KJ, Lind RC. Uncertainty and Evaluation of Public Investment Decisions. *Am Econ Rev* 1970;**60**(3):364-78
97. Deeks J, Higgins J, Altman DG, et al. Analysing data and undertaking meta-analyses. In: Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: Wiley-Blackwell, 2008:243-96.
98. O'Hagan A, Buck CE, Daneshkhah A, et al. *Uncertain Judgements: Eliciting Experts' Probabilities*. Chichester, West Sussex: Wiley, 2006.
99. Fleurence RL, Torgerson DJ. Setting priorities for research. *Health Policy* 2004;**69**(1):1-10
100. National Institute for Health and Care Excellence. Last Accessed <http://www.nice.org.uk/>.
101. Cookson R, McDaid D, Maynard A. Wrong SIGN, NICE mess: is national guidance distorting allocation of resources? *BMJ* 2001;**323**(7315):743-5
102. National Institute for Clinical Excellence. Multiple sclerosis - beta interferon and glatiramer acetate. (TA32). Last Accessed <http://www.nice.org.uk/TA32>.
103. Boggild M, Palace J, Barton P, et al. Multiple sclerosis risk sharing scheme: two year results of clinical cohort study with historical comparator. *BMJ* 2009;**339**:b4677 doi: 10.1136/bmj.b4677.
104. Pickin M, Cooper CL, Chater T, et al. The Multiple Sclerosis Risk Sharing Scheme Monitoring Study--early results and lessons for the future. *BMC Neurol* 2009;**9**:1 doi: 10.1186/1471-2377-9-1.
105. Rojas JI, Romano M, Ciapponi A, et al. Interferon Beta for primary progressive multiple sclerosis. *Cochrane Database Syst Rev* 2010(1):CD006643 doi: 10.1002/14651858.CD006643.pub3.
106. Oliver BJ, Kohli E, Kasper LH. Interferon therapy in relapsing-remitting multiple sclerosis: a systematic review and meta-analysis of the comparative trials. *Journal of the neurological sciences* 2011;**302**(1-2):96-105 doi: 10.1016/j.jns.2010.11.003.
107. Raftery J. Multiple sclerosis risk sharing scheme: a costly failure. *BMJ* 2010;**340**:c1672 doi: 10.1136/bmj.c1672.
108. Garrido MV, Kristensen FB, Nielsen CP, et al. *Health Technology Assessment and Health Policy Making in Europe*. Copenhagen, DK: World Health Organisation on behalf of the European Observatory on Health Systems and Policies, 2008.
109. Kristensen FB, Chamova J, Hansen NW. Toward a sustainable European Network for Health Technology Assessment. The EUnetHTA project. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 2006;**49**(3):283-5 doi: 10.1007/s00103-005-1229-9.

110. Lampe K, Makela M, Garrido MV, et al. The HTA core model: a novel method for producing and reporting health technology assessments. *Int J Technol Assess Health Care* 2009;**25 Suppl 2**:9-20 doi: 10.1017/S0266462309990638.
111. Pasternack I, Anttila H, Makela M, et al. Testing the HTA core model: experiences from two pilot projects. *Int J Technol Assess Health Care* 2009;**25 Suppl 2**:21-7 doi: 10.1017/S026646230999064X.
112. Turner S, Chase DL, Milne R, et al. The health technology assessment adaptation toolkit: description and use. *Int J Technol Assess Health Care* 2009;**25 Suppl 2**:37-41 doi: 10.1017/S0266462309990663.
113. Turner S, Chase DL, Milne R, et al. The adaptation of health technology assessment reports: identification of the need for, and development of, a toolkit to aid the process. *Int J Technol Assess Health Care* 2009;**25 Suppl 2**:28-36 doi: 10.1017/S0266462309990651.
114. Chase D, Rosten C, Turner S, et al. Development of a toolkit and glossary to aid in the adaptation of health technology assessment (HTA) reports for use in different contexts. *Health Technol Assess* 2009;**13**(59):1-142, iii doi: 10.3310/hta13590.
115. INAHTA. International Network of Agencies for Health Technology Assessment - Homepage. Last Accessed 04/04/2013. <http://www.inahta.net/>.
116. Hailey D. Development of the International Network of Agencies for Health Technology Assessment. *Int J Technol Assess Health Care* 2009;**25 Suppl 1**:24-7 doi: 10.1017/S0266462309090370.
117. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. General Methods for the Assessment of the Relation of Benefits to Costs, version 1.0. Cologne: Institute for Quality and Efficiency in Health Care, 2009.
118. Lopert R. Evidence-based decision-making within Australia's pharmaceutical benefits scheme. *Issue brief* 2009;**60**:1-13
119. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee v4.3. Canberra, Australia: Australian Government Department of Health and Ageing, 2008.
120. Institute of Medicine. Roundtable on Evidence Based Medicine, Learning What Works Best: The Nation's Need for Evidence of Comparative Effectiveness in Healthcare. Washington, D.C., 2007.
121. American College of P. Information on cost-effectiveness: an essential product of a national comparative effectiveness program. *Ann Intern Med* 2008;**148**(12):956-61
122. Chalkidou K, Tunis S, Lopert R, et al. Comparative effectiveness research and evidence-based health policy: experience from four countries. *Milbank Q* 2009;**87**(2):339-67 doi: 10.1111/j.1468-0009.2009.00560.x.
123. Hutton J, Trueman P, Henshall C. Coverage with evidence development: an examination of conceptual and policy issues. *Int J Technol Assess Health Care* 2007;**23**(4):425-32 doi: S0266462307070651 [pii] 10.1017/S0266462307070651.

124. Claxton K, Palmer S, Longworth L, et al. Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development. *Health Technol Assess* 2012;**16**(46):1-323 doi: 10.3310/hta16460.
125. Fenwick E, Claxton K, Sculpher M. The value of implementation and the value of information: combined and uneven development. *Med Decis Making* 2008;**28**(1):21-32 doi: 10.1177/0272989X07308751.
126. Claxton KP, Sculpher MJ. Using Value of Information Analysis to Prioritise Health Research: Some Lessons from Recent UK Experience. *Pharmacoeconomics* 2006;**24**(11):1055-68
127. Department of Health. NHS Reference Costs 2009-2010. Last Accessed 26 May. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_123459.
128. Mogyorosy Z, Smith P. The main methodological issues in costing health care services. A literature review. CHE Research Paper 7. York: University of York, 2005.
129. Finkler SA. *Issues in Cost Accounting for Health Care Organisations*. Maryland, USA, 1994.
130. Dickey B, Beecham J, Latimer E, et al. *Toolkit Estimating Per Unit Treatment Costs for Mental Health and Substance Abuse Programs*. Cambridge, MA: The Evaluation Center @ HSRI., 1999.
131. CCOHTA. A guidance document for the costing process v1.0. Ontario, Canada: Canadian Co-ordinating Office for Health Technology Assessment, 1996.
132. Beecham J. Collecting and estimating costs. In: Knapp M, ed. *The Economic Evaluation of Mental Health Care*. London, UK: Ashgate Publishing Ltd, 1995:61-82.
133. Hartley RJ, Keen EM, Large JA, et al. *On-line Searching: Principles and Practice* Butterworths, 1990.
134. Thomson Reuters. Web of Knowledge. Last Accessed 29/08/2013. <http://wok.mimas.ac.uk/>.
135. Steuten L, van de Wetering G, Groothuis-Oudshoorn K, et al. A Systematic and Critical Review of the Evolving Methods and Applications of Value of Information in Academia and Practice. *Pharmacoeconomics* 2013;**31**(1):25-48 doi: 10.1007/s40273-012-0008-3.
136. Moons CK, van Es GA, Stijnen T, et al. Efficiency optimization of the selection period in therapeutic trials. *J Clin Epidemiol* 1997;**50**(7):793-800
137. Falagas ME, Bliziotis IA. Cost-effectiveness of different study designs: an overlooked parameter in clinical research. *J Clin Epidemiol* 2007;**60**(3):215-6 doi: 10.1016/j.jclinepi.2006.05.008.
138. Thach CT, Fisher LD. Self-designing two-stage trials to minimize expected costs. *Biometrics* 2002;**58**(2):432-8

139. Griffin S, Welton NJ, Claxton K. Exploring the research decision space: the expected value of information for sequential research designs. *Med Decis Making* 2010;**30**(2):155-62 doi: 10.1177/0272989X09344746.
140. Hak E, Wei F, Grobbee DE, et al. A nested case-control study of influenza vaccination was a cost-effective alternative to a full cohort analysis. *J Clin Epidemiol* 2004;**57**(9):875-80 doi: 10.1016/j.jclinepi.2004.01.019.
141. Cohen D, Longo MF, Williams J, et al. Estimating the marginal value of 'better' research output: 'designed' versus 'routine' data in randomised controlled trials. *Health Econ* 2003;**12**(11):959-74 doi: 10.1002/hec.791.
142. Willan AR, Eckermann S. Optimal clinical trial design using value of information methods with imperfect implementation. *Health Econ* 2010;**19**(5):549-61 doi: 10.1002/hec.1493.
143. Hoomans T, Fenwick EA, Palmer S, et al. Value of information and value of implementation: application of an analytic framework to inform resource allocation decisions in metastatic hormone-refractory prostate cancer. *Value Health* 2009;**12**(2):315-24 doi: 10.1111/j.1524-4733.2008.00431.x.
144. Eckermann S, Willan AR. The option value of delay in health technology assessment. *Med Decis Making* 2008;**28**(3):300-5 doi: 0272989X07312477 [pii] 10.1177/0272989X07312477.
145. Karnon J. Planning the efficient allocation of research funds: an adapted application of a non-parametric Bayesian value of information analysis. *Health Policy* 2002;**61**(3):329-47
146. Willan AR. Optimal sample size determinations from an industry perspective based on the expected value of information. *Clin Trials* 2008;**5**(6):587-94 doi: 10.1177/1740774508098413.
147. Miller P. Role of pharmacoeconomic analysis in R&D decision making: when, where, how? *Pharmacoeconomics* 2005;**23**(1):1-12
148. Lal JA, Schulte In den Baumen T, Morre SA, et al. Public health and valorization of genome-based technologies: a new model. *Journal of translational medicine* 2011;**9**:207 doi: 10.1186/1479-5876-9-207.
149. Janssen MP, Koffijberg H. Enhancing value of information analyses. *Value Health* 2009;**12**(6):935-41 doi: 10.1111/j.1524-4733.2009.00548.x.
150. Groot Koerkamp B, Hunink MG, Stijnen T, et al. Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis. *Med Decis Making* 2007;**27**(2):101-11 doi: 10.1177/0272989X06297394.
151. Eckermann S, Briggs A, Willan AR. Health technology assessment in the cost-disutility plane. *Med Decis Making* 2008;**28**(2):172-81 doi: 10.1177/0272989X07312474.
152. Whitehead J. *The design and analysis of sequential clinical trials*. 2nd ed. New York: Ellis Horwood, 1992.
153. Evans WK, Dahrouge S, Stapleton J, et al. An estimate of the cost of conducting phase II trials in lung cancer. *Lung Cancer* 2000;**28**(2):85-95 doi: 10.1016/s0169-5002(99)00132-4.

