ECONOMIC EVALUATION OF HEALTH CARE TECHNOLOGIES: A COMPARISON OF ALTERNATIVE DECISION MODELLING TECHNIQUES

A thesis submitted for the degree of Doctor of Philosophy

by

Jonathan Daniel Karnon

Health Economics Research Group, Department of Economics and Finance, Brunel University

March 2001

Abstract

The focus of this thesis is on the application of decision models to the economic evaluation of health care technologies. The primary objective addresses the correct choice of modelling technique, as the attributes of the chosen technique could have a significant impact on the process, as well as the results, of an evaluation. Separate decision models, a Markov process and a discrete event simulation (DES) model are applied to a case study evaluation comparing alternative adjuvant therapies for early breast cancer. The case study models are built and analysed as stochastic models: whereby probability distributions are specified to represent the uncertainty about the true values of the model input parameters. Three secondary objectives are also specified. Firstly, the empirical application of the alternative decision models requires the specification of a 'modelling process' that is not well defined in the health economics literature. Secondly, a comparison of alternative methods for specifying probability distributions to describe the uncertainty in the model's input parameters is undertaken. The final secondary objective covers the application of methods for valuing the collection of additional information to inform the resource allocation decision.

The empirical application of the two relevant modelling techniques clarifies the potential advantages derived from the increased flexibility provided by DES over Markov models. The thesis concludes that the use of DES should be strongly considered if either of the following issues appear relevant: model parameters are a function of the time spent in particular states, or the data describing the timing of events are not in the form of transition probabilities. The full description of the modelling process provides a resource for health economists wanting to use decision models. No definitive process is established, however, as there exist competing methods for various stages of the modelling process. The main conclusion from the comparison of methods for specifying probability distributions around the input parameters is that the theoretically specified distributions are most likely to provide a common baseline for comparisons between evaluations. The central question that remains to be addressed is which method is the most theoretically correct? The application of a VoI analysis provides useful insights into the methods employed and leads to the identification of particular methodological issues requiring future research in this area.

List of Contents

List of Tables

List of Figures

Acknowledgements

The research reported in this thesis was funded by a North Thames training fellowship grant and a MRC project grant.

I am particularly grateful to my two supervisors, Professor Martin Buxton and Dr Jackie Brown, for their help and advice throughout the development of the thesis. I am also indebted to Ms Liz Fenwick and Dr Andrew Briggs for their advice on a range of issues included in the thesis, and to Dr John Yarnold for help in understanding the clinical area of breast cancer.

I would also like to thank Dr Mark Sculpher, Dr Karl Claxton, Dr Alec Miners, Mr Richard Grieve, Mr Andy Davies and Dr Kathy Johnston for their comments and input on draft versions of the thesis.

I especially would like to thank my wife, (Dr) Julie, for her patience, encouragement, comments, and advice during the course of the thesis. Finally, I thank Benny for providing light relief at the end of each day.

Publications

The research presented in Chapter 2 has been published in a peer-reviewed journal [Karnon and Brown, 1998]. One paper describing the case study evaluation (comparing tamoxifen plus chemotherapy versus tamoxifen alone in postmenopausal node-positive women with early breast cancer) has been accepted for publication (Pharmacoeconomics). Three other manuscripts – comparing the modelling techniques, the data analysis methods, and applying the value of information methodology – are currently under review.

Glossary

ABC	Adjuvant breast cancer
CAF	Cyclophosphamide, Adriamycin, 5-Fluorouracil
CMF	Cyclophosphamide, Methotrexate, 5-Fluorouracil
DES	Discrete event simulation
DFI	Disease free interval
DFS	Disease free survival
DNED	Death with no evidence of disease
ENBS	Expected net benefits of sampling
EVPI	Expected value of perfect information
EVSI	Expected value of sampling information
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
NB	Net benefits
NICE	National Institute for Clinical Excellence
OR	Operations research
QALM	Quality adjusted life month
QALY	Quality adjusted life year
RCT	Randomised controlled trial
TTP	Time to progression
VoI	Value of information

Chapter 1 Introduction

1.1 Introduction

The methodological focus of this thesis is the use of decision modelling techniques in the economic evaluation of health care technologies. The first two sections of this Chapter provide the background for the thesis, setting out the purpose of decision models in health technology assessment (HTA), and a brief history of the use of decision modelling in the economic evaluation of health care interventions. The final two sections describe the specific objectives addressed in this thesis, and the structure of the thesis.

1.2 The purpose of decision models

Decision models are most commonly employed to inform an immediate resource allocation decision, using only the data available at the time of the evaluation. However, decision models may also be used to advise on the collection of further data to make a more informed resource allocation decision. In the context of health care, decision models facilitate the comparison of alternative interventions through the following attributes:

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

- Structures for ordering and synthesising information - decision models can act as frameworks in which to organise evidence from a wide range of relevant sources, such as the incorporation of expert opinion alongside the results of primary and secondary data analyses.
- Explicit formulations of assumptions - an important dimension of decision modelling techniques is that the assumptions and definitions made in defining the structure of the model are explicit. All issues captured in the model are open to scrutiny by experts, both from clinical and economic perspectives, so any difficulties in the model are more likely to be raised and addressed. Through this refinement process, those assessing the model can increase the degree of confidence that they have in the model and the results that it produces.
- Test for the implications of uncertainty – there is uncertainty associated with all economic analyses of health technologies. The development and use of decision models provides an opportunity to explore the sensitivity of the results to variations in the assumptions that underpin the model. Such sensitivity analysis helps highlight areas in which further research is likely to be most useful.

1.3 The use of decision modelling in economic evaluation

Decision modelling, alternatively labelled 'decision analysis', originated from operations research (OR) and game theory in the 1950's and was taken up by the medical research community in the early 1970's. By 1987, a progress report on the use of decision analysis in clinical decision making hailed "several hundred articles" that had used decision analysis in medicine [Kassirer et al, 1987]. Kassirer *et al* [1987] cited improved computer accessibility that aided the design and analysis of decision models, as well as the development of clinically relevant measures of utility, as the main factors in increasing the use of such models. Since 1987 these factors have moved further forward so that a simple keyword search on Medline for 'decision analysis' highlighted 1341 relevant articles up to the year 2000.

One of the earliest applications of decision analysis to the economic analysis of health care interventions evaluated the costs and economic benefits of screening for spina bifida cystica [Hagard et al, 1976]. A rudimentary flow diagram, as a precursor to the now established decision tree, illustrated the patient pathways, but the key

characteristics of a decision analysis were clearly evident. The number of published economic decision analyses steadily increased throughout the 1980's. The majority of these early examples were undertaken in the US, which was probably due to the greater recognition of economic evaluation *per se* in the US.

In the UK, the need for a guided rationale in allocating available resources, which provides explicit reasons for the choices made, was a key development in the National Health Service (NHS) in the 1990's [Buxton, 1993]. The increased awareness of the economic aspects of health care led to a rapid increase in the number of published economic evaluations throughout the last decade. Using the Office of Health Economics' economic evaluation database [OHE, 2000] an elementary analysis of the proportion of economic evaluations using decision analysis in 1990 and 1999 was undertaken. In 1990 forty-four applied studies are recorded on the database, of which only 1 was labelled a modelling study (2.3%), in 1999 the number of recorded studies was 767 including 88 specified as modelling studies (11.5%). Such statistics are indicative of the growth in economic evaluation, but especially of the disproportionate increase in the use of modelling techniques to the application of economic evaluation.

Debates on the merits of basing clinical and policy decisions on the results of modelling studies have accompanied the expansion in the use of decision models. Opinion on the use of such models is divided. Views have been expressed that range from the sceptical to the cautious to the welcoming. Sheldon [1996] exemplified the former view stating that the results of model-based evaluation could be manipulated more easily and with more subtlety than other forms of evaluation. Drummond [1992] stated that the lack of well-defined methods made modelling studies easier to manipulate, and confidence was not enhanced by the 'black box' feel about the analysis of such models.

The cautious approach to the uptake of modelling studies was demonstrated by a paper that represented a synthesis of the ideas put forward by a group of senior health services researchers [Buxton et al, 1997]. The authors described the main instance in which decision models were employed in economic evaluation as situations where clinical trials had not been conducted or had not included economic data capture. They stated that economic analyses should be based on unbiased estimates of effect and that randomised controlled trials (RCTs) were likely to be the best source of such data.

Similar concerns to those raised by Sheldon were discussed and the use of pragmatic trials as an alternative to modelling was proposed. In conclusion, however, Buxton *et al* recognised that modelling is an unavoidable fact of life, but recommended that the use of modelling should be restricted to two scenarios. Firstly, to the early stages of development to identify the value of, and to inform, further research. Secondly, as a last resort when RCTs have not been possible or the available trials have not provided sufficient information.

Halpern *et al* [1998] complete the spectrum of opinion with their call for the use of decision models to be expanded. These authors stressed the fact that clinical trials are often not suitable for collecting relevant data as patient groups and the implemented treatment patterns may differ significantly between trials and the real-world. Decision models were also deemed to provide more timely information to aid decisions, rather than having to wait for the completion of clinical trials. Moreover, they claimed that the results from a prospective study might well be irrelevant by the time they are produced. Halpern *et al* suggested that the use of decision models is grounded in the economic theory of decision-making under uncertainty by providing an explicit outlet for the representation of the inherent uncertainty in all clinical resource allocation decisions. They accepted that there are problems associated with modelling studies, but argue that many of the cited problems could be countered through 'rigorous peer review and methodological development'.

1.4 Aims and objectives of the thesis

Methodological development in the application of decision models to the economic evaluation of health care technologies is necessary to overcome the expressed concerns about the use of such models to inform policy decisions. The range of potential problems stem from such issues as the inappropriate use of clinical data, possible biases in the use of observational data and expert opinion, the range of discretion available to the analysts, the potential for financial conflict between sponsors and researchers, and concerns about the transparency or validity of models [Buxton et al, 1997; Luce, 1995; Kassirer and Angell, 1994]. There have been attempts to define guidelines for modelling studies in recent years [Kassirer et al, 1987; Sonnenberg et al, 1994; Halpern

et al, 1998; Sculpher et al, 2000], though such papers only scratched the surface of the issues involved in the modelling process.

An important issue in the use of decision models to evaluate health care interventions is the correct choice of modelling technique, as the attributes of the chosen technique could have a significant impact on the process, as well as the results, of an evaluation. Modelling to assess the economic impact of health technologies is currently dominated by two techniques – the decision tree and the Markov model. Computer science experts have suggested that discrete event simulation (DES) offers important advantages over the other forms of modelling. DES is relatively untested in the field of HTA, though it has been used extensively in the OR field, including applications to the planning of health care services [Bolger and Davies, 1992; Lehaney and Paul, 1994].

The issue of choosing the correct modelling technique has been referred to in the health economics literature [Sonnenberg et al, 1994; Chaussalet et al, 1999], but the consequences of the choice have not been fully explored. This thesis aims to investigate the effect of the choice of decision modelling technique on the process and output of modelling projects, concentrating on a comparison of Markov models and DES to model extended time horizons. It was hypothesised that alternative modelling techniques could be compared on the basis of two broad criteria, flexibility and analytic input:

- Flexibility describes the representation of interrelationships between parameters, as well as the use of different forms of input data. This criterion reflects how closely a particular modelling technique allows the reality of patient pathways to be modelled.
- Analytic input relates to the complexity of the technique in terms of the level of expertise and the amount of time required.

The comparison of the modelling techniques is also presented in the context of a stochastic evaluation. A number of papers have described the use of probabilistic sensitivity analysis in economic evaluation [Felli and Hazen, 1998; Lord and Asante, 1999; Pasta et al, 1999]. Alternatively labelled as stochastic cost-effectiveness analyses [Briggs, 1999], the value of the input parameters within such models are described as

probability distributions. Distributions of each of the model's outputs are informed by randomly sampled sets of input parameter values from the specified probability distributions, which enables the statistical analysis of cost-effectiveness. Three secondary objectives for this thesis relate to the stochastic analysis of decision models. Firstly, the full process for developing an economic HTA model, from the point of specifying the project to the final experimentation with the model, is presented. Methods are incorporated from various disciplines, including health economics, clinical research, OR and the social sciences.

Methods for assembling the necessary probability distributions to populate stochastic decision models have been presented [Eddy et al, 1990a; Eddy et al, 1990b; Pasta et al, 1999], and the theoretical basis for the alternative methods have been reported [Lipton et al, 1995], but they have not been compared empirically. Another secondary objective of this thesis is to describe alternative input data analysis methods, and to apply them empirically to gain a better understanding of the differences between them, and of their respective merits.

Stochastic decision models also facilitate the statistical analysis of the value of conducting further research. Bayesian value of information (VoI) analyses aim to estimate the optimal sample size for a prospective trial to inform parameter values within a decision model. In the context of economic HTA models, parametric methods for the analysis of the VoI are complex and require strong assumptions [Claxton, 1999]. Non-parametric methods are available, which are in the early stages of development. The final secondary objective of this thesis is to apply these non-parametric methods empirically in order to highlight particular areas for future methodological research.

1.5 Structure of thesis

The thesis consists of nine further chapters, as set out below.

Chapter 2 introduces the alternative modelling techniques that were considered for comparison in this thesis, providing a commentary on the state of knowledge at the outset of the thesis. A general introduction to the three main techniques – decision trees, Markov models and DES models – is followed by a review of the decision modelling literature providing examples of the use of the different modelling techniques, which informs a preliminary assessment of the strengths and weaknesses of the alternatives. Chapter 2 also contains an assessment of possible criteria for the comparison of alternative modelling techniques. Five statements are defined in relation to modelling characteristics that could affect the choice of technique employed in an economic evaluation. These five statements form the basis for the comparison of the modelling techniques presented in this thesis.

Chapter 3 describes the modelling process from the point of considering a modelling project to the analysis of the model's outputs. Five stages of the modelling process are defined and discussed in chronological order:

1. Specifying the theoretical model;
2. Undertaking of a literature review to obtain data to populate the model;
3. Analysis of the identified data to populate the model;
4. Implementation of the model;
5. Experimentation with the model.

Guidance is drawn from the mainstream clinical and health economics literature, as well as from the OR and social sciences literature. Most areas of the modelling process are open to differential interpretation and analyst discretion, and where applicable, alternative options are presented.

Chapters 4 to 9 relate to the application of the methods described in Chapter 3 to a case study evaluation, which compared alternative adjuvant therapies for early breast cancer. **Chapter 4** reports the first stages of the case study evaluation, including a review of previous economic analyses of early breast cancer. The treatment area is described as part of the process for developing the preliminary model structure, which provides the

framework for the literature review. The description of the literature review focuses on the adopted inclusion and exclusion criteria, and the sources searched. The final section addresses the issue of sub-group analyses with the aggregate patient population.

Chapter 5 presents the reappraisal of the preliminary model structure and the process of harmonising the identified data. The final section describes the analysis of the data gathered from the literature review and the specification of probability distributions to describe the uncertainty around the true values of each input parameter. Three alternative approaches to specifying input distributions are described:

- Specification of theoretical distributions for each type of input parameter, using formulae derived by method of moments to estimate distribution parameters;
- Creation of weighted datasets, which are inputted to a software package that fits analytical distributions;
- Creation of weighted datasets used directly as input distributions.

Chapter 6 describes the construction of a Markov process and a DES model to evaluate alternative adjuvant therapies for early breast cancer. Decision trees are excluded at this stage, because they are not considered a realistic option to model extended time horizons. The differential aspects of implementing a Markov process and a DES model are presented in relation to the description of the health states within the model, but also to controlling the model inputs and outputs. Two issues relating to the analysis of the models are also addressed: an assessment of an adequate sample size to minimise the impact of first-order uncertainty on the results of the DES model, and a comparison of alternative methods for describing probabilistic 'length of time' input parameters.

Chapter 7 outlines the verification and validation processes employed to check that the models are internally consistent, and to check that the models' outputs are realistic, respectively. Three categories of verification are applied, which cover different aspects of the model's operations. The main form of validation involves the identification of a range of relevant sources of data, which are compared to the outputs from the case study evaluation. The corresponding outputs from the case study models are compared to the identified data and reasons for any differences between the compared outputs are sought

in the context of methodological and data differences between the case study evaluation and the identified studies.

Chapter 8 presents the results of the experimentation with the two decision models. The results of the analysis to inform an immediate allocation decision using only the identified data from both models and for all three data input analysis methods are presented. These results are presented as incremental cost-effectiveness ratios, but also as net benefits employing cost-effectiveness acceptability curves. The results of the VoI analysis are presented for both decision models.

Chapter 9 discusses the comparison of the DES model and the Markov process in the context of the case study evaluation, relating to the five model characteristics statements defined at the end of Chapter 2. The comparison of the alternative methods for specifying probability distributions is also discussed, as well as the methodological implications concerning the use of the alternative modelling techniques and methods for pooling and formatting the input data in a VoI analysis.

Chapter 10 addresses each of the objectives specified at the beginning of the thesis. The methodological insight gained from the development of a framework for the modelling process and the implications of the empirical evidence derived from the case study evaluation are presented, and areas for future research are recommended.

Chapter 2 Alternative modelling techniques

2.1 Introduction

The aim of this thesis is to compare the main decision modelling techniques that have been employed in the economic evaluation of health care technologies to date. This chapter provides a commentary on the state of knowledge regarding alternative modelling techniques prior to this thesis, concluding with the specification of criteria that are used as the basis for the comparison of the alternative decision models undertaken within this thesis.

Modelling, as a decision analytic tool, is a prominent resource in the health economists' toolbox [Buxton et al, 1997]. The critical evaluation of most aspects of the modelling process is under developed, but this problem appears to be particularly acute with respect to the appropriate choice of decision modelling technique. A recent paper reviewed modelling economic evaluations published in 1997 [Barton et al. forthcoming]. Of the 119 papers reviewed it was reported that 76 (64%) employed decision trees, 43 (36%) used Markov processes and only 2 (2%) reported results of discrete event simulation (DES) models (two studies reported results from both a decision tree and a Markov chain).

The characteristics of decision trees and Markov processes are very different and the choice between the two techniques in alternative treatment settings is relatively straightforward. With the introduction of DES to the field of economic evaluation in health care the issue of choosing the appropriate technique could become an important

decision in the initial stages of modelling projects. The choice of the correct modelling technique has been referred to in the health economics literature [Sonnenberg et al, 1994; Chaussalet et al, 1999], but the consequences of the choice have not been fully explored.

This chapter provides an introduction to the three forms of decision modelling referred to above – decision trees, Markov processes and DES models. A review of the clinical literature relating to modelling methodology was undertaken, as well as referring to work outside the discipline such as operations research and social science texts. The purpose of the literature review was to inform the characteristics of the alternative modelling techniques, but also to identify research covering the whole modelling process, which is described in chapter 3. Full details of the review are presented in Appendix 1. A large number of applied modelling economic evaluations have been published and a systematic review collating and reviewing all such published studies was not deemed to be a fruitful exercise. Instead, handpicked studies are used to illustrate particular aspects of the alternative techniques, as well as to highlight any limitations identified with respect to the characteristics of different treatment areas.

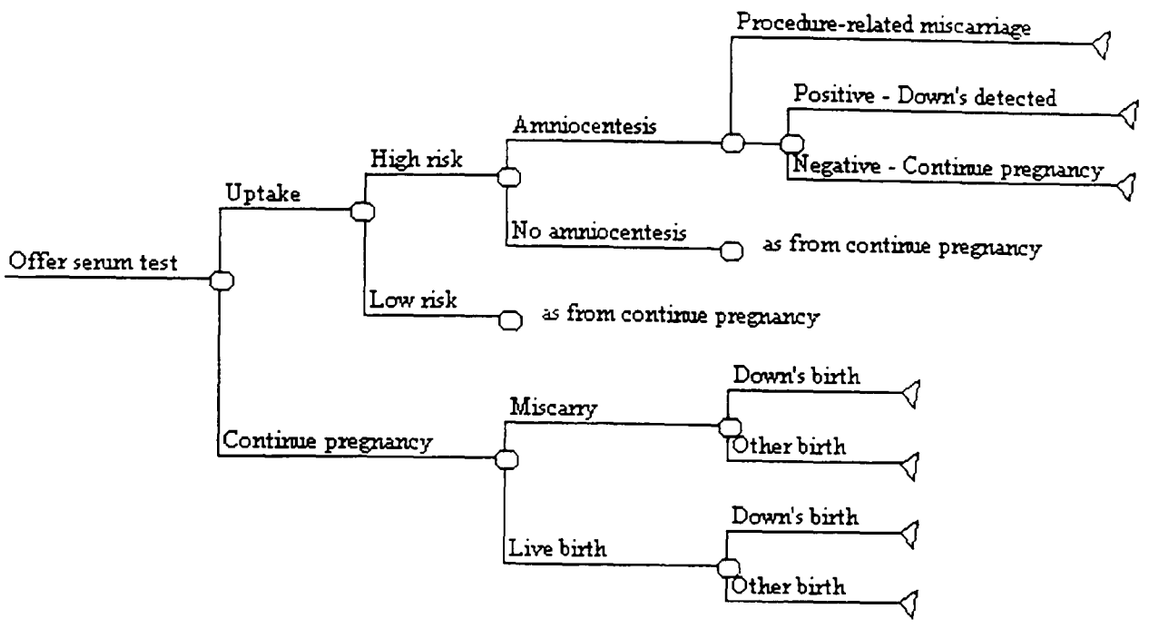
2.2 Decision trees

Decision trees are the simplest of the commonly used decision modelling techniques. As a tool for modelling relatively uncomplicated scenarios, decision trees provide a means of structuring a problem, and an effective method for combining data from various sources. Costs and effects are typically incorporated into a decision tree in different ways. The outcome measures of interest are generally attached to the endpoints of a tree, and the proportions of patients completing the tree at the respective endpoints are summed to give a measure of effect. Costs, however, may be attached to events within the tree, as well as to endpoints. To calculate total costs for each intervention, the costs associated with each unique pathway in the relevant section of the tree are summed. At each chance node, probabilities, conditional on the previous event, determine the proportion of patients progressing along each unique pathway in the tree.