154. Kilgore ML, Goldman DP. Drug costs and out-of-pocket spending in cancer clinical trials. *Contemporary Clinical Trials* 2008;**29**(1):1-8 doi: 10.1016/j.cct.2007.04.004.
155. Bennett CL, Stinson TJ, Vogel V, et al. Evaluating the financial impact of clinical trials in oncology: results from a pilot study from the Association of American Cancer Institutes/Northwestern University clinical trials costs and charges project. *J Clin Oncol* 2000;**18**(15):2805-10
156. Wagner JL, Alberts SR, Sloan JA, et al. Incremental costs of enrolling cancer patients in clinical trials: a population-based study. *J Natl Cancer Inst* 1999;**91**(10):847-53
157. Weiss JP, Parsons LS, Every NR, et al. Does enrollment in a randomized clinical trial lead to a higher cost of routine care? *American Heart Journal* 2002;**143**(1):140-44 doi: 10.1067/mhj.2002.119615.
158. Perrin RS, Lopez FJG. Incremental Drug Treatment Cost in HIV-Positive Patients in Industry-Sponsored Clinical Trials. *Annals of Pharmacotherapy* 2008;**42**(11):1586-91 doi: 10.1345/aph.1L156.
159. Emanuel EJ, Schnipper LE, Kamin DY, et al. The costs of conducting clinical research. *Journal of Clinical Oncology* 2003;**21**(22):4145-50 doi: 10.1200/jco.2003.08.156.
160. Wright JR, Levine MN. Researching the cost of research. *Journal of Clinical Oncology* 2003;**21**(22):4081-82 doi: 10.1200/jco.2003.08.972.
161. Roche K, Paul N, Smuck B, et al. Factors affecting workload of cancer clinical trials: Results of a multicenter study of the National Cancer Institute of Canada Clinical Trials Group. *Journal of Clinical Oncology* 2002;**20**(2):545-56 doi: 10.1200/jco.20.2.545.
162. Sargent DJ, George SL. Clinical Trials Data Collection: When Less Is More. *Journal of Clinical Oncology* 2010;**28**(34):5019-21 doi: 10.1200/jco.2010.31.7024.
163. Demol P, Weihrauch TR. The multinational clinical trials - Design, management and costs. *Medizinische Klinik* 1997;**92**(2):117-23
164. Urban N, Self S, Kessler L, et al. Analysis of the costs of a large prevention trial. *Controlled Clinical Trials* 1990;**11**(2):129-46 doi: 10.1016/0197-2456(90)90006-n.
165. Thornquist MD, Urban N, Tseng A, et al. Research cost analyses to aid in decision-making in the conduct of a large prevention trial, CARET. *Controlled Clinical Trials* 1993;**14**(4):325-39 doi: 10.1016/0197-2456(93)90229-7.
166. Eisenstein EL, Lemons PW, Tardiff BE, et al. Reducing the costs of phase III cardiovascular clinical trials. *American Heart Journal* 2005;**149**(3):482-88 doi: 10.1016/j.ahj.2004.04.049.
167. Eisenstein EL, Collins R, Cracknell BS, et al. Sensible approaches for reducing clinical trial costs. *Clinical Trials* 2008;**5**(1):75-84 doi: 10.1177/1740774507087551.
168. Vozdolska R, Sano M, Aisen P, et al. The net effect of alternative allocation ratios on recruitment time and trial cost. *Clinical Trials* 2009;**6**(2):126-32 doi: 10.1177/1740774509103485.

169. Allison DB, Allison RL, Faith MS, et al. Power and money: Designing statistically powerful studies while minimizing financial costs. *Psychological Methods* 1997;**2**(1):20-33 doi: 10.1037/1082-989x.2.1.20.
170. Vickers AJ, Scardino PT. The clinically-integrated randomized trial: proposed novel method for conducting large trials at low cost. *Trials* 2009;**10** doi: 10.1186/1745-6215-10-14.
171. Evans CJ, Crawford B. Data collection methods in prospective economic evaluations: how accurate are the results? *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2000;**3**(4):277-86 doi: 10.1046/j.1524-4733.2000.34005.x.
172. Howard M, Agarwal G, Lytwyn A. Accuracy of self-reports of Pap and mammography screening compared to medical record: a meta-analysis. *Cancer Causes & Control* 2009;**20**(1):1-13 doi: 10.1007/s10552-008-9228-4.
173. Rauscher GH, Johnson TP, Cho YI, et al. Accuracy of self-reported cancer-screening histories: A meta-analysis. *Cancer Epidemiology Biomarkers & Prevention* 2008;**17**(4):748-57 doi: 10.1158/1055-9965.epi-07-2629.
174. Bhandari A, Wagner T. Self-reported utilization of health care services: Improving measurement and accuracy. *Medical Care Research and Review* 2006;**63**(2):217-35 doi: 10.1177/1077558705285298.
175. Gama H, Correia S, Lunet N. Questionnaire design and the recall of pharmacological treatments: a systematic review. *Pharmacoepidemiology and Drug Safety* 2009;**18**(3):175-87 doi: 10.1002/pds.1703.
176. Evans C, Crawford B. Patient self-reports in pharmacoeconomic studies. Their use and impact on study validity. *Pharmacoeconomics* 1999;**15**(3):241-56
177. Goossens ME, Rutten-van Molken MP, Vlaeyen JW, et al. The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *J Clin Epidemiol* 2000;**53**(7):688-95 doi: S0895435699001778 [pii].
178. Merlo J, Berglund G, Wirfalt E, et al. Self-administered questionnaire compared with a personal diary for assessment of current use of hormone therapy: An analysis of 16,060 women. *American Journal of Epidemiology* 2000;**152**(8):788-92 doi: 10.1093/aje/152.8.788.
179. Patel A, Rendu A, Moran P, et al. A comparison of two methods of collecting economic data in primary care. *Family Practice* 2005;**22**(3):323-27 doi: DOI 10.1093/fampra/cmi027.
180. Petrou S, Murray L, Cooper P, et al. The accuracy of self-reported healthcare resource utilization in health economic studies. *Int J Technol Assess Health Care* 2002;**18**(3):705-10
181. Chapko MK, Liu CF, Perkins M, et al. Equivalence of two healthcare costing methods: bottom-up and top-down. *Health Econ* 2009;**18**(10):1188-201 doi: 10.1002/hec.1422.
182. Tan SS, Rutten FF, van Ineveld BM, et al. Comparing methodologies for the cost estimation of hospital services. *Eur J Health Econ* 2009;**10**(1):39-45 doi: 10.1007/s10198-008-0101-x.

183. Geue C, Lewsey J, Lorgelly P, et al. Spoilt for choice: implications of using alternative methods of costing hospital episode statistics. *Health Econ* 2012;**21**(10):1201-16 doi: 10.1002/hec.1785.
184. Chapko MK, Hedrick S. Cost as a study outcome: sensitivity of study conclusions to the method of estimating cost. *Med Care* 1999;**37**(4 Suppl Va):AS37-44
185. Knapp M, Beecham J. Reduced list costings: examination of an informed short cut in mental health research. *Health Econ* 1993;**2**(4):313-22
186. Whynes DK, Walker AR. On approximations in treatment costing. *Health Econ* 1995;**4**(1):31-9
187. Freer CB. Self-care: a health diary study. *Med Care* 1980;**18**(8):853-61
188. Freer CB. Health diaries: a method of collecting health information. *J R Coll Gen Pract* 1980;**30**(214):279-82
189. Verbrugge LM. Health diaries. *Med Care* 1980;**18**(1):73-95
190. Carp FM, Carp A. The validity, reliability and generalizability of diary data. *Exp Aging Res* 1981;**7**(3):281-96 doi: 10.1080/03610738108259811.
191. Norman GR, McFarlane AH, Streiner DL, et al. Health diaries: strategies for compliance and relation to other measures. *Med Care* 1982;**20**(6):623-9
192. West SL, Savitz DA, Koch G, et al. Recall accuracy for prescription medicines - self-report compared with database information. *American Journal of Epidemiology* 1995;**142**(10):1103-12
193. Beecham J, Knapp M. Costing psychiatric interventions. *Measuring Mental Health Needs*. London: Royal College of Psychiatrists, 2001:200-24.
194. Gandjour A. Assessing the value of a future study. *Pharmacoeconomics* 2008;**26**(10):891-2; author reply 92-4 doi: 26108 [pii].
195. Shavit O, Leshno M, Goldberger A, et al. Assessing the Value of a Future Study The Authors' Reply. *Pharmacoeconomics* 2008;**26**(10):892-94 doi: 10.2165/00019053-200826100-00009.
196. Williams A. The cost-benefit approach. *British medical bulletin* 1974;**30**(3):252-6
197. Drummond MF, Stoddart GL. Economic analysis and clinical trials. *Control Clin Trials* 1984;**5**(2):115-28
198. Philips Z, Claxton K, Palmer S. The half-life of truth: what are appropriate time horizons for research decisions? *Med Decis Making* 2008;**28**(3):287-99 doi: 0272989X07312724 [pii] 10.1177/0272989X07312724.
199. Eckermann S, Willan AR. Time and expected value of sample information wait for no patient. *Value Health* 2008;**11**(3):522-6 doi: VHE296 [pii] 10.1111/j.1524-4733.2007.00296.x.
200. Eckermann S, Willan AR. Expected value of information and decision making in HTA. *Health Econ* 2007;**16**(2):195-209 doi: 10.1002/hec.1161.

201. Eckermann S, Willan AR. Globally optimal trial design for local decision making. *Health Econ* 2009;**18**(2):203-16 doi: 10.1002/hec.1353.
202. Curtis L. *Unit Costs of Health and Social Care 2012*. Canterbury: University of Kent at Canterbury, 2013.
203. Devore JL, Peck R. *Introductory statistics*. 2nd ed. Minneapolis/St. Paul: West Pub., 1994.
204. Eckermann S, Karnon J, Willan AR. The value of value of information: best informing research design and prioritization using current methods. *Pharmacoeconomics* 2010;**28**(9):699-709 doi: 10.2165/11537370-000000000-00000.
205. Coyle D, Oakley J. Estimating the expected value of partial perfect information: a review of methods. *Eur J Health Econ* 2008;**9**(3):251-9 doi: 10.1007/s10198-007-0069-y.
206. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration, 2011.
207. Brennan A, Kharroubi S, O'Hagan A, et al. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Med Decis Making* 2007;**27**(4):448-70 doi: 10.1177/0272989X07302555.
208. Stevenson MD, Oakley J, Chilcott JB. Gaussian process modeling in conjunction with individual patient simulation modeling: a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis. *Med Decis Making* 2004;**24**(1):89-100 doi: 10.1177/0272989X03261561.
209. Griffin S, Claxton K, Hawkins N, et al. Probabilistic analysis and computationally expensive models: Necessary and required? *Value Health* 2006;**9**(4):244-52 doi: VHE107 [pii] 10.1111/j.1524-4733.2006.00107.x.
210. Bojke L, Claxton K, Palmer S, et al. Defining and characterising structural uncertainty in decision analytic models. *CHE Research Paper 9*. York: Centre for Health Economics, University of York, UK, 2006.
211. Drummond M, Sculpher M, Torrance G, et al. Economic Evaluation using Decision Analytic Modelling: The need to reflect all appropriate evidence. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford: Oxford University Press, 2005:280.
212. O'Connor D, Green S, Higgins J. Chapter 5: Defining the review question and developing criteria for including studies (section 5.1.2). In: Higgins J, Green S, eds. *Cochrane Handbook of Systematic Reviews of Interventions. Version 5.0.1 (updated September 2008)*: The Cochrane Collaboration, 2008.
213. Shemilt I, Mugford HM, Byford S, et al. Chapter 15: Incorporating economics evidence. In: Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration, 2011.
214. Gurusamy KS, Samraj K. Early versus delayed laparoscopic cholecystectomy for acute cholecystitis. *Cochrane Database Syst Rev* 2006(4):CD005440 doi: 10.1002/14651858.CD005440.pub2.

215. Gurusamy K, Samraj K, Wilson EC, et al. Early versus delayed laparoscopic cholecystectomy for acute cholecystitis - a meta-analysis of randomised clinical trials. (Submitted for publication) 2008
216. Greenhalgh T. How to read a paper. Getting your bearings (deciding what the paper is about). *BMJ* 1997;**315**(7102):243-6
217. Guyatt GH, Sackett DL, Sinclair JC, et al. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA* 1995;**274**(22):1800-4
218. O'Hagan A, Oakley J. SHELF: The SHEffield ELicitation Framework. Last Accessed 12th July 2009. <http://tonyohagan.co.uk/shelf/>.
219. Ades AE, Welton NJ, Caldwell D, et al. Multiparameter evidence synthesis in epidemiology and medical decision making. *J Health Serv Res Policy* 2008;**13**(S3):12-22
220. Cooper NJ, Sutton AJ, Abrams KR, et al. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health Econ* 2004;**13**(3):203-26 doi: 10.1002/hec.804.
221. Devore J, Peck R. Regression and correlation: descriptive methods. *Introductory Statistics*. St Paul, MN: West Publishing Company, 1994:417-58.
222. Purshouse R, Brennan A, Latimer N, et al. Modelling to Assess the Effectiveness and Cost-Effectiveness of Public Health Related Strategies and Interventions to Reduce Alcohol Attributable Harm in England Using the Sheffield Alcohol Policy Model Version 2.0. Sheffield, 2009.
223. Thokala P, Kruger J, Brennan A, et al. The Sheffield Type 1 Diabetes Policy Model. Sheffield, 2013.
224. Barton GR, Briggs AH, Fenwick EA. Optimal cost-effectiveness decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI). *Value Health* 2008;**11**(5):886-97 doi: VHE358 [pii] 10.1111/j.1524-4733.2008.00358.x.
225. Walsh B. Markov Chain Monte Carlo and Gibbs Sampling. Lecture Notes. Last Accessed 03/08/2010. http://membres-timc.imag.fr/Olivier.Francois/mcmc_gibbs_sampling.pdf.
226. Lunn DJ, Thomas A, Best N, et al. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comput* 2000;**10**(4):325-37
227. Cooper N, Abrams K, Sutton A, et al. A Bayesian approach to Markov modelling in cost-effectiveness analyses: application to taxane use in advanced breast cancer. *J R Statist Soc A* 2003;**166**(3):389-405
228. Chan S, Friedrichs K, Noel D, et al. Prospective randomized trial of docetaxel versus doxorubicin in patients with metastatic breast cancer. *J Clin Oncol* 1999;**17**(8):2341-54
229. Nelsen RB. *An Introduction to Copulas*. 2nd ed. New York: Springer Science+Business Media Inc., 2006.