Decision trees are most appropriate for modelling programmes in which the relevant events occur over a short time period, or evaluations that use an intermediate outcome measure. Decision trees are especially convenient for capturing a range of uni-dimensional outcomes. The following paragraphs provide examples of the use of decision trees in evaluations that covered short time horizons, either due to the nature of the interventions being evaluated or to the scope of the evaluation. Examples are also given of evaluations that extrapolated outcomes from the end of the decision tree to cover extended time horizons.

Evaluations of screening programmes provide good examples of the strengths of the simple decision tree [Lieu et al, 1994; Gessner et al, 1996]. Fletcher *et al* [1995] used a decision tree to synthesise data for a cost-effectiveness analysis of different antenatal screening policies for Down's syndrome. The decision tree covered the time period from the first point of contact with the screening programme (after confirmation of pregnancy) to the birth of the child (or termination of pregnancy). Figure 2.1 illustrates the use of the decision tree to describe the possible pathways following the offer of a serum test. Chance nodes related to the uptake of screening, high or low risk result, acceptance of amniocentesis, procedure related miscarriage, positive or negative test result, and a live birth or natural miscarriage. The decision tree simplified the expression of the relationships between these chance events in determining the outcomes. The frequencies of the following outcomes were recorded at the terminal nodes of the tree for each of the screening options: the number of live births with and without Down's syndrome, miscarriages with Down's syndrome, cases of Down's syndrome detected antenatally, amniocentesis performed, and the number of women offered screening. Costs were presented per Down's birth prevented, though the other important outcomes were also presented making the decision-maker aware of the negative outcomes.

Figure 2.1 Section of decision tree describing pathways associated with antenatal screening for Down's syndrome



Adapted from Figure 1 [Fletcher et al, 1995]

A good example of an evaluation of alternative therapeutic interventions using a decision tree is an evaluation of commonly used forms of prophylaxis for thromboembolism in patients undergoing hip replacement surgery in South Africa [Abdool-Carrim et al, 1997]. The outcome measure was the number of cases of deep vein thrombosis avoided. A decision tree was suited to such an evaluation because there was a well-defined period beyond which the outcome of interest would not occur. The treatment area, rather than the analyst, defined the time horizon for the model.

Alternatively, decision trees have been employed to model costs alone over short time horizons, assuming that outcomes are equivalent between the treatment options. For example, Jansen *et al* [1996] compared the cost of different non-steroidal anti-inflammatory drugs (NSAID's) relating to gastrointestinal (GI) complications. On the basis of expert opinion, the tree covered a 30-day treatment period for patients with osteoarthritis, which was judged to be sufficient to assess the incidence of NSAID-induced GI injury. The decision tree incorporated seven categories of mutually exclusive events ranging from no GI complications to hospitalisation.

A less commendable example of the use of a decision tree is a comparison of risperiodone versus haloperidol for the treatment of chronic schizophrenia [Davies et al. 1998]. This study covered a relatively long time horizon of two years, comparing the difference in costs with the number of patients responding at the end of the two years. A problem with this example was the arbitrary choice of time horizon at which the respective outcome measures were defined. Some authors have attempted to incorporate longer-term outcomes in decision trees by extrapolating from the short-term outcome at the end of the tree.

Midgette *et al* [1994] used a decision tree to compare intravenous streptokinase with conservative treatment for (suspected) acute myocardial infarction (MI). Patients entered the tree between 4 and 24 hours after the onset of symptoms, and left the tree within 35 days having either survived or died (fatal MI or other cause). Costs were associated with treatment-related adverse events – systemic bleeding, which may be fatal, and acute cerebrovascular accident, which may be fatal, disabling or non-disabling. The study incorporated survival by simply assuming a life expectancy of five years per survivor, and dividing the incremental costs by five to calculate a "cost per year of life saved".

The five-year life expectancy was obviously a very crude assumption designed to give a rough indication of the incremental cost per life year. Other studies have extrapolated from the short-term outcomes of a decision tree in greater detail. Kalish *et al* [1995] employed a decision tree to compare alternative thrombolytic therapies for acute MI. The tree described possible short-term events in some detail, including stroke (disabling and non-disabling), CABG, reinfarction, haemorrage, anaphylactic reaction and severe hypotension. Following any combination of these events the terminal nodes described five long-term health states – dead, stroke, reinfarction, stroke and reinfarction, and neither stroke nor reinfarction. From these endpoints constant costs and utility values were assigned to the surviving patients for an average life expectancy. This extrapolation reduced the explicit nature of the study because a long period of the treatment's effect, which had a large impact on the results of the study, was not analysed transparently. The authors had access to relevant data covering the longer-term outcomes of the patient group, but the decision tree did not facilitate the full incorporation of such data.

2.3 Markov models

Within Markov models, events are modelled as transitions from one health state to another. The time horizon covered by the model is split into cycles of equal length. At the end of each cycle a patient may move to a consequent health state, or remain in the same state (unless the current state is a tunnel state). This process of moving between states continues until a patient enters an absorbing state, such as the state 'dead'. Markov models are commonly described as being either a Markov chain or a Markov process, which differ in their representation of transition probabilities between health states. In a Markov chain the occurrence of events is determined by probabilities that are conditional only upon the current health state. In a Markov process, transition probabilities are also conditional on the current health state, but they may also vary according to the overall time spent in the model. Markov processes are particularly effective in clinical situations which involve continuous risk over an extended time horizon [Sonnenberg and Beck, 1993].

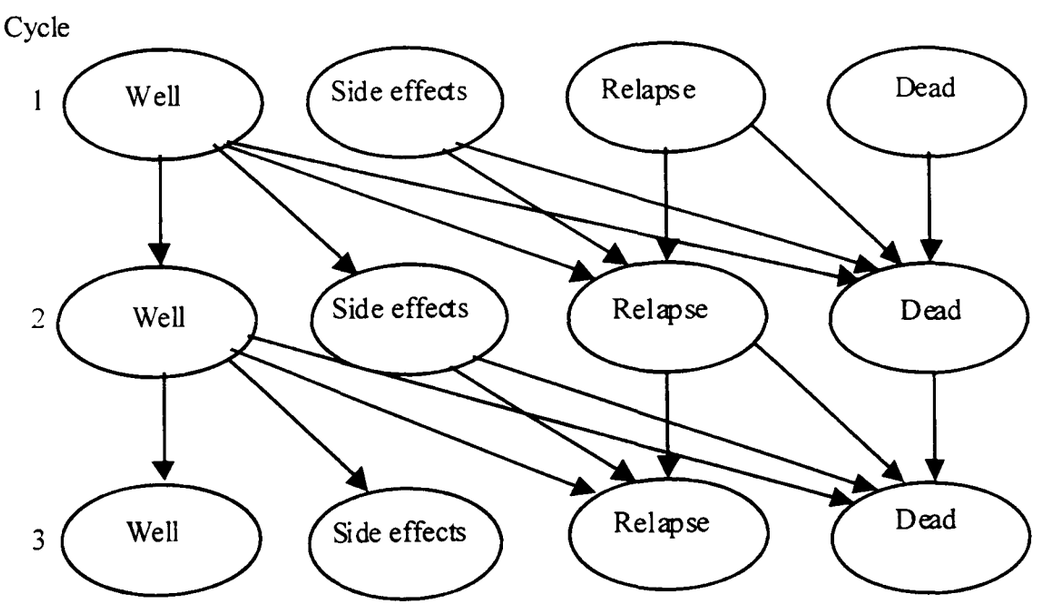
Utility values can be attached to each health state modelled, reflecting the severity of the state; similarly, costs are attached to individual health states to reflect the cost of remaining in a particular health state for the length of one cycle. A Markov model's outputs are estimated by multiplying the respective costs and utility values by the time spent in each health state, and then summing across all possible states.

There are three alternative methods available to evaluate Markov models. The matrix algebra solution is the purest (from a mathematician's perspective), but is rarely used due to the easy availability of computer power. A second approach analyses Markov models stochastically, whereby large numbers of individual patients are followed through the model. Monte Carlo simulation is used to randomly generate values from distributions that reflect the relevant transition probabilities. (This application of Monte Carlo simulation is referred to as first-order simulation. It should not be confused with second-order simulation, which samples the actual values of the transition probabilities from probability distributions (see section 3.4.3)). For example, if patients can move from state A to either state B or C and the respective probabilities are 0.75 and 0.25, each patient in state A will sample a value representing either state B or C from a

commensurately defined binary distribution. However, the most common approach in the economic evaluation of health care technologies is the cohort method, which is a deterministic analytic approach that follows a cohort of patients through the model. At the end of each cycle the proportion of patients remaining in a state is multiplied by the relevant transition probabilities to determine how many patients move to each consequent state. It should be noted that the choice of evaluative method does not affect the characteristics of the modelling technique.

Figure 2.2 presents a simplified version of a Markov process used to evaluate the cost-effectiveness of various adjuvant therapies applied to a range of breast cancer patients [Hillner and Smith, 1991; Hillner and Smith, 1992a; Hillner and Smith, 1992b]. The arrows between the health states indicate the possible pathways, until patients reach the absorbing state 'dead'. The authors stated that their reasons for taking a Markov approach were 'the relatively long time-frame and the time-dependent nature of the events considered' [Hillner and Smith, 1991].

Figure 2.2 An illustrative example of a Markov model comparing adjuvant therapies for early breast cancer



Adapted from Figure 1 [Hillner and Smith, 1992a]

The remainder of this section on Markov models describes some of their limitations relating to the lack of memory and the need to specify constant cycle lengths, as well as means of overcoming these problems. The final illustration describes the combination of decision trees and Markov models.

Firstly, an often-cited drawback of Markov models is their lack of memory, which means that the probability of moving from a particular state is not influenced by the route taken to arrive in the state. This assumption of the irrelevance of the history of patients is known as the Markovian assumption [Klein et al, 1993]. The following examples illustrate the differing impacts of this assumption. Danese *et al* [1996] modelled a screening policy for mild thyroid failure against a no screening scenario. The model consisted of a wide range of health states encompassing initial clinical states, clinical states of mild thyroid failure and related states (uncomplicated angina, MI, intermittent claudication and cerebrovascular accident). The model focussed on the immediate health states following screening - the clinical states of mild thyroid failure - with little detail given to the related cardiovascular health states. The analysis of the related overt hypothyroidism and cardiovascular health states could have been expanded, but it is likely that this would have required stronger assumptions to compensate for the lack of memory within the model. It is not clear whether the model structure was unduly influenced by the presence of the Markovian assumption.