230. Meneguzzo D, Vecchiato W. Copula sensitivity in collateralized debt obligations and basket default swaps. *Journal of Futures Markets* 2004;**24**(1):37-70
231. Kumar P, Shoukri MM. Copula based prediction models: an application to an aortic regurgitation study. *BMC Med Res Methodol* 2007;**7**:21 doi: 1471-2288-7-21 [pii] 10.1186/1471-2288-7-21.
232. Reed SD, Anstrom KJ, Li Y, et al. Updated estimates of survival and cost effectiveness for imatinib versus interferon-alpha plus low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukaemia. *Pharmacoeconomics* 2008;**26**(5):435-46 doi: 2657 [pii].
233. Ramsey SD, Clarke LD, Roberts CS, et al. An economic evaluation of atorvastatin for primary prevention of cardiovascular events in type 2 diabetes. *Pharmacoeconomics* 2008;**26**(4):329-39 doi: 2645 [pii].
234. Teerawattananon K, lewsakul S, Yenjitir C, et al. Economic evaluation of treatment administration strategies of ganciclovir for cytomegalovirus retinitis in HIV/AIDS patients in Thailand: a simulation study. *Pharmacoeconomics* 2007;**25**(5):413-28 doi: 2555 [pii].
235. Charlesworth G, Shepstone L, Wilson E, et al. Does befriending by trained lay workers improve psychological well-being and quality of life for carers of people with dementia, and at what cost? A randomised controlled trial. *Health Technol Assess* 2008;**12**(4):iii, v-ix, 1-78 doi: 99/34/07 [pii].
236. Price D, Musgrave S, Wilson E, et al. A pragmatic single-blind randomised controlled trial and economic evaluation of the use of leukotriene receptor antagonists in primary care at steps 2 and 3 of the national asthma guidelines (ELEVATE study). *Health Technol Assess* 2011;**15**(21):1-132 doi: 10.3310/hta15210.
237. Peek GJ, Mugford M, Tiruvoipati R, et al. Efficacy and economic assessment of conventional ventilatory support versus extracorporeal membrane oxygenation for severe adult respiratory failure (CESAR): a multicentre randomised controlled trial. *Lancet* 2009;**374**(9698):1351-63 doi: S0140-6736(09)61069-2 [pii] 10.1016/S0140-6736(09)61069-2.
238. Rubin DB. *Mutliple Imputation for Nonresponse in Surveys*. Hoboken, New Jersey: Wiley, 2004.
239. Wilson EC, Price D, Musgrave SD, et al. Cost Effectiveness of Leukotriene Receptor Antagonists versus Inhaled Corticosteroids for Initial Asthma Controller Therapy: A Pragmatic Trial. *Pharmacoeconomics* 2010;**28**(7):585-95
240. Wilson EC, Price D, Musgrave SD, et al. Cost Effectiveness of Leukotriene Receptor Antagonists versus Long-Acting Beta-2 Agonists as Add-On Therapy to Inhaled Corticosteroids for Asthma: A Pragmatic Trial. *Pharmacoeconomics* 2010;**28**(7):597-608 doi: 10.2165/11537550-000000000-00000.
241. Price D, Musgrave SD, Shepstone L, et al. Leukotriene antagonists as first-line or add-on asthma-controller therapy. *N Engl J Med* 2011;**364**(18):1695-707 doi: 10.1056/NEJMoa1010846.
242. R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2010.

243. Clemen RT, Reilly T. *Making hard decisions with decision tools*. Pacific Grove, CA.: Duxbury Press, 2001.
244. Phillips Ld. Group elicitation of probability distributions: are many heads better than one? In: Shanteau J, Mellers B, Schum DA, eds. *Decision Science and Technology: Reflections on the Contributions of Ward Edwards*. Norwell, MA: Kluwer Academic Publishers, 1999:313-30.
245. O'Hagan A. Eliciting a Univariate Distribution from a Several Experts. Sheffield: University of Sheffield, 2010.
246. O'Hagan A. Eliciting a Distribution for Several Variables. Sheffield: University of Sheffield, 2010.
247. Oakley JE, O'Hagan A. SHELF: The Sheffield Elicitation Framework (version 2.0). Sheffield, UK: School of Mathematics and Statistics, University of Sheffield, 2010.
248. *British National Formulary 49*. 49 ed. London: British Medical Association & Royal Pharmaceutical Society of Great Britain, 2005.
249. NHS Health and Social Care Information Centre. Prescription Cost Analysis, 2005. Last Accessed <http://www.ic.nhs.uk/pubs/prescostanalysis2005>.
250. Carlson JJ, Garrison LP, Ramsey SD, et al. The potential clinical and economic outcomes of pharmacogenomic approaches to EGFR-tyrosine kinase inhibitor therapy in non-small-cell lung cancer. *Value Health* 2009;**12**(1):20-7 doi: 10.1111/j.1524-4733.2008.00415.x.
251. Raiffa H, Schlaifer R. *Applied Statistical Decision Theory*. Boston, MA: Harvard Business School, 1961.
252. Thompson MS. Decision-analytic determination of study size. The case of electronic fetal monitoring. *Med Decis Making* 1981;**1**(2):165-79
253. Claxton K, Thompson KM. A dynamic programming approach to the efficient design of clinical trials. *J Health Econ* 2001;**20**(5):797-822 doi: S0167-6296(01)00093-5 [pii].
254. Department of Health Research and Development Directorate. *Attributing the costs of health and social care Research and Development (AcoRD)*, 2012.
255. Gross CP, Anderson GF, Powe NR. The relation between funding by the National Institutes of Health and the burden of disease. *N Engl J Med* 1999;**340**(24):1881-7
256. Michaud CM, Murray CJ, Bloom BR. Burden of disease--implications for future research. *JAMA* 2001;**285**(5):535-9 doi: jsc00344 [pii].
257. Phelps CE, Parente ST. Priority setting in medical technology and medical practice assessment. *Med Care* 1990;**28**(8):703-23
258. Wetterslev J, Thorlund K, Brok J, et al. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol* 2008;**61**(1):64-75 doi: S0895-4356(07)00147-3 [pii] 10.1016/j.jclinepi.2007.03.013.
259. Buxton M, Hanney S. How can payback from health services research be assessed? *J Health Serv Res Policy* 1996;**1**(1):35-43

260. Townsend J, Buxton M. Cost effectiveness scenario analysis for a proposed trial of hormone replacement therapy. *Health Policy* 1997;**39**(3):181-94 doi: S0168851096008706 [pii].
261. Davies L, Drummond M, Papanikolaou P. Prioritizing investments in health technology assessment. Can we assess potential value for money? *Int J Technol Assess Health Care* 2000;**16**(1):73-91
262. Mitton C, Donaldson C. *Priority setting toolkit : a guide to the use of economics in healthcare decision making*. London: BMJ Books, 2004.
263. Armitage P. Sequential analysis in therapeutic trials. *Annu Rev Med* 1969;**20**:425-30 doi: 10.1146/annurev.me.20.020169.002233.
264. Pocock S. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977;**64**(2):191-9
265. National Institute for Health Research. *Feasibility and Pilot Studies*.
266. Eldridge S, Bond C, Campbell M, et al. Definition and reporting of pilot and feasibility studies. *Trials* 2013;**14**(Suppl 1):O18
267. Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *The Lancet* 2014
268. Jones A, Conroy E, Williamson P, et al. The use of systematic reviews in the planning, design and conduct of randomised trials: a retrospective cohort of NIHR HTA funded trials. *BMC Medical Research Methodology* 2013;**13**(1):50
269. Torgerson D, Campbell M. Unequal randomisation can improve the economic efficiency of clinical trials. *J Health Serv Res Policy* 1997;**2**(2):81-5
270. Cher DJ, Miyamoto J, Lenert LA. Incorporating risk attitude into Markov-process decision models: importance for individual decision making. *Med Decis Making* 1997;**17**(3):340-50
271. Knapp M, Prince M, Albanese E, et al. *Dementia UK: A report into the prevalence and cost of dementia prepared by the Personal Social Services Research Unit (PSSRU) at the London School of Economics and the Institute of Psychiatry at King's College London, for the Alzheimer's Society*. London, 2007.
272. Asthma UK. *Where do we stand? Asthma in the UK today.*, 2004.
273. Thomas M, Murray-Thomas T, Fan T, et al. Prescribing patterns of asthma controller therapy for children in UK primary care: a cross-sectional observational study. *BMC Pulm Med* 2010;**10**:29 doi: 1471-2466-10-29 [pii] 10.1186/1471-2466-10-29.
274. Soriano JB, Kiri VA, Maier WC, et al. Increasing prevalence of asthma in UK primary care during the 1990s. *Int J Tuberc Lung Dis* 2003;**7**(5):415-21
275. Office of National Statistics. *Key Population and Vital Statistics (Correction Notices)*. 2004 Series VS No.31, PP1 No.27. London: Office of National Statistics., 2008.
276. Rubenfeld GD, Caldwell E, Peabody E, et al. Incidence and outcomes of acute lung injury. *N Engl J Med* 2005;**353**(16):1685-93 doi: 353/16/1685 [pii] 10.1056/NEJMoa050333.

APPENDIX A: Calculation of Beneficial Populations and Research Costs

A1.1 Potential beneficial population and research costs: BECCA.

In 2006 there were an estimated 750,000 people with dementia in the UK, projected to rise to 1,450,000 by 2050.²⁷¹ Assuming a linear increase, the estimated prevalence in 2011 is 829,546, with an incidence of 12,727 each year. Assuming 80% of dementia patients have a carer who could potentially benefit from the BECCA intervention, over a ten year period this equates to a total of 790,909, or 769,484 (discounted at 3.5%; Table A-1).

Table A-1: BECCA Beneficial Population

Year	PwD	Carers	df	discounted
0	829545	663636	1.000	663636
1	15909	12727	0.966	12297
2	15909	12727	0.934	11881
3	15909	12727	0.902	11479
4	15909	12727	0.871	11091
5	15909	12727	0.842	10716
6	15909	12727	0.814	10354
7	15909	12727	0.786	10004
8	15909	12727	0.759	9665
9	15909	12727	0.734	9338
10	15909	12727	0.709	9023
		790909		769484

Table A-2 shows the original budget allocations over the five years of the project, divided into a fixed and variable component. The allocation between fixed and variable costs is somewhat arbitrary and therefore required judgement. For example, staff costs may be considered fixed within a particular project as the cost will not vary directly with the number of patients enrolled. However, trials may not recruit at the expected rate, requiring extensions to contracts, thus such costs are not completely independent of sample size. The original budget estimated a total cost of £463,586. Uprating this to current costs, I estimate that the budget would be £673,680 if the trial was commenced in 2011 (Table A-3). Actual expenditure on the BECCA trial was £642,903 (Personal communication, University of East Anglia finance office). Therefore the costs were multiplied by $642903/463586 = 1.39$ to account for this, yielding a total expected cost of $£673,680 * 1.39 = £934,263$. This equates to a fixed cost of £469,731 and variable costs of £2,131 per patient.