Jonsson *et al* [1993, 1995] looked at treatment intervention in patients with established osteoporosis in order to reduce the risk of fractures. The Markovian assumption hindered the incorporation of differential probabilities for patients who had, and had not, experienced a fracture. The model was based on yearly calculations of risk. Patients could move from a healthy state to either death or a fracture state. Following a hip fracture, they were assumed to remain in one of three hip fracture states, reflecting varying quality of life, until death. The risk of a second hip fracture, re-operation, etc., was included in the associated cost and quality of life values of the hip fracture state. Following a year spent in any of the alternative fracture states all patients returned to the healthy state with no continuing impact on their risk of mortality, and were subject to the fracture risks again. This study appears to employ stronger assumptions about the risks of fractures that could be overcome if the patients' treatment histories could be recorded.

Technically, the Markovian assumption may be overcome by splitting health states so as to describe the path taken to reach the present state, for example, state C could become "state C after state A' and "state C after state B". An alternative approach is demonstrated by Sharples *et al* [forthcoming], who used linked Markov chains to

analyse the cost of the main clinical events following lung tansplantation. Following transplantation patients could be in one of five health states – well, rejection, cytomegalovirus, infection (non-CMV) or dead. The first Markov chain covered the first six months post-transplant. The second model covered the subsequent six months and included the possibility of developing BOS, a condition that alters the probability of experiencing the other health states within the model. At the point at which patients developed BOS they were transferred to a subsequent model with the same structure but alternative probabilities of experiencing the events described above.

The memoryless nature of the Markov model also impacts on the ability of the model to analyse subsections of an aggregate patient group who might have differential characteristics that influence their pathway probabilities.

Johannesson *et al* [1991] used a Markov model to investigate the cost-effectiveness of cardiovascular disease prevention using 8-year logistic risk equations for Coronary Heart Disease (CHD) and stroke. The values of age, sex, diastolic blood pressure, serum cholesterol, smoking, glucose intolerance and left ventricular hypertrophy, for each patient, were determined at the start of the model. Patients were subject to three forms of risk – CHD, stroke, and death from other causes. CHD was divided into five separate states – immediate death, recognised MI, unrecognised MI, angina pectoris and coronary insufficiency. From each of these health states, as well as from stroke, the only possibility was to remain in that state or to die. The values of the risk factors were varied according to the use of different interventions, but not to alternative patient characteristics. If the policy question referred to the aggregate patient group, separate analyses of the constituent groups would need to be combined to calculate the relevant cost-effectiveness results.

Another limitation of the standard Markov model is that a single time period for a cycle must be chosen after which patients are allowed to move to the next health state [Klein et al, 1993]. The length of a cycle is chosen to represent a clinically meaningful time period and to reflect the available data [Beck and Pauker, 1983]. For example, in choosing one month as an appropriate cycle length to evaluate the cost of glaucoma treatment Kobelt and Jonsson [1999] stated that longer cycles would not reflect clinical

management accurately, whilst shorter cycles would misrepresent clinical observation in glaucoma.

If the events modelled are subject to different meaningful periods a Markov model can be inflexible. Anton and Revicki [1995] evaluated the cost-effectiveness of antidepressants for major depression. The time horizon for the model was the lifetime of patient's, who were assumed to be 30-year old women. A Markov cycle of one year was chosen to reflect the probability of patients recovering from depression or recurring from remission over the course of their lifetime. Within the course of a year spent receiving treatment for an episode of depression the effects of five separate states with varying cost and utility values were calculated outside of the model.

It is possible to link separate Markov processes using differing lengths of cycles, each describing a particular section of a patient's possible prognosis, as described above [Sharples et al, forthcoming], though such an approach is dependent on suitable patient pathways, and does increase the complexity of the modelling task.

Markov models are often used effectively in conjunction with decision trees. In evaluations that comprise numerous events within the early part of the model, Markov models are then attached to the terminal nodes of the decision tree to describe less complex patient pathways in the following years. This approach utilises the simpler mechanisms of the decision tree to describe events that occur only at the beginning of an extended time horizon.

Eckman *et al* [1995] combined a decision tree and a Markov chain to examine the cost-effectiveness of approaches to the diagnosis and treatment of patients with type 2 (non-insulin-dependent) diabetes mellitus who have foot infections and suspected osteomyelitis. The pathway of patients up to a constant health state, including diagnosis and initial treatment of the condition, was modelled within a decision tree format. The initial tree differentiated between medical and surgical treatment options, following a variety of testing strategies that related to either a high sensitivity or a high specificity of the diagnostic procedure. The tree also incorporated the possibility of recurrent infection or initial treatment failure. Thereafter, the surviving patients entered a Markov chain where patients could die of diabetes, of a co-morbid disease, or according to the

age-specific mortality rate of the general population. The Markov chain was very basic. The lack of descriptive states in the Markov chain appears to reveal more about the authors' concentration on the initial treatment section of the model, and their assumption of subsequent constant costs and utility, than the shortcomings of Markov chains.

2.4 Discrete event simulation (DES)

Both DES models and Markov processes are forms of simulation, though DES allows more complicated representations of the system being modelled [Hillier and Lieberman, 1995]. Within DES, patients move through the model experiencing events at any discrete time period after the previous event. DES models may only be analysed stochastically using first-order Monte Carlo simulation as described in the previous section. This is because DES is event-orientated, whereby the model asks what and when is the next event for every patient at each event, rather than a Markov model, which asks what events are occurring at regular intervals. The analysis of second-order uncertainty is similar for both modelling techniques, consisting of testing alternative combinations of input parameters by undertaking multiple runs of the models.

The following section describes and provides examples of the main advantages of DES over Markov models, namely increased flexibility over data requirements and an ability to overcome the restrictions of the Markovian assumption. Two potential disadvantages are also described, concerning the dangers of overspecifying models and the need for increased analytic input.

The greater flexibility of DES with respect to the required data was demonstrated by Hart *et al* [1997] who used DES to estimate the direct lifetime costs of an insulin-dependent diabetes mellitus (IDDM) patient. Only those chronic conditions assumed to have the greatest economic impact were modelled – nephropathy, retinopathy and cardiovascular complications. Unaffected patients entered the system and were assigned a probability of developing IDDM, as well as a corresponding time of diagnosis. Individual patients' experience of chronic health states associated with IDDM were sampled from probability distributions representing incidence rates and times of onset. An overall time to cardiovascular complications for a proportion of all IDDM patients was calculated. If the simulated time to death after diagnosis was

greater than the average time to cardiovascular complications an adjustment to costs and life expectancy was made. For retinopathy complications the proportion of patients requiring photocoagulation increased with duration of the disease, for which costs were incorporated into the model using an algorithm describing the increasing incidence over time.

Urban *et al* [1997] used a stochastic simulation model to simulate the natural progression of ovarian cancer. A screening programme was then superimposed, and the change in life years gained and costs evaluated. A public use database was used to estimate the proportion of women diagnosed in 5-yearly groups and in each stage of diagnosis (local, regional or distant). Assuming a uniform distribution within the 5-yearly groups, an exact age (month) at diagnosis for each patient was sampled, as well as a stage from the calculated distribution of stages at diagnosis. The length of each subsequent stage was determined in relation to the length of the local stage. Women were then randomly assigned a percentile in the relevant age- and stage-specific empirical survival distribution. Age at death was calculated by adding survival time to age at diagnosis. An average cost for each disease stage was assigned to the time spent in each stage. Various combinations of transvaginal sonography and/or the tumour antigen CA 125 were inserted into the model. The stage, timing within a stage, and the sensitivity of each of the screening tests were all defined in relation to specific distributions. When screening resulted in earlier detection, a new age at death was assigned by indexing the new relevant age- and stage-specific survival distributions, and adding survival to age at diagnosis following detection by screening. The input data used to populate the model were inputted in their original formats, such as the specific length of time to the occurrence of another event, rather than transforming data as would be required using a Markov model. This increases the realism of the model.

DES overcomes the limitations of the Markovian assumption by assigning attributes to patients that describe relevant characteristics or their treatment history, which may influence their pathway through the model [Davies and Davies, 1987]. For example, patients may be assigned an age or stage of cancer, prior to entering the model, they may also acquire attributes as certain events within the model are experienced. In addition, patients may have attributes that influence the costs and/or utility values associated with a particular state. In the absence of patient attributes, costs and utility

effects are incorporated in a similar manner to Markov models, whereby costs and utility values are attached to each health state within the model and the time spent in each state is multiplied by their respective values.

The MISCAN (MIcrosimulation SCreening ANalysis) model, developed by Habbema *et al* [1984, 1987] to simulate models of mass screening for disease, demonstrates the use of patient attributes whereby the experience of previous health states, and their duration, influenced the duration in subsequent health states. The simulation described here was based on models of the natural history of cervical cancer, and the change in the natural history due to screening. Disease history was represented by a sequence of disease states and the ages at which a person entered these states. The model was split into seven categories of health states – not at risk of cervical cancer due to previous hysterectomy, screen-eligible states (normal, pre-invasive or pre-clinical invasive), clinical states (clinically or screen-detected cancer) and end states (death from cancer and death from other causes). Transitions between states were based on probabilities, and each transition had a corresponding probability distribution describing when the transition will occur. The attribute "age" could also influence transition probabilities. An additional feature of the model was that the time spent in a state could depend on the dwelling time in the previous state, which would be useful, for example, if there was a relationship between the growth rates of different stages of the disease.

Warner *et al* [1996] used a "stochastic, discrete event, object-oriented simulation" to investigate the health and economic implications of a smoking-cessation programme in the workplace. Model complexity was reduced due to the assignment of attributes to workers, both before and during their time in the system, which influenced their transition probabilities. The events within the model related to the actions of quitting smoking, leaving the firm, retiring and dying. The model was run twice, with and without the smoking-cessation programme in place. On activating the programme, smokers had a higher probability of quitting smoking, which sequentially influenced other variables of interest (mortality, medical costs, etc.). Each worker in the model was assigned an age, sex, time at firm and smoking characteristics that influenced their progression through the model. At the end of every year age was increased by one year, as well as other variables, such as years of smoking or smoking cessation, which in turn influenced progression through the model. If a worker neither died, retired, nor left the

firm, he/she continued cycling through the model. Former workers who either left the firm or retired were subject to only two probabilistic events - background (non-programme-related) smoking cessation and dying.

A possible drawback of DES is that an analyst may attempt to make a decision model overly complex due to the increased flexibility of DES [Balmer and Paul, 1986]. Ideally, a model would describe reality as closely as possible. However, there are two constraints on data collection, and hence, on the complexity of a decision model: limited research resources and the need to inform policy within a limited timeframe. Though the most important elements of a model cannot be objectively tested until the final analysis of the model, it is possible to identify parameters on the periphery of the model that could be safely excluded. An example of underplaying the capabilities of DES because the identified data did not warrant a more complex representation of events is provided by the model used to estimate the costs associated with IDDM [Hart et al, 1997]. As described above, this model included the impact of cardiovascular and retinopathy complications as affecting a proportion of all patients if death did not occur before the average time to the occurrence of such complications. Such a portrayal of events was all that the available data supported, though the events could have been modelled in much more detail given the capabilities of DES, but more complexity may have distorted the available data.

Finally, DES models are generally more complicated to build as they are based on the use of programming code. Three of the four examples quoted above used complex programming languages (Gauss, FORTRAN and Objective-C). Though the advent of 'visual interactive modelling systems' assists the development of DES models, without the need for 'proper' computer programming [Pidd, 1998], the additional features of DES do increase the complexity of the modelling task.

2.5 Criteria for comparison

Two broad criteria on which decision modelling techniques can be compared are the quality of the decision model and the corresponding analytic input. The following two sections describe the issues relating to these two criteria, which form the basis for the empirical comparison of the alternative modelling techniques undertaken in this thesis.

2.5.1 Model quality

It has been argued that the true test of the quality of a decision model is in terms of its value to the decision-maker(s) as the ultimate aim of the model is to help the user reach a better informed and rational decision [McCabe and Dixon, 2000]. In principle the quality of a model could be tested by randomising decision-makers to use and not to use a model, with quality being proven if the outcomes associated with the model-based decisions are better than those associated with the control group [Sculpher et al, 2000]. Unfortunately, in practice it is probably not feasible to test models in this manner and softer, more subjective measures of quality are necessary. Recent guidelines on estimating cost-effectiveness in health and medicine referred to decision models being as good as their ability to represent reality 'at the level needed to draw useful conclusions', noting that this depends upon model structure and the assumptions regarding relationships between model parameters [Mandelblatt et al, 1996].

Indeed, most of work that has attempted to describe positive aspects of decision model design has been reduced to general statements that can only be interpreted subjectively. The following quotes provide examples:

'models should be kept as simple as possible to aid understanding by decision makers.' [Buxton et al, 1997]
'it is important that the possible pathways described by the model are feasible and sensible.' [McCabe and Dixon, 2000]
'the analyst should choose the type of model ... selecting the simplest format which adequately reflects the disease.' [Sculpher et al, 2000]

The main conclusion drawn from these examples is that good modelling practice can only be developed in general terms because the application of the derived principles 'requires knowledge of the clinical domain and involves subjective judgement' [Sonnenberg et al, 1994]. The derivation of criteria for the comparison of alternative modelling techniques is also based on this principle of generalisability. The approach sought to link the desirable characteristics of a model with the differential characteristics of the modelling techniques and determine the importance of each

characteristic in the representation of patient pathways. Sculpher *et al* [2000] grouped the dimensions of quality into three categories – the structure of the model, data, and consistency – which described the areas in which the choice of modelling technique could affect the quality of a modelling project. Sonnenberg *et al* [1994] classified alternative models by their design and use of time as a model function. The discussion around model quality, therefore, covers three broad areas; representation of time, structural considerations and data flexibility.

The main characteristics of the decision tree are that any event within the tree may happen only once and that the timing of the events within the model can only be incorporated in the terminal nodes. As the other two modelling techniques handle long time horizons more adequately the first statement over the choice of modelling technique relates to the required time horizon.

1. *If anything other than short-term outcomes are to be modelled decision trees are an inappropriate choice of modelling technique.*

This does not mean that decision trees are necessarily the right choice to model short time horizons, but the following discussion concentrates on the representation of extended time horizons within decision models. At present the main techniques available to incorporate such time horizons are Markov models and DES.

The structure of the model refers to the actual health states included in the model, as well as the relationships between them. The inclusion of relevant health states would appear to be similar in Markov models and DES, but differences do emerge in the ability of the two techniques to represent certain relationships between health states. The crux of the difference is the Markovian assumption that informs the relationship between any two health states only on the basis of the current health state and the overall time spent in the model. This assumption is not necessarily a hindrance as the natural history of many chronic diseases can be approximated by state of health and time since diagnosis [Beck et al, 1994]. Problems are apparent in other disease areas, such as cancer, where the transition probabilities between stages depend on the length of time spent in each state [Beck and Scardino, 1994]. Similarly, the probability of

patients with a history of depression experiencing further episodes of depression is dependent on the length of time they spend as asymptomatic [Nuijten et al, 1995].

Alternatively, there are evaluative areas in which the route taken to reach a particular state will influence the pathways out of the state. For example, in an evaluation of a hospital at home scheme versus conventional inpatient care for elderly patients, the probability of experiencing a readmission following discharge differentiated between patients admitted to the scheme through the accident and emergency department and from inpatient wards [Campbell et al, 2000]. Such instances do not necessarily preclude the use of Markov models as subsequent states can be described with reference to the experience of previous states.

The following quality statements set out the basis for the structure of the model to correspond as far as possible to the reality of the disease.

2. *If model parameters are a function of the time spent in particular states DES will more accurately reflect the true relationships between health states.*

3. *If the specification of similar health states that differ only with respect to the experience of previous states compromises the clarity of the model, the use of DES should be considered.*

The main issue around the use of data within decision models relates to the accurate depiction of the timing of events. The probability of an event occurring within a Markov model must always be defined as a function of cycles of fixed duration. This format can cause problems if an event is likely to occur within a specified time period and the state from which the event occurs is not an initial state. Another data-related issue concerns the specification of the length of a cycle in a Markov model. This characteristic means that when two events of widely differing duration are included in a model the representation of the timing of one of the events is likely to be unrealistic. For example, a model describing the lifetime treatment costs of patients with sickle cell disorders would necessarily include a range of acute and chronic conditions [Karnon et al, 2001]. Acute conditions typically occur sooner and more frequently than chronic conditions. If a cycle length of 1-month is chosen to reflect the length of acute events

then the time spent in the chronic states must also be represented as monthly probabilities. In a DES model the timing of events is completely flexible. Such data can be represented as the probability of an event occurring at specified times or the time to an event can be stated directly.

The fourth quality statement covers the flexibility of the modelling technique with respect to the realistic representation of the data.

4. *If the data describing the timing of events are not in the form of transition probabilities then DES will provide a truer representation of reality.*

2.5.2 Analytic input

The level of analytic input associated with the alternative decision modelling techniques encompasses two aspects. Firstly, the issue of available expertise may be relevant. Though increasingly user-friendly software for the development of DES models is available, there remains a steeper learning curve associated with the use of DES as opposed to the relative simplicity of developing and analysing Markov models. Secondly, Markov models and DES may differ significantly in the time required to build, and experiment with, the model. Obviously, the time spent building the model will be a function of the expertise of the analyst, but the greater flexibility of DES is only achieved through added complexity, which increases the likelihood of modelling error. The cohort method of analysing Markov models is adequate for the majority of economic evaluations in health care (this issue is developed further in Chapter 3). Experimentation with a DES model can only be undertaken using first-order Monte Carlo simulation, which requires much longer experimentation times than cohort analyses. It should be borne in mind that the time to build and analyse a decision model covers the testing, verification, validation, baseline analysis and sensitivity analyses undertaken. The following statement, regarding analytic input, relates to the availability of such input.

5. *The advantages of DES need to be weighed against the additional resource requirements. Realistic assessments of the necessary inputs should inform the choice of modelling technique.*

Conclusions

The preceding sections described the varying characteristics of three alternative modelling techniques that have been used to model the cost-effectiveness of health care technologies and possible criteria for the comparison of these alternatives. Preliminary inferences can be drawn from the discussion with respect to the general characteristics of treatment areas that are best suited to the alternative modelling techniques.

Simple scenarios occurring over a short time horizon may best be modelled using a decision tree. Markov models allow longer time periods to be analysed, in which the risks of events are continuous, and the timing of an event is uncertain [Sonnenberg et al. 1994]. Such models do not impose unduly restrictive assumptions unless the treatment area of interest conflicts with the Markovian assumption of pathway independence. In such cases, DES may be viewed as a superior technique. The biggest advantage of DES appears to be that it allows the analyst to model more complex and dynamic systems than other types of modelling, as well as permitting experimentation that might not be feasible otherwise. The greater flexibility inherent in the use of discrete event simulation may also enable the model to capture more detail about the uncertainty in the system being modelled [Bolger and Davies, 1992].

The choice of modelling technique should be based on an in-depth assessment of the particular disease area. The process of matching the characteristics of the alternative modelling techniques to the disease area, as described above, should provide an indication of the specific differences in the quality of prospective models. Such differences should then be traded against the more objectively defined variation in the required analytic inputs.

Five statements linking issues of model quality and analytic input to the main modelling techniques employed in the economic evaluation of health care have been defined in this chapter. The statements, presented below, will form the basis for the comparison of the alternative modelling techniques that will be developed in the remainder of this thesis: following the empirical application of the competing techniques the associated modelling process and cost-effectiveness results will be assessed on the basis of these five statements. The application of this approach to the retrospective identification of a

'good' model in the case study presented in this thesis will test this process for the prospective determination of the appropriate modelling technique for individual disease areas.

1. *If anything other than short-term outcomes are to be modelled decision trees are an inappropriate choice of modelling technique.*
2. *If model parameters are a function of the time spent in particular states DES will more accurately reflect the true relationships between health states.*
3. *If the specification of similar health states that differ only with respect to the experience of previous states compromises the clarity of the model, the use of DES should be considered.*
4. *If the data describing the timing of events are not in the form of transition probabilities then DES will provide a truer representation of reality.*
5. *The advantages of DES need to be weighed against the additional resource requirements. Realistic assessments of the necessary inputs should inform the choice of modelling technique.*

Chapter 3 The modelling process

3.1 Introduction

This Chapter aims to provide the framework for the empirical application of alternative modelling techniques that will be described in the following Chapters. The full process for developing and analysing an economic health technology assessment (HTA) decision model is described from the point of setting up a project to the final experimentation with a computer-based model. The formulated process will be adopted explicitly in the case study evaluations, which will enable the process itself to be evaluated in the final Chapters of this thesis. At this stage the choice of modelling technique is left open because the underlying process of a modelling project is similar regardless of the type of model employed.

To inform the modelling process a wide-ranging literature search was undertaken, incorporating texts from the disciplines of medicine, social science and operations research. The details of the search are provided in Appendix 1.

A well known source of guidance for clinical decision analyses is the seminal publication by Weinstein and Fineberg [1980], though earlier papers had applied the techniques of decision analysis to clinical research. In particular, Kassirer published an introduction to decision analysis in a clinical context in 1976. The main limitation of both these sources was that they concentrated on the principles of decision trees, only providing a brief summary of other aspects of the process such as the assignment of probabilities. Indeed, the majority of the identified texts covering the use of decision

models within the evaluation of health care technologies focussed on the principles of the use of specific modelling techniques rather than other aspects of the modelling process [McNeil and Pauker, 1984; Detsky et al, 1997; Beck and Pauker, 1983; Briggs and Sculpher, 1998].