Table A-2: BECCA trial original budget

	% FTE	months	year 1	year 2	year 3	year 4	year 5	total
Fixed costs								
RA	100	57	23559	24726	25774	26884	20928	121871
Clerical assistant	20	57	3022	3116	3211	3250	2437	15036
Project management meetings			570	570	570	570	570	2850
Computer	2*2000		4000					2850
Health Economics & proj mgt overheads @ 40%			13000	6000	3000	3000	4000	29000
Total			54783	45549	44149	45758	37281	226370
Variable								
BF facilitator, 1-3	100	36	25774	26884	28064			80722
BF facilitator, 4-5	50	18				14618	7432	22050
Home visits to carers: 4000x£4			3200	3200	3200	3200	3200	16000
Visits to GP practices: 100*£20			400	400	400	400	400	2000
Consumables (printing, phone etc)			750	780	810	560	435	3335
Volunteer OOP expenses			4000	8000	8000	8000	4000	32000
GP reimbursement for time			7000	3000				10000
GP reimbursement for admin costs			15000					15000
MD training for lay volunteers overheads @ 40%			3000	5000	5000	2000		15000
Total			69434	58018	56700	34625	18440	237216
Grand total:								463586

Table A-3: BECCA Budget, 2010 £

	% FTE	months	year 1	year 2	year 3	year 4	year 5	total
Fixed costs								
RA ^a	100	57	34649	35704	36791	37908	39063	184115
Clerical assistant ^b	20	57	4445	4499	4584	4583	4549	22659
Project management meetings ^c			694	694	694	694	694	3468
Computer ^d	2*1000		2000					2000
Health Economics & proj mgt ^b overheads @ 40%			19120	8664	4282	4230	7466	43762
Total			76544	65642	62900	64411	69216	338714
Variable								
BF facilitator, 1-3 ^b	100	36	37907	38820	40060			116787
BF facilitator, 4-5 ^b	50	18				20612	13872	34484
Home visits to carers: 4000x£4 ^c			3894	3894	3894	3894	3894	19472
Visits to GP practices: 100*£20 ^c			487	487	487	487	487	2434
Consumables (printing, phone etc) ^c			913	949	986	682	529	4058.70
Volunteer OOP expenses ^c			4868	9736	9736	9736	4868	38944
GP reimbursement for time ^b			10295	4332				14627.09
GP reimbursement for admin			22061					22061

costs^b

MD training for lay volunteers ^b	4412	7220	7137	2820		21589.48
overheads @ 40%	15163	15528	16024	8245	5549	60508.41
Total	100000	80967	78324	46476	29200	334965.7
Grand total						673680

a. mid point RA starting at 1b4 - spine point 28

b. uprated from original by same proportion as RA

c. uprated by CPI 2001-2010

d. reduced to reflect price fall

A1.2 Potential beneficial population and research costs: ELEVATE.

In 2004 there were an estimated 5.2m people with asthma in the UK,²⁷² of which 90.6% are at 'step 2' (controlled with inhaled corticosteroids).²⁷³ During the period 1990 - 1998, GPRD data suggests the prevalence of asthma in the UK general population rose from approximately 3% to 5%.²⁷⁴ This equates to an increase of approximately 0.025% per annum. Assuming a linear increase, and based on a UK population in 2004 of 59,834,300,²⁷⁵ the estimated prevalence of step 2 patients in 2011 is approximately 5,660,000, with an incidence of 135,500 each year. Over a ten year period therefore, the potential population who could benefit from this trial is 7,015,120 or 6,786,978 (discounted at 3.5%; Table A-4).

Table A-4: ELEVATE Beneficial Population

Year	Patients	df	discounted
0	5659873	1.000	5659873
1	135525	0.966	130942
2	135525	0.934	126514
3	135525	0.902	122236
4	135525	0.871	118102
5	135525	0.842	114108
6	135525	0.814	110249
7	135525	0.786	106521
8	135525	0.759	102919
9	135525	0.734	99439
10	135525	0.709	96076
	7015120		6786978

Table A-5 and Table A-6 show the original and updated budget allocations over the four years of the project, divided into a fixed and variable component. The total original budget for the trial in 2000 was £416,293. In 2011 figures, I estimate this to be £744,525. Actual expenditure on the trial was £840,790, therefore 2010 trial cost estimates were increased by a factor of 2.02 (=840790/416293) to reflect this. This yielded a total fixed cost estimate of £1,305,470 (=646366*2.02) and variable costs of £198,253 (=98159*2.02) for 687 patients, or £289 per patient for a new trial.

Table A-5: ELEVATE trial original budget

	% FTE	months	year 1	year 2	year 3	year 4	total
Fixed costs							
<i>RA x1.5</i>	150	78	34912	37991	41478	8402	122783
<i>Secretary</i>	100	40	15549	16674	17874	9360	59457
<i>Consumables+ computer equip</i>			6100	50	50	250	6450
<i>Statistician</i>	20	40	6295	6803	7352	1910	22360
<i>Project supervision</i>			10342	10342	10342	3447	34473
<i>Steering cttee meetings, travel to mtngs & background work</i>			5560	5561	5561	1853	18535
<i>overheads @ 40%</i>			22702	24587	26682	7869	81840
Total			101460	102008	109339	33091	345898
Variable							
<i>RA practice visits x30</i>			2000	2000	2000		6000
<i>Practice research assistants</i>			9315	9315	9315		27945
<i>GP costs</i>			12150	12150	12150		36450
Total			23465	23465	23465	0	70395
Grand total:							416293

Table A-6: ELEVATE Budget, 2010 £

	% FTE	months	year 1	year 2	year 3	year 4	total
Fixed costs							
<i>RA x1.5^a</i>	150	78	51973.5	53556	55186.5	56862	217578
<i>Secretary^b</i>	100	40	23148	23505	23781	63345	133780
<i>Consumables+ computer equip^c</i>			5816.1	61.55	61.55	307.75	6246.95
<i>Statistician^b</i>	20	40	9371	9590	9782	12926	41669.67
<i>Project supervision^b</i>			15396	14579	13760	23328	67063.49
<i>Steering cttee meetings, travel to mtngs & background work^d</i>			6844	6846	6846	2281	22816.59
<i>overheads @ 40%</i>			33797	34661	35500	53253	157211.1
Total			146346	142798	144917	212304	646365.7

Variable				
<i>RA practice visits x30^d</i>	2462	2462	2462	7386
<i>Practice research assistants^b</i>	13867	13131	12394	39392.23
<i>GP costs^b</i>	18088	17128	16166	51381.17
Total	34417	32721	31021	98159.4
Grand total				744525

a. mid point RA starting at 1b4 - spine point 28

b. uprated from original by same proportion as RA

c. computer price reduced to 1000 each. Other costs increased by CPI

d. costs increased by CPI

A1.3 Potential beneficial population and research costs: CESAR.

The incidence of acute lung injury has been estimated at 86.2 per 100,000 person years (age adjusted).²⁷⁶ Given a projected UK population of 62,761,000 in 2011, this equates to 54,100 patients who could potentially benefit from the information in an additional trial per annum. Over 10 years the total population is 595,100, or 504,028 discounted at 3.5% per annum (Table A-7).

Table A-7: CESAR Beneficial Population

Year	Patients	Df	discounted
0	54100	1.000	54100
1	54100	0.966	52271
2	54100	0.934	50503
3	54100	0.902	48795
4	54100	0.871	47145
5	54100	0.842	45551
6	54100	0.814	44010
7	54100	0.786	42522
8	54100	0.759	41084
9	54100	0.734	39695
10	54100	0.709	38352
	595100		504028

Table A-8 and Table A-9 show the original budget allocations over the five years of the project. divided into a fixed and variable component. The original budget estimated a total cost of £9,466,273. Based on a 2011 start date, I estimate a new trial would have fixed costs of £1,827,720 and variable costs of £65,102 per patient.

Table A-8: CESAR trial original budget

	% FTE	months	year 1	year 2	year 3	year 4	year 5	total
Fixed costs								
<i>Study co-ordinator</i>	50	54	17198	17660	17660	17660	8831	79009
<i>Data manager</i>	50	54	15361	15935	16373	16373	8186	72228
<i>Statistician</i>	20	54	6844	7089	7370	7744	4007	33054
<i>Clinical research fellow</i>	100	39	33207	34919	37260	9586		114972
<i>Data & Admin Officer</i>	50	54	11129	11684	12205	12666	6542	54226
<i>Secretary</i>	20	54	2908	2996	3084	3176	1614	13778
<i>Secretary</i>	10	54	1635	1635	1635	1635	829	7369
<i>Secretary</i>	10	54	1454	1498	1542	1588	807	6889
<i>Researcher</i>	50	54	14741	15624	16562	17556	9205	73688
<i>Economist</i>	50	54	10858	11339	11894	12407	6338	52836
<i>Resp Physician sessional</i>	30	36	2000	12000	12000	10000		36000
<i>Steering cttee</i>								15000
<i>collaborators meetings</i>								49500
<i>consultation</i>								650
<i>DMC</i>								500
<i>conference attendance</i>								2000
<i>7 computers + 1 printer</i>								7500
<i>exceptional items*</i>								9084
<i>Overheads</i>			46934	52951.6	55034	44156.4	18543.6	217619.6
Total								845902.6
Variable								
<i>site visits and resp ass.</i>								32000
<i>consumables</i>								55334
<i>service support costs</i>			186300	253368	253368			693036
<i>excess treatment costs</i>			2240000	2800000	2800000			7840000
Total			2426300	3053368	3053368	0	0	8620370
Grand total:								9466273

*(randomisation, rent, QoL scales)

Table A-9: CESAR Budget, 2010 £

	% FTE	months	year 1	year 2	year 3	year 4	year 5	total
Fixed costs								
<i>Study co-ordinator^a</i>	50	54	34649	35704	36791	37908	39063	184115
<i>Data manager^b</i>	50	54	30948	32216	34110	35145	36210	168630
<i>Statistician^b</i>	20	54	13789	14332	15354	16623	17725	77822
<i>Clinical research fellow^b</i>	100	39	66903	70597	77624	20577		235700.17
<i>Data & Admin Officer^b</i>	50	54	22422	23622	25427	27188	28938	127596.4
<i>Secretary^b</i>	20	54	5859	6057	6425	6817	7139	32297.597
<i>Secretary^b</i>	10	54	3294	3306	3406	3510	3667	17182.389
<i>Secretary^b</i>	10	54	2929	3029	3212	3409	3570	16148.799
<i>Researcher^b</i>	50	54	29699	31588	34504	37685	40717	174192.23
<i>Economist^b</i>	50	54	21876	22925	24779	26632	28035	124246.68
<i>Resp Physician sessional^b</i>	30	36	4029	24261	25000	21465		74755
<i>Steering ctte^c collaborators meetings^c consultation^c DMC^c conference attendance^c 7 computers + 1 printer^d exceptional items^c</i>								18465 60935 800 616 2462 7500 11182
<i>Overheads</i>			94558.45	107054.6	114652.1	94783.74	82025.66	493074.52
Total								1827720.4
Variable								
<i>site visits and resp ass.^c</i>								39392
<i>consumables^c</i>								68116
<i>service support costs^c</i>								853127
<i>excess treatment costs^c</i>								9651040
Total								10611675
Grand total								12439396

a. Mid point RA starting at 1b4 - spine point 28

b. Uprated from original by same proportion as RA

c. Uprated by CPI

d. Assumed same cost as original

APPENDIX B: Correlation Coefficients amongst Data components in the BECCA study

Table B-1 shows the correlation coefficients between sections of the resource use questionnaire used in the BECCA study. 'Intervention' refers to the cost of the intervention. Note that none of the correlation coefficients exceed $|0.478|$.

Table B-1: Correlation Coefficients between cost components of BECCA data

rho	3.1	3.2	3.3	3.4	3.6	3.7	3.8	3.9	3.11	3.12	3.13	4	5
3.1	1.000												
3.2	-0.163	1.000											
3.3	0.069	0.281	1.000										
3.4	0.147	-0.099	0.108	1.000									
3.6	0.033	-0.023	0.056	0.037	1.000								
3.7	-0.061	-0.067	0.017	0.062	0.465	1.000							
3.8	0.032	-0.009	0.121	0.008	0.090	0.046	1.000						
3.9	0.009	-0.003	0.131	0.026	0.035	0.048	0.125	1.000					
3.11	-0.024	0.254	0.329	-0.050	0.053	0.051	0.189	0.139	1.000				
3.12	-0.056	-0.025	-0.041	-0.068	0.044	-0.007	0.000	-0.050	-0.046	1.000			
3.13	0.093	0.006	0.087	0.051	0.005	0.030	0.097	0.239	0.157	-0.029	1.000		
4	0.333	-0.209	0.175	0.200	0.094	0.064	0.129	0.028	0.167	-0.048	0.106	1.000	
5	-0.086	0.011	0.029	-0.008	-0.112	-0.078	0.108	0.034	0.082	-0.039	-0.004	-0.026	1.000
8.1	0.139	-0.010	0.059	0.066	0.060	0.136	0.147	-0.004	0.159	-0.007	0.080	0.154	-0.063
8.3	-0.025	0.094	0.092	-0.048	-0.053	-0.050	0.076	0.137	-0.022	0.029	0.000	0.080	-0.022
8.4	-0.054	0.150	0.038	0.123	0.088	0.143	0.153	-0.072	-0.013	-0.052	0.004	0.025	-0.007
8.5	0.056	-0.037	0.041	0.373	0.118	0.109	0.015	-0.076	-0.022	-0.014	0.001	0.045	-0.005
8.7	0.009	0.033	0.026	0.381	0.074	0.032	0.023	0.025	-0.053	-0.005	-0.015	0.018	-0.018
8.8	-0.041	0.191	0.020	-0.067	0.067	0.050	0.118	0.029	0.233	0.037	0.018	0.027	-0.051
8.9	-0.018	0.017	0.153	0.204	0.107	0.070	0.285	0.096	0.070	-0.011	0.025	0.052	-0.046
8.10	0.007	-0.026	-0.016	0.074	-0.046	0.000	-0.017	-0.035	0.001	-0.041	-0.025	0.065	-0.044
8.11	-0.115	-0.021	0.011	0.003	0.025	0.039	-0.040	0.105	-0.100	-0.045	-0.039	-0.122	0.080
8.12	-0.005	0.148	0.088	0.014	-0.008	0.058	0.207	0.010	0.189	0.029	0.021	0.061	-0.033
8.13	-0.060	0.014	0.105	0.097	0.001	0.026	0.330	0.115	0.033	-0.022	0.002	-0.071	-0.043
9	-0.193	0.026	-0.047	-0.105	-0.067	-0.069	-0.088	-0.056	-0.028	-0.013	-0.032	-0.332	0.005
Intervention	-0.018	-0.052	-0.026	-0.029	-0.001	0.043	0.125	-0.012	-0.033	-0.023	0.007	-0.011	-0.035

Table B-1 continued

rho	8.1	8.3	8.4	8.5	8.7	8.8	8.9	8.10	8.11	8.12	8.13	9	I
3.1													
3.2													
3.3													
3.4													
3.6													
3.7													
3.8													
3.9													
3.11													
3.12													
3.13													
4													
5													
8.1	1.000												
8.3	0.101	1.000											
8.4	0.063	0.028	1.000										
8.5	0.069	0.019	0.402	1.000									
8.7	0.080	0.027	0.217	0.306	1.000								
8.8	0.171	-0.034	0.113	-0.023	0.019	1.000							
8.9	0.149	0.008	0.432	0.153	0.478	0.071	1.000						
8.10	-0.065	-0.018	-0.026	0.068	-0.013	-0.005	-0.011	1.000					
8.11	0.075	-0.033	0.055	0.068	0.041	-0.030	0.094	-0.031	1.000				
8.12	0.331	0.013	0.174	-0.002	0.040	0.031	0.242	-0.016	0.063	1.000			
8.13	0.105	-0.034	0.120	0.089	-0.020	0.200	0.138	0.131	-0.092	0.110	1.000		
9	-0.074	-0.045	0.005	-0.014	0.015	-0.020	-0.010	-0.002	-0.075	-0.043	-0.015	1.000	
Intervention	0.096	0.049	-0.078	-0.038	-0.070	-0.015	-0.001	0.018	-0.097	0.131	0.142	0.097	1.000

APPENDIX C: Full details of calculations for Chapter 4

For source data see Tables 4-1 and Table 4-2 on page 166.

Box C-1: Population EVPI

Population EVPI

$$P_{EVPI} = N \cdot \sigma^* \cdot L_{N^*} \left(\frac{b_0}{\sigma^*} \right)$$

$$= N \cdot \sigma^* \cdot \left(\phi \left(\frac{b_0}{\sigma^*} \right) - \frac{b_0}{\sigma^*} \left[\Phi \left(-\frac{b_0}{\sigma^*} \right) - I\{b_0 < 0\} \right] \right)$$

Expected reduction in standard error of incremental net benefit

$$\sigma^* = \sqrt{v_0 - v_1} = \sqrt{217.15^2 - 0^2} = 217.15$$

Therefore population EVPI

$$= 6,786,978 * 217.15 * (\phi(0.26) - 0.26[\Phi(-0.26) - 0])$$

$$= 6,786,978 * 217.15 * 0.28$$

$$= £416,256,000$$

Box C-2: ENBS, INB at n=8,589 per arm

Expected net benefit of sampling n observations per arm

$$ENBS(n) = (N - 2n) \cdot \sigma_n^* \cdot L_{N^*}(D_n^*) - (K_s + 2k_s n + nb_0)$$

Beneficial population

$$(N - 2n) = (6,786,978 - 2 * 8589) = 6,769,800$$

Expected reduction in standard error of incremental net benefit

$$\sigma_n^* = \sqrt{217.15^2 - \frac{1}{\frac{1}{217.15^2} + \frac{1}{3,097.47^2}}} = 214.62$$

Normalised mean at which to calculate unit normal loss

$$D_n^* = \frac{b_0}{\sigma_n^*} = \frac{56.41}{214.62} = 0.26$$

Unit normal loss

$$L_{N^*}(D_n^*) = (\phi(D_n^*) - D_n^*[\Phi(-D_n^*) - I\{b_0 < 0\}]) = 0.28$$

Cost of sampling

$$K_s + 2k_s n + nb_0$$

$$= £1,305,470 + 2 * 8589 * 288.58 + 8589 * 56.41 = £6,262,667$$

Expected net benefit of sampling

$$\therefore ENBS(8,589) = 6,764,294 * 214.62 * 0.28 - 6,262,667$$

$$\cong £401,859,000$$

Box C-3: EVPPI, QALYs

Population expected value of partial perfect information on QALYs

$$P_{EVPPI_{QALYs}} = N\sigma^*L_{N^*}(D^*, b_0)$$

$$= N \cdot \sqrt{v_0 - v_1} \cdot \left(\phi\left(\frac{b_0}{\sqrt{v_0 - v_1}}\right) - \frac{b_0}{\sqrt{v_0 - v_1}} \left[\Phi\left(-\frac{b_0}{\sqrt{v_0 - v_1}}\right) - I\{b_0 < 0\} \right] \right)$$

Pre-posterior variance of INB and its components

$$v_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda\rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1}$$

$$v(\Delta C)_1 = v(\Delta C)_0 = 78.11^2$$

$$v(\Delta E)_1 = 0$$

$$\rho_{\Delta E, \Delta C, 1} = \rho_{\Delta E, \Delta C, 0} = -0.036$$

$$\therefore v_1 = 0 + 78.11^2 - 0 = \pounds 6101.77$$

Expected reduction in standard error of INB

$$\sqrt{v_0 - v_1} = \sqrt{\pounds 47,155.03 - \pounds 6101.77} = \sqrt{\pounds 41,053.26} = \pounds 202.62$$

EVPPI

$$P_{EVPPI_{QALYs}} = 6,786,978 * 202.62$$

$$* \left(\phi\left(\frac{56.41}{202.62}\right) - \frac{56.41}{202.62} \left[\Phi\left(-\frac{56.41}{202.62}\right) - I\{56.41 < 0\} \right] \right)$$

$$= 6,786,978 * 202.62 * 0.28$$

$$= \pounds 378,296,000$$

Box C-4: ENGS, QALYs @ n=9458 per arm

Expected net gain of sampling data on QALYs, n=9458 per arm

$$ENGS(n) = (N - 2n) \cdot \sigma_n^* \cdot L_{N^*}(D_n^*) - (K_s + 2k_s n + nb_0)$$

Beneficial population

$$(N - 2n) = (6,786,978 - 2 * 9,458) = 6,768,062$$

Expected reduction in standard error of incremental net benefit

$$\sigma_n^* = \sqrt{v_0 - v_1}$$

Preposterior variance of INB and its components

$$v_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1}$$

$$v(\Delta C)_1 = v(\Delta C)_0 = 78.11^2$$

$$v(\Delta E)_1 = \left(\frac{1}{v(\Delta E)_0} + \frac{n}{s(\Delta E)^2} \right)^{-1} = \left(\frac{1}{0.04^2} + \frac{9458}{0.536^2} \right)^{-1} = 0.0000298$$

$$\rho_{\Delta E, \Delta C, 1} = \rho_{\Delta E, \Delta C, 0} = -0.036$$

$$\therefore v_1 = 5000^2 * 0.0000298 + 78.11^2 - 2\lambda(-0.036)\sqrt{0.0000298}\sqrt{78.11^2} = \pounds 6997.80$$

Therefore expected reduction in standard error of INB

$$\sigma_n^* = \sqrt{47,155.03 - 6997.80} = 200.39$$

Normalised mean at which to calculate unit normal loss

$$D_n^* = \frac{b_0}{\sigma_n^*} = \frac{56.41}{200.39} = 0.28$$

Unit normal loss

$$L_{N^*}(D_n^*) = (\phi(D_n^*) - D_n^*[\Phi(-D_n^*) - I\{b_0 < 0\}]) = 0.27$$

Cost of sampling

$$K_s + 2k_s n + nb_0 = \pounds 1,305,470 + 2 * 9458 * 192.39 + 9458 * 56.41 = \pounds 4,944,634$$

Expected net
benefit of
sampling

$$\begin{aligned}\therefore ENBS(9,458) &= 6,768,062 * 200.39 * 0.27 - 4,944,634 \\ &\cong \text{£}365,991,000\end{aligned}$$

Box C-5: EVPPI, Cost

Population expected value of partial perfect information on cost

$$P_{EVPPI_{Cost}} = N\sigma^*L_{N^*}(D^*, b_0)$$

$$= N \cdot \sqrt{v_0 - v_1} \cdot \left(\phi \left(\frac{b_0}{\sqrt{v_0 - v_1}} \right) - \frac{b_0}{\sqrt{v_0 - v_1}} \left[\Phi \left(-\frac{b_0}{\sqrt{v_0 - v_1}} \right) - I\{b_0 < 0\} \right] \right)$$

Pre-posterior variance of INB and its components

$$v_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1}$$

$$v(\Delta E)_1 = v(\Delta E)_0 = 0.04^2$$

$$v(\Delta C)_1 = 0$$

$$\rho_{\Delta E, \Delta C, 1} = \rho_{\Delta E, \Delta C, 0} = -0.036$$

$$\therefore v_1 = 5000^2 0.04^2 + 0 - 0 = \pounds 39,940.91$$

Expected reduction in standard error of INB

$$\sqrt{v_0 - v_1} = \sqrt{\pounds 47,155.03 - \pounds 39,940.91} = \sqrt{\pounds 7,214.12} = \pounds 84.94$$

EVPPI

$$= 6,786,978 * 84.94$$

$$* \left(\phi \left(\frac{56.41}{84.94} \right) - \frac{56.41}{84.94} \left[\Phi \left(-\frac{56.41}{84.94} \right) - I\{56.41 < 0\} \right] \right)$$

$$= 6,786,978 * 84.94 * 0.15$$

$$= \pounds 87,476,000$$

Box C-6: ENGS, Cost @ n=6735 per arm

Expected net gain of sampling data on cost, n=6735 per arm

$$ENGS(n) = (N - 2n) \cdot \sigma_n^* \cdot L_{N^*}(D_n^*) - (K_s + 2k_s n + nb_0)$$

Beneficial population

$$(N - 2n) = (6,786,978 - 2 * 6735) = 6,773,508$$

Expected reduction in standard error of incremental net benefit

$$\sigma_n^* = \sqrt{v_0 - v_1}$$

Preposterior variance of INB and its components

$$v_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1}$$

$$v(\Delta E)_1 = v(\Delta E)_0 = 0.040^2$$

$$v(\Delta C)_1 = \left(\frac{1}{v(\Delta C)_0} + \frac{n}{s(\Delta C)^2} \right)^{-1} \\ = \left(\frac{1}{78.11^2} + \frac{6735}{1049.35^2} \right)^{-1} = \pounds 159.23$$

$$\rho_{\Delta E, \Delta C, 1} = \rho_{\Delta E, \Delta C, 0} = -0.036$$

$$\therefore v_1 = 5000^2 * 0.040^2 + 159.23 - 2\lambda(-0.036) * 0.040\sqrt{159.23} \\ = \pounds 40,279.83$$

Therefore expected reduction in standard error of INB

$$\sigma_n^* = \sqrt{47,155.03 - 40,279.83} = 82.92$$

Normalised mean at which to calculate unit normal loss

$$D_n^* = \frac{b_0}{\sigma_n^*} = \frac{56.41}{82.92} = 0.68$$

Unit normal loss

$$L_{N^*}(D_n^*) = (\phi(D_n^*) - D_n^*[\Phi(-D_n^*) - I\{b_0 < 0\}]) = 0.15$$

Cost of sampling

$$K_s + 2k_s n + nb_0 = \pounds 1,305,470 + 2 * 6,735 * 192.39 + 6,735 * 56.41 \\ = \pounds 3,896,902$$

Expected net benefit of sampling

$$\therefore ENBS_{Cost}(6,735) = 6,773,508 * 82.92 * 0.15 - 3,896,902 \\ \cong \pounds 78,673,000$$

Box C-7: EVPPI, non-drug cost

Population expected value of partial perfect information on non-drug cost

$$P_{EVPPI_{Non-drug\ cost}} = N\sigma^*L_{N^*}(D^*, b_0)$$

$$= N \cdot \sqrt{v_0 - v_1} \cdot \left(\phi\left(\frac{b_0}{\sqrt{v_0 - v_1}}\right) - \frac{b_0}{\sqrt{v_0 - v_1}} \left[\Phi\left(-\frac{b_0}{\sqrt{v_0 - v_1}}\right) - I\{b_0 < 0\} \right] \right)$$