The most useful publications relating to the process of modelling projects in health care were those that did not focus solely on the principles of the modelling techniques, but rather on a framework for modelling studies. Three sources were identified that aimed to improve the general process of undertaking medical decision analytic models [Eddy, 1985; Sonnenberg et al, 1994; Sculpher et al, 2000], whilst another paper aimed to increase standardisation with regard to modelling practices [Halpern et al, 1998]. Each study produced good practice recommendations. In addition, general texts from the operations research and social science literature were sought to inform various aspects of the modelling process.

The remainder of this Chapter is split into five sections covering the modelling process in chronological order:

1. Specifying the theoretical model;
2. Undertaking of a literature review to obtain data to populate the model;
3. Analysis of the identified data to populate the model;
4. Implementation of the model;
5. Experimentation with the model.

3.2 Specifying the theoretical model

Sonnenberg *et al* [1994] constructed a framework of progression from the decision problem to the resulting model. Using current knowledge, their first step required the development of a theoretical model that represents the initial understanding of the possible pathways experienced by patients. The problem should be structured in an attempt to understand the issues that need to be addressed within the model [Pidd, 1998], by specifying a set of assumptions that take the form of logical or mathematical relationships comprising possible patient pathways [McHaney, 1991]. The development of the theoretical model may be clarified using simple diagrams to represent the relationships, such as influence diagrams [Marshall and Oliver, 1995], or

activity cycle diagrams [Paul and Balmer, 1993]. Gottfried lists several early questions to guide the construction of the conceptual model [Gottfried, 1984]:

- What are the relevant problem parameters? What are the state variables, the decision variables and the system parameters?
- What are the pertinent cause and effect relationships? How can they be expressed in mathematical terms?
- What data are required? Are the data readily available? If not, how much effort is required to obtain meaningful data?

The main parameters of interest to the health economist are the state variables, which are chosen to represent patient pathways. In the vast majority of economic evaluations, the only decision variable relates to the objective of the evaluation. Two types of system parameters are apparent in economic HTA models. One system parameter is the patient group of interest. The definition of the relevant patient group for an evaluation is obviously an important topic and is discussed below under 'study inclusion criteria'. The second system parameter describes the scope of model, which determines the range of events included in the model. For example, limited scope models concentrate only on those events that have cost implications for the health service, which can be broadened to include events that impact on the costs incurred by the patient and society in general.

Interrelationships within an economic HTA model are possible in the sense that the occurrence of an event along a patient pathway can influence the likelihood of the patient experiencing another particular event(s). Such knowledge is useful as an indicator of the data that will be required to populate the model and may be incorporated into the strategy adopted for the identification of data. It may also affect the choice of decision modelling technique. Goodwin and Wright [1998] highlight the problem of biased assessment of correlation as a source of modelling error. People often see non-existent associations between events because they can easily imagine the events occurring together. It is, therefore, particularly important that any assumed relationships between variables within the model are confirmed by available data.

Most modelling textbooks encourage the development of models with the least amount of detail that maintains the veracity of the model [McHaney, 1991]. A cyclical process may be enacted whereby increasing detail is added until the structure of model satisfies all parties involved in the evaluation.

3.3 The literature review

Most economic HTA decision models collect and integrate data from the existing literature. Ideally, such data would be identified using similar methods to those used in systematic reviews, being comprehensive, rigorous and explicit. The aim of a literature review has been interpreted as the avoidance of missing *useful* studies placed in sources 'outside one's habitual purview', not identifying every paper that is somehow related to the area of interest [Sutton et al, 1998].

There is a substantial methodological literature on systematic reviews, covering each of the main stages: study questions, study inclusion criteria, search strategies, study validity (quality), data extraction, and data analysis. The proposed methods for dealing with each of these issues should be described explicitly in the review protocol, which should be specified before the main study commences. The following sub-sections describe the development of the study questions, study inclusion criteria, search strategies, and data extraction, whilst methods for pooling the data (including a discussion around data quality) are covered in the next main section (Analysing the literature review data). However, prior to the development of the review protocol some background work is required to inform its structure [NHS Centre for Reviews and Dissemination, 1996]. This initial task consists of a brief literature search for two purposes. Firstly, it is necessary to estimate the volume of literature in the field so that an estimate of the size of the review can be made. Secondly, it is useful to have an idea of the most frequently employed study designs in the area of interest in order to inform the study designs included in the review.

3.3.1 Study questions

A series of study questions should be raised for separate elements of patients' pathways within the preliminary model, which can be treated as separate reviews. For example, a

typical economic HTA decision model might require separate questions to describe treatment side effects, response to treatment, and progression from response, as well as for the collection of resource use, cost, and utility-based data.

Real associations between parameters can be missed because there was no prior expectation of such a relationship [Goodwin and Wright, 1998]. It may be advisable to facilitate the collection of qualitative data within the review process so that potential relationships may be uncovered even if they were not hypothesised *a priori*.

3.3.2 Study inclusion criteria

Sutton *et al* [1998] identified four areas in which study inclusion criteria should be specified:

- Relevant patient groups
- Health intervention/technology of interest
- Outcome measures used
- Types of study design

Sonnenberg *et al* [1994] promoted the use of generic models that are applicable to a representative population or a broad category of patients, which would appear sensible to maximise the potential outputs of a model. It is important to recognise that the relevant patient group for reviews informing parameters within the model will probably be a subset of the aggregate patient group. For example, the original patient group may be split between responders and non-responders further down the patient pathways.

The explicit limiting of a literature search by defining inclusion criteria relating to the sources of data is not recommended in guidelines for systematic reviews or meta-analyses [Sutton et al, 1998; NHS Centre for Reviews and Dissemination, 1996]. However, the specification of multiple study questions relating to different events within a decision model means that the research burden can escalate very quickly, especially in treatment areas for which data are plentiful. In such cases, the need to populate the model with the 'best available data' [Buxton et al, 1997], must be balanced against the time and resources available to complete the study; 'best' may be redefined as 'sufficiently robust to engender confidence in the model inputs'. Additional

inclusion criteria may be specified that relate to the intensity of the search, such as restricting the search by setting a cut-off year, or excluding sources on the basis of methodology, for example, the review could be restricted to trial-based data. Alternatively, non-English language studies could be excluded. Research into the hypothesised Tower of Babel bias has been undertaken [Gregoire et al. 1995], but no evidence of the expected bias against negative studies was identified. An obvious inclusion criteria is to seek only published work, though such a criterion could be subject to publication biases, whereby negative results are less likely to be published. However, focusing solely on published research is legitimate in two circumstances. Firstly, if the quantity of published research is large it is unlikely that the direction of the aggregated results will be wrong, although the magnitude may be overestimated. The magnitude of any differences between intervention should then be interpreted cautiously. Secondly, the parameter of interest to the study question may not be the primary focus of the research identified. The bias towards significant results in publications is unlikely to apply to secondary results [Cooper, 1998].

These extra criteria can be relaxed with respect to some threshold of data quantity or quality for the individual study questions. For example, if few data, or data from poor quality studies are available during the initial review, then the inclusion criteria can be progressively widened until sufficient quality data are identified. The collection of additional data may also form part of an iterative procedure, whereby the model can be improved by obtaining additional data on key parameters that are identified by a sensitivity analysis of a populated model.

3.3.3 The literature search

The literature search defines the sources that are inspected for relevant data based on the specified inclusion criteria for each study question. It is quite feasible that different ranges of sources will be employed for study questions aimed at populating alternative sections of the decision model. The following paragraphs describe various data sources, though comparisons are difficult because there has been little empirical research on their respective merits [Cooper, 1998]. The easy access available to electronic databases, such as Medline, via the Internet, makes them an obvious starting point, though the EMBASE and SCISEARCH databases are recommended for inclusion in any

systematic review [NHS Centre for Reviews and Dissemination, 1996]. Other databases include the database of abstracts of reviews of effectiveness and the NHS economic evaluation database, both provided by the Centre for Reviews and Dissemination, and HEED, the economic evaluation database provided by the Office of Health Economics. Variable success has been reported for searches of electronic databases, which may be due to narrow searches or inadequate indexing on the part of database compilers or the relevant authors [Sutton et al, 1998].

Other sources include research registers, such as the NHS Research Register and the Cochrane library. Registers can contain information on research studies, based in the past, present and/or future. The precision of such databases is improving continuously and hopefully they will soon become the most reliable source of data, both published and forthcoming [NHS Centre for Reviews and Dissemination, 1996]. Additional studies are often unearthed through handsearching journals and scanning the reference lists of retrieved studies. Similarly, scanning the contents of relevant conference proceedings can provide details of the current state of knowledge and what research has recently been completed, usually before it has been published [Rosenthal, 1994]. Data may also be identified through informal channels that include the personal solicitation of research, by contacting individual researchers about any relevant research in which they have been involved.

If no usable data for a particular model parameter is identifiable it will be necessary to seek the opinions of experts in the field of interest. Two formal techniques have been identified that elicit the opinions of experts. The Delphi method involves sending an initial questionnaire to the chosen experts, the results of which inform the development of a second questionnaire. The second questionnaire is sent, along with the results of the first questionnaire, to the same experts. The attachment of the results from the first rounding of questioning empowers each expert with the same information for completing the second questionnaire [Hillier and Lieberman, 1995].

The expert group technique involves bringing together a group of experts who interact with each other to produce a consensus. The experts can be asked to specify a probability distribution. However, the Delphi method is generally preferred because it

does not attempt to forge a group consensus, the conclusions reached by an expert group may reflect the personalities in the group [Sculpher et al, 2000].

Whichever sources are employed in a literature search an important criterion is that the results obtained should be reproducible. Thus, full details of the search methodologies should be provided enabling the interested reader to obtain similar results.

3.3.4 Data extraction

The process for extracting data from the identified studies should use a specified data extraction form. In the context of a series of literature reviews to populate an economic HTA decision model the development of an evidential database to collect and manage the relevant data is a useful tool. The development of such a database is described in Appendix 2.

Recommended methods for the extraction of data state that it should be undertaken by at least two people [NHS Centre for Reviews and Dissemination, 1996], though the combination of limited resources and quantity of data may militate against such ideals.

3.4 Analysing the literature review data

The process of analysing the literature review data encompasses three broad activities, which are described in this section. Firstly, in the light of the identified data, both quantitative and qualitative, a reappraisal of the theoretical model should be undertaken. Secondly, the identified data should be appraised to ensure that the relevant variable definitions are the same. If not, it may be possible to harmonise the definitions by making adjustments to the reported data. Finally, the identified data for each input parameter should be pooled and arranged into a form that can be used to populate the model.

3.4.1 Reappraising the model structure

The theoretical model is developed primarily to inform the literature review. There are two reasons that the structure of the initial model may be altered. Firstly, additional

qualitative data may be identified that tends towards an alternative set of pathways to describe the progression of patients through the model. Alterations driven by an increased understanding of the treatment area are uncontroversial as they can only add to the validity of the model.

The alternative source of change to a model's structure is the existence of insufficient or inadequate quantitative data to populate the model in its initial form. The extent to which the model structure should be based on the available data is a moot point. Sonnenberg *et al* [1994] define the practical model as 'the most detailed model that can be constructed given the limitations of available data', reflecting that such changes are 'necessary and useful compromises' pgJS54. Sculpher *et al* [2000] warn that structuring models on the basis of the quality of data available could cause the loss of important clinical events. In their view the use of expert opinion to inform parameter values is always preferable to changing the original structure of the model because the sensitivity of the results to changes in the parameter values can be assessed.