Pre-posterior variance of INB and its components

$$v_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1}$$

$$v(\Delta E)_1 = v(\Delta E)_0 = 0.04^2$$

$$v(\Delta C)_1 = v(\Delta C_n)_1 + v(\Delta C_d^A)_1 + 2\rho_{\Delta C_n, \Delta C_d^A, 1} \sqrt{v(\Delta C_n)_1} \sqrt{v(\Delta C_d^A)_1}$$

$$v(\Delta C_d^A)_1 = v(\Delta C_d^A)_0 = £45.36^2$$

$$v(\Delta C_n)_1 = 0$$

$$\rho_{\Delta C_n, \Delta C_d^A, 1} = \rho_{\Delta C_n, \Delta C_d^A, 0} = 0.352$$

$$\therefore v(\Delta C)_1 = 0 + £45.36^2 + 0 = £45.36^2$$

$$\rho_{\Delta E, \Delta C, 1} = \rho_{\Delta E, \Delta C, 0} = -0.036$$

$$\therefore v_1 = \lambda^2 0.04^2 + £45.36^2 - 2\lambda * (-0.036) \sqrt{0.04^2} \sqrt{£45.36^2}$$

$$= £42,644.50$$

Expected reduction in standard error of INB

$$\sqrt{v_0 - v_1} = \sqrt{£47,155.03 - £42,644.50} = \sqrt{4510.53} = £67.16$$

EVPPI

$$= 6,786,978 * 67.16$$

$$* \left(\phi\left(\frac{56.41}{67.16}\right) - \frac{56.41}{67.16} \left[\Phi\left(-\frac{56.41}{67.16}\right) - I\{56.41 < 0\} \right] \right)$$

$$= 6,786,978 * 67.16 * 0.112$$

$$= £51,037,000$$

Box C-8: ENGS, non-drug cost @ n=9456

Expected net gain of sampling data on non-drug cost, n=9456 per arm

$$ENGS(n) = (N - 2n) \cdot \sigma_n^* \cdot L_{N^*}(D_n^*) - (K_s + 2k_s n + nb_0)$$

Beneficial population $(N - 2n) = (6,786,978 - 2 * 9456) = 6,768,066$

Expected reduction in standard error of incremental net benefit

$$\sigma_n^* = \sqrt{v_0 - v_1}$$

Preposterior variance of INB and its components

$$v_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1}$$

$$v(\Delta E)_1 = v(\Delta E)_0 = 0.040^2$$

$$v(\Delta C)_1 = v(\Delta C_n)_1 + v(\Delta C_d^A)_1 + 2\rho_{\Delta C_n, \Delta C_d^A, 1} \sqrt{v(\Delta C_n)_1} \sqrt{v(\Delta C_d^A)_1}$$

$$v(\Delta C_d^A)_1 = v(\Delta C_d^A)_0 = £45.36^2$$

$$v(\Delta C_n)_1 = \left(\frac{1}{v(\Delta C_n)_0} + \frac{n}{s(\Delta C_n)^2} \right)^{-1} = \left(\frac{1}{49.60^2} + \frac{6735}{666.75^2} \right)^{-1} = £46.13$$

$$\rho_{\Delta C_n, \Delta C_d^A, 1} = \rho_{\Delta C_n, \Delta C_d^A, 0} = 0.352$$

$$\therefore v(\Delta C)_1 = £46.13 + £45.36^2 + 2(0.352)\sqrt{£46.13}\sqrt{£45.36^2} = £2320.74$$

$$\rho_{\Delta E, \Delta C, 1} = \rho_{\Delta E, \Delta C, 0} = -0.036$$

$$\therefore v_1 = 5000^2 * 0.040^2 + 2320.74 - 2\lambda(-0.036) * 0.040\sqrt{2320.74} = £42,947.65$$

Therefore expected reduction in standard error of INB

$$\sigma_n^* = \sqrt{47,155.03 - 42,947.65} = 64.86$$

Normalised mean at which to calculate unit normal loss

$$D_n^* = \frac{b_0}{\sigma_n^*} = \frac{56.41}{64.86} = 0.87$$

Unit normal loss

$$L_{N^*}(D_n^*) = (\phi(D_n^*) - D_n^*[\Phi(-D_n^*) - I\{b_0 < 0\}]) = 0.11$$

Cost of sampling $K_s + 2k_s n + nb_0 = \pounds 1,305,470 + 2 * 9,456 * 96.19 + 9,456 * 56.41$
 $= \pounds 3,124,667$

Expected net benefit of sampling $\therefore ENBS_{Non-drug\ cost}(9,456) = 6,773,508 * 64.86 * 0.11 - 3,124,667$
 $\cong \pounds 42,934,000$

Box C-9: EVPPI, Drug Cost

Population expected value of partial perfect information on non-drug cost

$$P_{EVPPI_{drugs\ cost}} = N\sigma^*L_{N^*}(D^*, b_0)$$

$$= N \cdot \sqrt{v_0 - v_1} \cdot \left(\phi\left(\frac{b_0}{\sqrt{v_0 - v_1}}\right) - \frac{b_0}{\sqrt{v_0 - v_1}} \left[\Phi\left(-\frac{b_0}{\sqrt{v_0 - v_1}}\right) - I\{b_0 < 0\} \right] \right)$$

Pre-posterior variance of INB and its components

$$v_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1}$$

$$v(\Delta E)_1 = v(\Delta E)_0 = 0.04^2$$

$$v(\Delta C)_1 = v(\Delta C_n)_1 + v(\Delta C_d^A)_1 + 2\rho_{\Delta C_n, \Delta C_d^A, 1} \sqrt{v(\Delta C_n)_1} \sqrt{v(\Delta C_d^A)_1}$$

$$v(\Delta C_n)_1 = v(\Delta C_n)_0 = £49.60^2$$

$$v(\Delta C_d^A)_1 = 0$$

$$\rho_{\Delta C_n, \Delta C_d, 1} = \rho_{\Delta C_n, \Delta C_d, 0} = 0.352$$

$$\therefore v(\Delta C)_1 = £49.60^2 + 0 + 0 = £49.60^2$$

$$\rho_{\Delta E, \Delta C, 1} = \rho_{\Delta E, \Delta C, 0} = -0.036$$

$$\therefore v_1 = \lambda^2 0.04^2 + £49.60^2 - 2\lambda * (-0.036) \sqrt{0.04^2} \sqrt{£49.60^2}$$

$$= £43,106.96$$

Expected reduction in standard error of INB

$$\sqrt{v_0 - v_1} = \sqrt{£47,155.03 - £43,106.96} = \sqrt{4048.07} = £63.62$$

EVPPI

$$= 6,786,978 * 63.62$$

$$* \left(\phi\left(\frac{56.41}{63.62}\right) - \frac{56.41}{63.62} \left[\Phi\left(-\frac{56.41}{63.62}\right) - I\{56.41 < 0\} \right] \right)$$

$$= 6,786,978 * 63.62 * 0.103$$

$$= £44,439,000$$

Box C-10: ENGS, drug cost @ n=9197

Expected net gain of sampling data on drug cost $ENGS(n) = (N - 2n) \cdot \sigma_n^* \cdot L_{N^*}(D_n^*) - (K_s + 2k_s n + nb_0)$

Beneficial population $(N - 2n) = (6,786,978 - 2 * 9197) = 6,768,584$

Expected reduction in standard error of incremental net benefit

$$\sigma_n^* = \sqrt{v_0 - v_1}$$

Preposterior variance of INB and its components

$$v_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1}$$

$$v(\Delta E)_1 = v(\Delta E)_0 = 0.040^2$$

$$v(\Delta C)_1 = v(\Delta C_n)_1 + v(\Delta C_d^A)_1 + 2\rho_{\Delta C_n, \Delta C_d^A, 1} \sqrt{v(\Delta C_n)_1} \sqrt{v(\Delta C_d^A)_1}$$

$$v(\Delta C_n)_1 = v(\Delta C_n)_0 = £49.60^2$$

$$v(\Delta C_d^A)_1 = \left(\frac{1}{v(\Delta C_d^A)_0} + \frac{n}{s(\Delta C_d^A)^2} \right)^{-1}$$

$$= \left(\frac{1}{45.36^2} + \frac{6735}{608.03^2} \right)^{-1} = £39.43$$

$$\rho_{\Delta C_n, \Delta C_d^A, 1} = \rho_{\Delta C_n, \Delta C_d^A, 0} = 0.352$$

$$\therefore v(\Delta C)_1 = £49.60^2 + £39.43 + 2(0.352)\sqrt{£49.60^2}\sqrt{£39.43}$$

$$= £2718.53$$

$$\rho_{\Delta E, \Delta C, 1} = \rho_{\Delta E, \Delta C, 0} = -0.036$$

$$\therefore v_1 = 5000^2 * 0.040^2 + 2718.53 - 2 * 5000(-0.036)$$

$$* 0.040\sqrt{2718.53} = £43,401.92$$

Therefore expected reduction in standard error of INB

$$\sigma_n^* = \sqrt{47,155.03 - 43,401.92} = 61.26$$

Normalised mean at which to calculate unit normal loss

$$D_n^* = \frac{b_0}{\sigma_n^*} = \frac{56.41}{61.26} = 0.92$$

Unit normal loss	$L_{N^*}(D_n^*) = (\phi(D_n^*) - D_n^*[\Phi(-D_n^*) - I\{b_0 < 0\}]) = 0.10$
Cost of sampling	$K_s + 2k_s n + nb_0 = \pounds 1,305,470 + 2 * 9,197 * 96.19 + 9,197 * 56.41$ $= \pounds 3,074,839$
Expected net benefit of sampling	$\therefore ENBS_{drug\ cost}(9,197) = 6,773,508 * 61.26 * 0.10 - 3,074,839$ $\cong \pounds 36,485,000$

Box C-11: ENGS, drug cost, two processes @ $n_a=9081$, $n_b=240$

Prior mean incremental cost of drugs from processes A and B

$$\begin{bmatrix} (\Delta C_d^A)_0 \\ (\Delta C_d^B)_0 \end{bmatrix} = \begin{bmatrix} \pounds 102.54 \\ \pounds 289.82 \end{bmatrix}$$

Prior variance/covariance matrix

$$\mathbf{V}' = \begin{bmatrix} \pounds 2,058 & \pounds 1,806 \\ \pounds 1,806 & \pounds 2,328 \end{bmatrix}$$

Inverse of prior matrix

$$\mathbf{H}' = \mathbf{V}'^{-1} = \frac{1}{2058 * 2328 - 1806^2} \begin{bmatrix} 2328 & -1806 \\ -1806 & 2058 \end{bmatrix} \\ = \begin{bmatrix} 0.0015 & -0.0012 \\ -0.0012 & 0.0013 \end{bmatrix}$$

Inverse of sample var/covar matrix

$$\mathbf{H} = \begin{bmatrix} n_A/v_A & 0 \\ 0 & n_B/v_B \end{bmatrix} = \begin{bmatrix} 9081/369,700 & 0 \\ 0 & 240/414,702 \end{bmatrix} \\ = \begin{bmatrix} 0.0246 & 0 \\ 0 & 0.0006 \end{bmatrix}$$

Inverse of pre-posterior var/covar matrix

$$\mathbf{H}'' = \mathbf{H}' + \mathbf{H} = \begin{bmatrix} 0.0261 & -0.0012 \\ -0.0012 & 0.0019 \end{bmatrix}$$

Pre-posterior var/covar matrix

$$\mathbf{V}'' = \mathbf{H}''^{-1} = \frac{1}{0.0261 * 0.0019 - 0.0012^2} \begin{bmatrix} 0.0019 & 0.0012 \\ 0.0012 & 0.0261 \end{bmatrix} \\ = \begin{bmatrix} \pounds 39.43 & \pounds 24.20 \\ \pounds 24.20 & \pounds 534.58 \end{bmatrix}$$

Therefore pre-posterior variance of incremental cost of drugs using process A

$$\therefore v(\Delta C_d^A)_1 = \mathbf{V}''_{11} = \pounds 39.43$$

Expected net gain of sampling data on drug cost (n_A, n_B) observations with each process per arm

$$ENGS(n_u, n_b) = (N - 2(n_u + n_b)) \cdot (\sigma^* L_N^*(D^*) - (k_{su}n_u + k_{sb}n_b + K_S I\{n_u > 0 \cup n_b > 0\} + (n_u + n_b)b_0))$$

Beneficial population

$$(N - 2(n_u + n_b)) = (6,786,978 - 2 * (9081 + 240)) = 6,768,336$$

Expected reduction in standard error of incremental net

$$\sigma_n^* = \sqrt{v_0 - v_1}$$

benefit

Preposterior
variance of INB and
its components

$$v_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1}$$

$$v(\Delta E)_1 = v(\Delta E)_0 = 0.040^2$$

$$v(\Delta C)_1 = v(\Delta C_n)_1 + v(\Delta C_d^A)_1 + 2\rho_{\Delta C_n, \Delta C_d^A, 1} \sqrt{v(\Delta C_n)_1} \sqrt{v(\Delta C_d^A)_1}$$

$$v(\Delta C_n)_1 = v(\Delta C_n)_0 = £49.60^2$$

$$v(\Delta C_d^A)_1 = £39.43$$

$$\rho_{\Delta C_n, \Delta C_d^A, 1} = \rho_{\Delta C_n, \Delta C_d^A, 0} = 0.352$$

$$\begin{aligned} \therefore v(\Delta C)_1 &= £49.60^2 + £39.43 + 2(0.352) \sqrt{£49.60^2} \sqrt{£39.43} \\ &= £2718.55 \end{aligned}$$

$$\rho_{\Delta E, \Delta C, 1} = \rho_{\Delta E, \Delta C, 0} = -0.036$$

$$\begin{aligned} \therefore v_1 &= \lambda^2 * 0.040^2 + 2718.55 - 2\lambda * (-0.036) * 0.040 \sqrt{2718.55} \\ &= £43,401.93 \end{aligned}$$

Therefore expected
reduction in
standard error of
INB

$$\sigma_n^* = \sqrt{47,155.03 - 43,401.93} = 61.26$$

Normalised mean
at which to
calculate unit
normal loss

$$D_n^* = \frac{b_0}{\sigma_n^*} = \frac{56.41}{61.26} = 0.92$$

Unit normal loss

$$L_{N^*}(D_n^*) = (\phi(D_n^*) - D_n^*[\Phi(-D_n^*) - I\{b_0 < 0\}]) = 0.10$$

Cost of sampling

$$\begin{aligned} k_{SA}n_A + k_{SB}n_B + K_S I\{n_A > 0 \cup n_B > 0\} + (n_A + n_B)b_0 \\ = 96.19 * 9081 + 9.62 * 240 + £1,305,470 * 1 \\ + (9081 + 240) * 56.41 = £3,057,140 \end{aligned}$$

Expected net
benefit of sampling

$$\begin{aligned} \therefore ENBS(9081, 240) &= 6,778,142 * 61.26 * 0.10 - 3,057,140 \\ &\cong £36,494,000 \end{aligned}$$

Box C-12: ENGS of optimal trial design

Prior mean incremental cost of drugs from processes A and B

$$\begin{bmatrix} (\Delta C_d^A)_0 \\ (\Delta C_d^B)_0 \end{bmatrix} = \begin{bmatrix} \pounds 102.54 \\ \pounds 289.82 \end{bmatrix}$$

Prior variance/covariance matrix

$$\mathbf{V}' = \begin{bmatrix} \pounds 2,058 & \pounds 1,806 \\ \pounds 1,806 & \pounds 2,328 \end{bmatrix}$$

Inverse of prior matrix

$$\mathbf{H}' = \mathbf{V}'^{-1} = \frac{1}{2058 * 2328 - 1806^2} \begin{bmatrix} 2328 & -1806 \\ -1806 & 2058 \end{bmatrix} \\ = \begin{bmatrix} 0.0015 & -0.0012 \\ -0.0012 & 0.0013 \end{bmatrix}$$

Inverse of sample var/covar matrix

$$\mathbf{H} = \begin{bmatrix} n_A/v_A & 0 \\ 0 & n_B/v_B \end{bmatrix} = \begin{bmatrix} 3693/608.03^2 & 0 \\ 0 & 904/643.97^2 \end{bmatrix} \\ = \begin{bmatrix} 0.010 & 0 \\ 0 & 0.002 \end{bmatrix}$$

Inverse of pre-posterior var/covar matrix

$$\mathbf{H}'' = \mathbf{H}' + \mathbf{H} = \begin{bmatrix} 0.012 & -0.001 \\ -0.001 & 0.004 \end{bmatrix}$$

Pre-posterior var/covar matrix

$$\mathbf{V}'' = \mathbf{H}''^{-1} = \frac{1}{0.012 * 0.004 - 0.001^2} \begin{bmatrix} 0.004 & 0.001 \\ 0.001 & 0.012 \end{bmatrix} \\ = \begin{bmatrix} \pounds 89.96 & \pounds 30.13 \\ \pounds 30.13 & \pounds 293.76 \end{bmatrix}$$

Therefore pre-posterior variance of incremental cost of drugs using process A

$$\therefore v(\Delta C_d^A)_1 = \pounds 89.96$$

Expected net gain of sampling data on drug cost (n_A, n_B) observations with each process per arm, n_E observations on QALYs and n_n on non-drug costs

$$ENGS(n_{\Delta E}, n_{\Delta C_n}, n_{\Delta C_d^A}, n_{\Delta C_d^B}) \\ = [N - 2 * \max(n_{\Delta E}, n_{\Delta C_n}, n_{\Delta C_d^A} + n_{\Delta C_d^B})] * [\sigma^* L_N^*(D^*) \\ - (k_{s\Delta E} n_{\Delta E} + k_{s\Delta C_n} n_{\Delta C_n} + k_{s\Delta C_d^A} n_{\Delta C_d^A} + k_{s\Delta C_d^B} n_{\Delta C_d^B} \\ + K_s I \{n_{\Delta E} > 0 \cup n_{\Delta C_n} > 0 \cup n_{\Delta C_d^A} > 0 \cup n_{\Delta C_d^B} > 0\} \\ + \max(n_{\Delta E}, n_{\Delta C_n}, n_{\Delta C_d^A} + n_{\Delta C_d^B}) b_0]$$

Beneficial population

$$N - 2 * \max(n_{\Delta E}, n_{\Delta C_n}, n_{\Delta C_d^A} + n_{\Delta C_d^B}) \\ = (6,786,978 - 2 * \max(10,787, 4,264, 3,693 \\ + 904) = 6,765,404$$

Expected reduction
in standard error of
incremental net
benefit

$$\sigma_n^* = \sqrt{v_0 - v_1}$$

Preposterior
variance of INB and
its components

$$v_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E, \Delta C, 1} \sqrt{v(\Delta E)_1} \sqrt{v(\Delta C)_1}$$

$$v(\Delta E)_1 = \left(1/0.04^2 + 10,787/0.536^2\right)^{-1} = 0.00003$$

$$v(\Delta C)_1 = v(\Delta C_n)_1 + v(\Delta C_d^A)_1 + 2\rho_{\Delta C_n, \Delta C_d^A, 1} \sqrt{v(\Delta C_n)_1} \sqrt{v(\Delta C_d^A)_1}$$

$$v(\Delta C_n)_1 = \left(1/49.60^2 + 4264/666.75^2\right)^{-1} = 100.02$$

$$v(\Delta C_d^A)_1 = V_{11}^{\Delta} = \text{£}89.96$$

$$\rho_{\Delta C_n, \Delta C_d^A, 1} = \rho_{\Delta C_n, \Delta C_d^A, 0} = 0.352$$

$$\therefore v(\Delta C)_1 = 100.02 + 89.96 + 2(0.352)\sqrt{100.02}\sqrt{89.96} = \text{£}256.78$$

$$\rho_{\Delta E, \Delta C, 1} = \rho_{\Delta E, \Delta C, 0} = -0.036$$

$$\begin{aligned} \therefore v_1 &= \lambda^2 0.00003 + 256.78 - 2\lambda(-0.036)\sqrt{0.00003}\sqrt{256.78} \\ &= \text{£}939.99 \end{aligned}$$

Therefore expected
reduction in
standard error of
INB

$$\sigma_n^* = \sqrt{\text{£}47,155.03 - \text{£}939.99} = 214.98$$

Normalised mean
at which to
calculate unit
normal loss

$$D_n^* = \frac{b_0}{\sigma_n^*} = \frac{56.41}{214.98} = 0.26$$

Unit normal loss

$$L_{N^*}(D_n^*) = (\phi(D_n^*) - D_n^*[\Phi(-D_n^*) - I\{b_0 < 0\}]) = 0.28$$

$$\begin{aligned}
\text{Cost of sampling} & k_{SE}n_E + k_{SC_n}n_{C_n} + k_{SC_d^A}n_{C_d^A} + k_{SC_d^B}n_{C_d^B} \\
& + K_s I \{n_E > 0 \cup n_{C_n} > 0 \cup n_{C_d^A} > 0 \cup n_{C_d^B} > 0\} \\
& + \max(n_E, n_{C_n}, n_{C_d^A}, n_{C_d^B}) b_0 \\
& = 2 * (96.19 * 10787 + 96.19 * 4264 + 96.19 \\
& * 3693 + 9.62 * 904) + \pounds 1,305,470 + (10787) \\
& * 56.41 = \pounds 5,537,462
\end{aligned}$$

$$\begin{aligned}
\text{Expected net} & \therefore ENBS(10787, 4264, 3693, 904) \\
\text{benefit of sampling} & = 6,765,404 * 214.98 * 0.28 - 5,537,462 \\
& \cong \pounds 403,722,000
\end{aligned}$$

APPENDIX D: Exploration of impact of reducing uncertainty in the correlation coefficient

In Chapter 4 I made a simplifying assumption that trialists can choose to measure the cost of drugs using either process A or process B, but not both together. In this appendix I explain why this simplifying assumption is necessary, and why relaxing it is not so straightforward to incorporate into the analysis.

Allowing both data processes to be observed in the same observations would change the algebra of section 4.2 as follows:

$$E \begin{bmatrix} \Delta C_d^A \\ \Delta C_d^B \end{bmatrix} = \begin{bmatrix} \Delta C_{d,0}^A \\ \Delta C_{d,0}^B \end{bmatrix}, \quad V \begin{bmatrix} \Delta C_d^A \\ \Delta C_d^B \end{bmatrix} = \begin{bmatrix} V(\Delta C_d^A)_0 & Cov(\Delta C_d^A, \Delta C_d^B)_0 \\ Cov(\Delta C_d^A, \Delta C_d^B)_0 & V(\Delta C_d^B)_0 \end{bmatrix} \quad [D-1]$$

$$E \left(\begin{bmatrix} \Delta C_{d,s}^A \\ \Delta C_{d,s}^B \end{bmatrix} \middle| \begin{bmatrix} \Delta C_{d,0}^A \\ \Delta C_{d,0}^B \end{bmatrix} \right) = \begin{bmatrix} \Delta C_{d,0}^A \\ \Delta C_{d,0}^B \end{bmatrix},$$

$$V \left(\begin{bmatrix} \Delta C_{d,s}^A \\ \Delta C_{d,s}^B \end{bmatrix} \middle| \begin{bmatrix} \Delta C_{d,0}^A \\ \Delta C_{d,0}^B \end{bmatrix} \right) = \begin{bmatrix} v(\Delta C_d^A)_s/n_A & Cov(\Delta C_d^A, \Delta C_d^B)_s/n_{A \cap B} \\ Cov(\Delta C_d^A, \Delta C_d^B)_s/n_{A \cap B} & v(\Delta C_d^B)_s/n_B \end{bmatrix} \quad [D-2]$$

The joint posterior variance/covariance matrix is the inverse of \mathbf{H}' thus the posterior distribution is summarised in Equations [D-6] and [D-7], where \mathbf{m} is the mean from each data process (Equation [4-12]).