A general rule might state that extensions to the initial model informed by the identified data, which increase the detail of the model, should be accepted at this stage of the modelling process. Conversely, in cases where few data to populate certain sections of the model are identified, subtle modifications to the structure of the model may be enacted that rearrange the relationships relating to the 'difficult' parameters to reconcile the format of any available data. If such modifications fail to accommodate the available data then expert opinion may be sought to fill the void. It should be borne in mind that the process of obtaining the opinions of experts may result in such widely conflicting estimates that their inclusion detracts from the overall predictions of the model.

The extent to which the model structure will need altering will be influenced by the scope of knowledge used to inform the initial structure and the complexity of the treatment area. The reassessment of the model structure at this stage of the modelling process should be regarded as a pre-emptive effort to improve the validity of the model, which will be affected by both the model structure and the input data.

3.4.2 Harmonising the identified data

Combining the data involves the specification of the patient groups and interventions of interest, but also the tight definition of events included in the model. If large amounts of data are available it may be possible to use only the data that correspond to an exact definition of a particular event. If relevant data are not plentiful the analyst may be reluctant to discard data on the basis of subtle differences in the definition of certain events and sometimes steps can be taken to improve the comparability of data from alternative sources. The process of harmonising the data makes explicit assumptions about the differences in the data presented by different studies and attempts to revise the results of outlying studies to a baseline definition of an event. Taking a hypothetical example, the common definition of an objective response to treatment may include a stabilisation of disease, but one identified study restricted the definition of objective response to an improvement in the disease condition. If the number of stable responses within this study is reported they can combined with the positive responses. If stabilisation is not reported, then to avoid losing the data from this study an estimate of the number of stable responses could be made by fitting the proportion of stable cases reported by the remaining studies to the odd study on the basis of study characteristics.

The decision to harmonise data should be made carefully. For example, the analyst should be clear that the harmonised studies reflect the same underlying endpoint, otherwise the data may describe unwarranted variability in the observed value of an input parameter. If there appears to be underlying variation between the studies it may be necessary to control for heterogeneity using the methods presented in section 3.4.3.1. The process of harmonising the data cannot be subject to hard and fast rules as the adjustments made to the data will depend on the event described and the format of the available data. Some general principles can be identified, such as the estimation of missing values, but harmonisation is mainly determined by the individual circumstances of the modelling project. More details of the process of harmonisation are described in Chapter 5, where such methods are applied to the case study.

3.4.3 Pooling and formatting the data

The final stage of the input data analysis requires quantitative techniques to arrange the data into a suitable format for populating the decision model that allows for the

representation of uncertainty in the values of the input parameters. Unless the observed estimates for each parameter are identical it will be necessary to analyse the effect of the uncertainty in the values of the input parameters on the decision model's outputs. Methods for pooling and formatting data to populate deterministic and stochastic models are described in separate sections, following a brief discussion of alternative types of uncertainty.

Exogenous sources of uncertainty relate to the structure of the model. The US Panel on cost-effectiveness advised that such uncertainty should be tested using alternative specifications of the model employing alternative assumptions with respect to the functional form of the relationships described in the model [Manning et al, 1996]. Briggs [2000] argues that assessing the multitude of assumptions that create a decision model is beyond the scope of a single analytic team and that structural forms of uncertainty should be assessed by different researchers. However, given limited research funds it may not be feasible to employ multiple research teams to simultaneously investigate the same research topic and some form of 'internal' sensitivity analysis of the model's structure may add to the comprehensiveness of an evaluation. Examination of this source of uncertainty is beyond the scope of this thesis, suffice to say that such sensitivity analysis would add significantly to the research burden of any modelling evaluation.

Endogenous uncertainty must be handled within the experimentation of each defined model. There are two orders of such uncertainty - first and second. For any given set of input parameter values, first-order uncertainty describes the variation in the results of a decision model at the level of the individual patient. Each patient is subject to the same probabilities of experiencing an event from a particular state at any given time, so the variation observed between patients is due entirely to the 'inherent uncertainty of the probabilistic structure of the model', which can be interpreted in a similar manner to the variation observed in a clinical trial [Briggs, 2000]. First-order uncertainty can only be measured using decision models that follow individual patients through the model and record the individual estimates of costs and effects.

Running larger numbers of patients through the model will provide a more accurate representation of the mean outcomes for the population of patients, but the estimates of

variation (the standard deviations) will be a function of the number of patients run through the model. Some analysts have used the number of patients expected in a particular location in a given time period to estimate the standard deviations in order to represent uncertainty more realistically, reflecting the context of the analysis [Sharples et al, 1996; Szeto and Devlin, 1996]. However, the description of first-order uncertainty is irrelevant to the representation of uncertainty for resource allocation using cost-effectiveness data [Briggs, 2000]. Such decisions are based on the mean differences in costs and effects for the populations of patients receiving alternative treatments. It is, therefore, the uncertainty around the mean costs and effects for a particular intervention that is of interest to the decision-maker. This is termed second-order uncertainty. Second-order uncertainty is represented by the estimation of the mean value of the model outputs for alternative values of the input parameters.

3.4.3.1 Input values for deterministic models

Historically, deterministic analysis has been mostly employed in the economic evaluation of health care interventions, whereby non-random inputs lead to non-random outputs. The stochastic nature of the problem is not captured in the output of the model. Weinstein and Fineberg [1980] state that it is 'only the averaged-out, or expected, probability that matters at any given chance node. This means that the "spread" around a probability does not matter' (pg174). However, it is now commonly accepted that sensitivity analysis should be undertaken to measure how important individual input parameters are to the model's outcome. Common forms of deterministic sensitivity analysis include one-way, multivariate, threshold and extreme analyses [Briggs et al, 1994]. Other than threshold analysis, each of the methods involves specifying feasible ranges for the values of the input parameters that are being tested. The relevant point estimate for each input parameter should be specified as the weighted mean value derived from all the identified studies, whilst the values to be employed in sensitivity analyses of deterministic models should be informed by the spread of the weighted parameter values.

The methods available for weighting data depend on the format of the identified data. If good primary data that incorporates measures of the within sample variance are available then methods for weighting data based on meta-analytic methods may be

employed. The combination of survival curves from the literature adapts the meta-analytic methods slightly and is presented in Appendix 4. Alternatively, for parameter values derived from sources that do not report sample variance more subjective means for combining the data need to be pursued. The following two sections describe methods that can be employed to weight data in these two scenarios. Firstly, however, a unique scenario is addressed, which concerns parameter values for which there are no direct empirical estimates, but which may have been estimated previously using decision modelling techniques. Various options, differing with respect to the required analytic input, are available. If a number of separate studies (of adequate validity/quality) are available, employing alternative modelling techniques and structures, then each study's parameter estimate could be incorporated using equal weights. However, the analyst should check the similarity of the data inputted into each model with respect to the timing and comprehensiveness of the data collection process. If some data appear to be out of date it may be necessary to re-model some or all of the observed data using more up-to-date data obtained from the other studies, or from a fresh review. The same approach could feasibly apply to the synthesis of all modelling studies, including full cost-effectiveness studies.

Weighting parameters with known sample variance

The techniques used in meta-analysis to weight data from separate studies follow a general approach that can be adapted to most data types [Sutton et al, 1998]. The formulae for weighting data using the fixed effects and random effects models are presented in Appendix 3, as well as a test for the extent of heterogeneity in the available data. Methods for weighting data according to their assumed precision can be linked to the level of heterogeneity between the studies. If the selected studies are thought to display only limited heterogeneity then the fixed effects model provides an adequate representation of variation, representing sampling error alone. If unwarranted heterogeneity remains in the dataset, more complex methods to weight the data might be considered, such as those used in random effects models. The basic aim in using random effects models is to account for the increased variation between studies 'when assuming the studies are estimating different (underlying) effect sizes' [Sutton et al, 1998]pg69.

Using the random effects weighting procedure will always produce a greater variance in the parameter estimates, but its use is not universally accepted. Raudenbush *et al* [1994] suggested that the fixed effects model is the sensible choice if only a few studies exist as between study variation will be very poorly estimated by the random effects model. Others claim that the random effects model replaces the 'implausible assumptions of fixed effects analysis [with] untenable assumptions of its own', such as between study heterogeneity can be represented by a single variance, and that the between trial distribution is normal [Thompson and Pocock, 1991]. Peto [1987] claimed that such models move the objective (of the weighting process) to answering an unimportant question concerning randomly chosen treatments and patient populations. However, Spector and Thompson [1991] specified perhaps the most pragmatic approach, treating the random effects method as a type of sensitivity analysis looking at how the results change as the distribution of weights between the incorporated studies are altered.

Attempts may also be made to adjust the weight of a study's input according to the perceived quality of the study. Quality may be defined with respect to the study design and to the conduct of the study. For example, experimental studies are generally ranked higher than observational designs, but a poorly conducted RCT may be less reliable than a sturdy observational study [Sutton et al, 1998]. However, much disagreement exists over the use of quality weights and the only practical recommendation on this issue is that a subjective assessment of the quality of the identified studies should be made. A quality threshold of inclusion may then be employed rather than a system of differential weighting.

Weighting parameters with unknown sample variance

If within-study variance is unknown then the choice of weighting technique will depend on the sources available for the individual estimates. If data are available from primary studies that report the associated sample size informing the estimated parameter values then the respective sample sizes can be used to weight the individual estimates.

Other parameter values may be derived from secondary sources that do not report any information that can be used to objectively weight the individual estimates. Most

commonly resource use and cost data are derived from such sources, such as Trust returns or charges data. Two options are available to weight such data:

- each estimate can be assumed to be of equal validity;
- a subjective judgement of validity can be made.

Subjective judgements are analogous to the quality weights discussed in the previous section, but their use may be justified in this context because no statistical basis is available for the application of weights. The relevance to particular evaluations of alternative parameter estimates may differ according to their source. For example, estimates derived from the country in which the model is to be used to inform policy will probably be of more relevance to the evaluation and receive a larger weight.

A two-stage process for weighting data obtained from secondary sources is proposed here. The first step is to rank the available estimates in increasing order of subjective validity. The second step involves weighting the data on the basis of the ranks. The simplest approach would be to weight the data linearly with respect to the ranks. For example, from a series of 10 estimates the top ranked estimate would receive a weight of 10, whilst the least confident estimate receives a weight of 1. Alternatively, the assigned weights could be specified with respect to a baseline value that is taken to be most (least) relevant point estimate. For example, if estimate x is given a weight of 1 and estimate y is thought to be half as relevant as x then it receives a weight of 0.5.

Using data derived from expert opinion, the mean value specified by either process (Delphi panels or expert groups) will represent parameter point estimates. To represent uncertainty the individual estimates produced by the Delphi method automatically provide a range of values, whilst the expert group can be asked to specify a feasible range.