$$\mathbf{H}' = \begin{bmatrix} H'_{11} & H'_{12} \\ H'_{21} & H'_{22} \end{bmatrix} = \begin{bmatrix} v(\Delta C_d^A)_0 & Cov(\Delta C_d^A, \Delta C_d^B)_0 \\ Cov(\Delta C_d^A, \Delta C_d^B)_0 & v(\Delta C_d^B)_0 \end{bmatrix}^{-1} \quad [D-3]$$

$$= \frac{1}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2} \begin{bmatrix} v(\Delta C_d^B)_0 & -Cov(\Delta C_d^A, \Delta C_d^B)_0 \\ -Cov(\Delta C_d^A, \Delta C_d^B)_0 & v(\Delta C_d^A)_0 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} n_A/v_A & n_{A \cap B}/Cov(\Delta C_d^A, \Delta C_d^B)_s \\ n_{A \cap B}/Cov(\Delta C_d^A, \Delta C_d^B)_s & n_B/v_B \end{bmatrix} \quad [D-4]$$

$$\begin{aligned} \mathbf{H}'' &= \mathbf{H}' + \mathbf{H} \\ &= \begin{bmatrix} H'_{11} + n_A/v_A & H'_{12} + n_{A \cap B}/Cov(\Delta C_d^A, \Delta C_d^B)_s \\ H'_{21} + n_{A \cap B}/Cov(\Delta C_d^A, \Delta C_d^B)_s & H'_{22} + n_B/v_B \end{bmatrix} \end{aligned} \quad [D-5]$$

$$\begin{bmatrix} \Delta C_{d,1}^A \\ \Delta C_{d,1}^B \end{bmatrix} = \mathbf{H}''^{-1} \left(\mathbf{H}' \begin{bmatrix} \Delta C_{d,0}^A \\ \Delta C_{d,0}^B \end{bmatrix} + \mathbf{H}\mathbf{m} \right) \quad [D-6]$$

$$\mathbf{V}'' = \begin{bmatrix} v(\Delta C_d^A)_1 & Cov(\Delta C_d^A, \Delta C_d^B)_1 \\ Cov(\Delta C_d^A, \Delta C_d^B)_1 & v(\Delta C_d^B)_1 \end{bmatrix} = \mathbf{H}''^{-1} \quad [D-7]$$

$$\mathbf{m} = \begin{bmatrix} m_A \\ m_B \end{bmatrix} \quad [D-8]$$

As previously, the the problem is to choose (n_A, n_B) that maximises the ENGS (Equation [D-9]).

$$\begin{aligned} ENGS_{n_A, n_B} &= (N - 2(n_A + n_B)) \cdot \sigma^* L_{N^*}(D^*, b_0) \\ &\quad - [k_{sA}n_A + k_{sB}n_B + K_s I\{n_A > 0 \cup n_B > 0\} + (n_A + n_B)b_0] \end{aligned} \quad [D-9]$$

where:

b_0 = prior mean incremental net benefit

$\sigma^* = \sqrt{v_0 - v_1}$ = expected reduction in standard error of incremental net benefit

$$D^* = \frac{|b_0|}{\sigma^*}$$

$L_{N^*}(\cdot)$ = unit normal linear loss integral

The posterior estimate of the variance of ΔC_d^A (cell 1,1 of matrix \mathbf{V}'') is shown in Box D-1 (based on equations [D-3] - [D-7]).

Box D-1: Preposterior variance of mean of Process A

$$\begin{aligned} v(\Delta C_d^A)_1 &= V''_{11} = H''_{11}^{-1} \\ &= \frac{H'_{22} + \frac{n_B}{v(\Delta C_d^B)_s}}{\left(H'_{11} + \frac{n_A}{v(\Delta C_d^A)_s} \right) \left(H'_{22} + \frac{n_B}{v(\Delta C_d^B)_s} \right) - \left(H'_{12} + \frac{n_{A \cap B}}{Cov(\Delta C_d^A, \Delta C_d^B)_s} \right) \left(H'_{21} + \frac{n_{A \cap B}}{Cov(\Delta C_d^A, \Delta C_d^B)_s} \right)} \end{aligned}$$

$$\text{Noting that } H'_{12} = H'_{21} = \frac{-Cov(\Delta C_d^A, \Delta C_d^B)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2}$$

thus:

$$v(\Delta C_d^A)_1 = \frac{H'_{22} + \frac{n_B}{v(\Delta C_d^B)_s}}{\left(H'_{11} + \frac{n_A}{v(\Delta C_d^A)_s}\right)\left(H'_{22} + \frac{n_B}{v(\Delta C_d^B)_s}\right) - \left(\frac{-Cov(\Delta C_d^A, \Delta C_d^B)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0} + \frac{n_{A \cap B}}{Cov(\Delta C_d^A, \Delta C_d^B)_s}\right)^2}$$

Substituting in equations for H'_{11} and H'_{22} :

$$v(\Delta C_d^A)_1 = \frac{\frac{v(\Delta C_d^A)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0} + \frac{n_B}{v(\Delta C_d^B)_s}}{\left(\frac{v(\Delta C_d^B)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0} + \frac{n_A}{v(\Delta C_d^A)_s}\right)\left(\frac{v(\Delta C_d^A)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0} + \frac{n_B}{v(\Delta C_d^B)_s}\right) - \left(\frac{-Cov(\Delta C_d^A, \Delta C_d^B)_0}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0} + \frac{n_{A \cap B}}{Cov(\Delta C_d^A, \Delta C_d^B)_s}\right)^2}$$

This is more clearly written by noting that:

$$|\mathbf{H}'| = v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0^2$$

Thus:

$$v(\Delta C_d^A)_1 = \frac{\frac{v(\Delta C_d^A)_0}{|\mathbf{H}'|} + \frac{n_B}{v(\Delta C_d^B)_s}}{\left(\frac{v(\Delta C_d^B)_0}{|\mathbf{H}'|} + \frac{n_A}{v(\Delta C_d^A)_s}\right)\left(\frac{v(\Delta C_d^A)_0}{|\mathbf{H}'|} + \frac{n_B}{v(\Delta C_d^B)_s}\right) - \left(\frac{-Cov(\Delta C_d^A, \Delta C_d^B)_0}{|\mathbf{H}'|} + \frac{n_{A \cap B}}{Cov(\Delta C_d^A, \Delta C_d^B)_s}\right)^2}$$

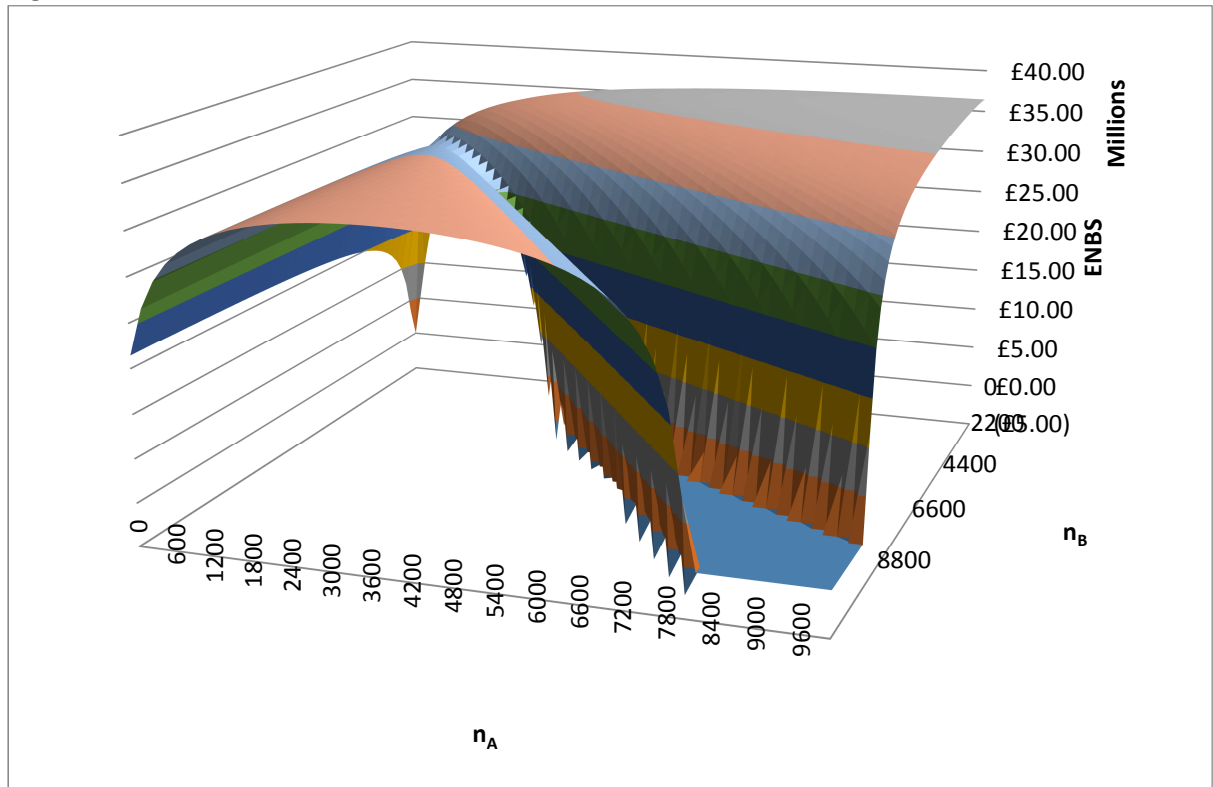
Results

This analysis identifies an optimal result at 8816 observations in each arm using process A, and of these, process B should also be measured in 344. However, this does not make intuitive sense: given that process A is the 'preferred' process, there can be no benefit from observing process B in a subset of observations on which process A is being measured: the results would only make intuitive sense if the numbers of observations on process B were greater than process A, and A was measured in a subset of those, i.e. if $n_B > n_A$.

Exploration of this in more detail suggests that the reduction in uncertainty is being overestimated as for some combinations of sample sizes, the ENGS is undefined (figure D-1, undefined values plotted as zeroes). The reason for this is that the estimated reduction in variance is greater than the prior variance, leading to a negative preposterior variance and thus an error in the resulting calculation of EVSI and hence ENGS.

This is most easily demonstrated for situations where an equal number of observations are proposed using each process. Box D-2 below shows the calculations, with the resulting negative preposterior variance. This flaw appears to have been identified by Pratt, Raiffa and Schalifer,⁴ but unfortunately they do not elaborate on this, simply stating that “another technique will be required” to analyse such data. For the purpose of this thesis then, I limit the analysis to situations where either process A or process B are observed in a particular patient. Whilst limited in applicability, this is mathematically valid. Relaxing this assumption and estimating the expected return on investment when both processes are observed in the same patient remains an important area for future research.

Figure D-1: ENGS for combinations of observations collected on n_A and n_B



Box D-2: Estimate of preposterior variance when $n_A=n_B=500$

Prior mean
incremental cost of
drugs from
processes A and B

$$\begin{bmatrix} (\Delta C_d^A)_0 \\ (\Delta C_d^B)_0 \end{bmatrix} = \begin{bmatrix} \pounds 102.54 \\ \pounds 289.82 \end{bmatrix}$$

Prior
variance/covariance
matrix

$$V' = \begin{bmatrix} \pounds 2,058 & \pounds 1,806 \\ \pounds 1,806 & \pounds 2,328 \end{bmatrix}$$

Inverse of prior
matrix

$$H' = V'^{-1} = \frac{1}{2058 * 2328 - 1806^2} \begin{bmatrix} 2328 & -1806 \\ -1806 & 2058 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0015 & -0.0012 \\ -0.0012 & 0.0013 \end{bmatrix}$$

Inverse of sample
var/covar matrix

$$H = \begin{bmatrix} n_A/v_A & n_{A \cap B}/Cov(\Delta C_d^A, \Delta C_d^B)_s \\ n_{A \cap B}/Cov(\Delta C_d^A, \Delta C_d^B)_s & n_B/v_B \end{bmatrix}$$

$$= \begin{bmatrix} 5000/369,700 & 5000/320,990 \\ 5000/320,990 & 5000/414,702 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0135 & 0.0156 \\ 0.0156 & 0.0121 \end{bmatrix}$$

Inverse of pre-
posterior var/covar
matrix

$$H'' = H' + H = \begin{bmatrix} 0.0150 & 0.0144 \\ 0.0144 & 0.0134 \end{bmatrix}$$

Pre-posterior
var/covar matrix

$$V'' = H''^{-1} = \frac{1}{0.0150 * 0.0134 - 0.0144^2} \begin{bmatrix} 0.0150 & 0.0144 \\ 0.0144 & 0.0134 \end{bmatrix}$$

$$= \begin{bmatrix} -\pounds 2400.18 & -\pounds 2578.17 \\ -\pounds 2578.17 & -\pounds 2694.73 \end{bmatrix}$$

Therefore pre-
posterior variance
of incremental cost
of drugs using
process A

$$\therefore v(\Delta C_d^A)_1 = V''_{11} = -\pounds 2400.18$$

APPENDIX E: Power calculation

The formula for determining the sample size for a study with two independent outcomes and a continuous outcome measure is:

$$n_i = 2 \left(\frac{Z\sigma}{E} \right)^2$$

Where:

n_i = number of patients per arm

Z = standard normal distribution evaluated at desired confidence level.

E = minimally important difference in outcome measure between treatment groups.

$$\sigma \cong s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}}$$

If the minimally important difference in net benefit is £100, and 95% confidence is desired, the appropriate sample size, based on data reported in Table 4-1 and Table 4-2 is:

$$\sigma \cong s_p = \sqrt{\frac{(175 - 1)2010.64^2 + (184 - 1)2356.20^2}{(175 + 184 - 2)}} = 2194.58$$

$$n_i = 2 \left(\frac{1.96 * 2194.58}{100} \right)^2 \cong 3700$$