3.4.3.2 Input values for stochastic models

A number of papers have described the use of probabilistic sensitivity analysis in economic evaluation [Felli and Hazen, 1998; Lord and Asante, 1999; Pasta et al, 1999]. Alternatively labelled as stochastic cost-effectiveness analyses [Briggs, 1999], the values of the input parameters within such models are described as probability

distributions. The objective of a stochastic analysis is to obtain a distribution for each of the model's outputs that is informed by randomly sampled sets of input parameter values from the specified probability distributions. Stochastic cost-effectiveness analyses incorporate the Bayesian paradigm where the parameter is seen as a variable, with its own probability distribution, since the true value of the parameter is unknown. In the Bayesian context the input parameter distributions are posterior distributions informed by the available data [Parmigiani et al, 1996], which represent second-order uncertainty as defined above. As long as there is uncertainty about the particular value of a parameter, the parameter is treated as a variable and probabilities are used to express the uncertainties [Iverson, 1984].

Stochastic analyses may be viewed as the ultimate application of multivariate deterministic sensitivity analyses, where the aim is to test the impact of every combination of parameter values for every parameter within the model. If the definition of uncertainty for each of the input parameters is valid, then stochastic analyses present the most comprehensive and appropriate form of sensitivity analysis.

Four methods for defining probability distributions to represent the sampling distributions of the mean values of the input parameters are described in the following sections. There is no consensus on the appropriateness of the four methods. The first three methods (theoretically defined distributions, empirical distributions, and fitted distributions using statistical fitting software) were applied to the case study. The applied methods are described in Chapter 5, the alternative results in Chapter 8, whilst the implications derived from the use of the alternative methods are discussed in Chapter 9. The fourth method (bootstrapped primary data) applies to parameters informed by patient-level data and is not relevant to the employed case study.

Theoretical distributions based on parameter type

The first option applies purely theoretical considerations to the choice of probability distribution for different categories of input parameters. The characteristics of the different types of parameters included in a decision model are examined and a probability distribution with properties that match those of the input parameter is assigned. The same type of probability distribution is applied to groups of input

parameters. The full process for choosing appropriate probability distributions and fitting the distribution parameters is described in Appendix 5, though an abbreviated narrative is provided below.

The choice of probability distributions to describe alternative parameter types is an individual choice, as is the categorisation of the input parameters. The following exposition was formed through discussion with colleagues and with reference to Bayesian statistics texts [Iverson, 1984; Berry and Stangl, 1996; Gelman et al. 1995]. Four categories of parameters were identified - proportions, survival times (length of time to the next event), costs and utility values, for which appropriate probability distributions were chosen:

1. Proportions describe the probability that a patient will experience an event, which is bounded by 0 and 1. The beta distribution, which is bounded by 0 and 1, provides the most realistic representation of proportions, as it can lie in a wide variety of shapes when the two distribution parameters (alpha and beta) are varied [Iverson, 1984];
2. Survival times describe the length of survival (or time to next event), they are bounded by 0. The gamma distribution is bounded by zero and approximates the normal distribution at large samples. It is also extremely flexible, using a shape parameter to describe the available data accurately [Rice, 1995];
3. Cost parameters have been described by the lognormal distribution [Pasta et al, 1999; Fenwick et al, 2000], though the gamma distribution may provide a more flexible description of the sampling distribution of costs [Parmigiani et al, 1996];
4. Utility values portray similar properties to a proportion though 0 and 1 do not strictly bind them. The beta distribution is still advocated as a scale parameter can be fitted to the beta distribution to incorporate a larger range than 0 to 1.

The method of moments specifies formulae for estimating parameters for alternative probability distributions by finding expressions for them in their lowest order moments, then substituting sample moments into the expression [Rice, 1995]. The available data can be combined directly with these formulae to estimate relevant parameters for the chosen probability distributions.

Empirical distributions

The easiest approach to specifying a probability distribution is to use the collected data 'as is', this is known as a trace simulation [McHaney, 1991]. However, the individual parameter estimates should be incorporated on the basis of their relative weights as estimated using the methods described above. A weighted dataset comprising replications of each parameter estimate according to their respective weights can be created that may be inputted directly in to the decision model. The inverse of the variances may be large, so a consistent method for creating weighted datasets is to limit its size to 100 observations. The following steps will create such a dataset:

1. Calculate the weight for each of the identified parameter estimates W_i, where $i = 1 \ldots n$;
2. Divide each W_i by the sum of the W_i and multiply by 100 (value = X_i);
3. Create X_i copies of each parameter estimate.

Using the available data directly means that the model is based on exactly what has been observed in the past, which may increase the credibility of the model's conclusions. However, the fact that no values other than those observed could be sampled may misrepresent the full set of possible values. This is particularly likely if a limited dataset is available [Kelton et al, 1998]. Empirical distributions have been used in previous stochastic models when the available data were extensive [Parmigiani et al, 1996].

Fitted distributions

Rather than using the observed data directly, probability distributions can be specified as a projection of the sampling distribution of the parameters. Lipton *et al* [1995] describe the manual process for specifying distributions and estimating the respective parameters. However, there are an increasing number of computer-based packages that will analyse the available data and choose the best fitting distributions, as well as the appropriate distribution parameters [Crystal Ball, 2000; Stat::Fit, 1996]. The aim of this procedure is to select a probability distribution with random samples that is indistinguishable from the collected data [McHaney, 1991]. The choice of distribution can be tested on the basis of goodness-of-fit tests, such as the chi-square and Kolmogorov-Smirnov tests, though these tests are known to have very low power so the

probability of rejecting a fit is small, even when the distribution postulated is wrong [Bratley et al, 1987].

It may be appropriate to consider the properties of computer-based distributions, as the software does not have knowledge of the intended use of the specified distribution. These problems can be minimised by careful consideration of the characteristics of different probability distributions. Most statistical fitting packages allow the user to select a range of theoretically sound distributions, so improper distributions can be excluded at the start of the process. Alternatively, restrictions imposed within the decision model can be used to block the use of infeasible parameter values within the model. Fitted distributions allow a more rounded description of the data, though it is possible to sample values that are not feasible or to lose important characteristics of the data such as sequential patterns [Kelton et al, 1998].

Bootstrapped distributions

The definition of bootstrapped distributions involves taking a large number of repeated samples of size n, with replacement, from a dataset with n observations. The set of mean values provides an empirical approximation to the sampling distribution of the 'true population' value [Lord and Asante, 1999]. Pasta *et al* [1999] recommended using the bootstrap sampling distribution directly. This approach was deemed to have a number of advantages including the fact that the estimated value would never lie outside the range of values identified in the available studies. It was also noted that the divergence of the bootstrap sample reflects the divergence in the original sample. Alternatively, the bootstrap dataset could be used as the basis for the choice of a fitted distribution. If the bootstrap sample is normally distributed then confidence intervals calculated parametrically can be applied to a normal distribution [Lord and Asante, 1999], otherwise a fitted distribution can be specified using the bootstrap sample as discussed in the previous section.

Unfortunately, when the number of different observations from the literature is small the bootstrap means do not differ greatly from the mean of the weighted dataset. Bootstrapped distributions are only applicable in cases where a large number of observations are available [Lipton et al, 1995]. Lipton *et al* give no indication of what

constitutes a small sample, which would appear to be an empirical question that relates to the variation within individual datasets. The main circumstance in which bootstrapping should be used to inform input probability distributions is when patient-level data from a primary study is available. Variation between patients represents first-order uncertainty, but model parameters should represent uncertainty about the value of the population mean (second-order uncertainty). Bootstrapped distributions based on patient-level data represent second-order uncertainty.

3.5 Implementing the model

The implementation of a decision model refers to the act of preparing an analysable form of the model through the transfer of the theoretical model to a computer-based software package. This stage of the process does not question the states included in the model or the relationships assumed between them, rather it is concerned with the optimal strategy for describing the structure of the model. Consideration of the appropriate modelling technique should only be addressed at this stage of the modelling process. The current state of knowledge regarding the choice of modelling technique was discussed in Chapter 2. The decision should be based on a number of factors, relating to both the characteristics of the treatment area and to the research resources available. If the final choice of modelling technique does not accommodate all of the prior assumptions made with respect to the description of patient pathways, then modifications to the model structure can be implemented.

The topics covered in following sections include the process of building a computer-based model, the verification of the model to check the internal operation (section 3.5.2), and the validation of the model where it's realism is tested (section 3.5.3). The methods described are applied to the case study in Chapters 6 (building the model) and 7 (verifying and validating).

3.5.1 Building the model

Sonnenberg *et al* [1994] and Detsky *et al* [1997] provided a series of good practice guidelines towards the construction of a decision model. For example, it was recommended that embedded decisions within the model should be avoided. Though

decisions made within the model can be automated, separate analyses of the model will be required for each estimate of the decision-makers criterion for adopting the more effective technology. For stochastic models especially, the required time for experimentation may be infeasibly extended. Linkage between probabilities within models was also suggested as good practice. For example, where the probability of an event influences separate events within the model that probability should be incorporated into the formula for the probabilities associated with each of the respective events. The common probability can then be varied simultaneously, during stochastic or deterministic sensitivity analyses, for each of the associated probabilities.

Three additional issues must be considered in models covering extended time horizons. Firstly, allowance must be made for the outputs of the model to be adjusted with respect to differential timing. The discounting of costs is an uncontroversial issue, though the theory underpinning the discounting of health benefits is less secure it is advised that such model outputs also require discounting [Lipscomb et al, 1996]. The discount rate will not necessarily be the same for costs and effects so separate discount rate parameters should be specified [Briggs and Sculpher, 1998].

Secondly, using a time-oriented model, i.e. a Markov model, a cycle length will need to be specified. The length of the cycle should be chosen to represent a clinically meaningful time interval [Sonnenberg and Beck, 1993]. If the available data describes the probabilities of events over longer time periods than that chosen for the cycle length they will need converting to probabilities that describe the chance of an event in a shorter period. This conversion requires that the original probabilities be transformed to rates of incidence, which inform the instantaneous probability of an event rather than the probability of an event occurring over a specified period. The rate is then used to calculate the probability of an event over the shorter period of time [Sonnenberg and Wong, 1993]. If it is assumed that transition rates remain constant over the longer period, the formula 3.1 can be used to convert the probabilities directly [Miller and Homan, 1994].

$$P_i = 1 - [1 - P(t_i, t_{i+1})]^{1/i} \qquad (3.1)$$

where $P(t_i, t_{i+1})$ is the probability of an event between time period i and i+1, and i represents the number of shorter time intervals within the originally specified time period. The assumption of constant rates of occurrence is often acceptable, but if there are good reasons to assume that rates are not constant then the longer period should be split into shorter periods that best reflect the expected variability in rates. For example, converting annual mortality data to monthly probabilities it may be expected that patients are more likely to die in the latter half of any particular year. If so, the proportion of patients dying in the full year can be split unevenly between the first and second six months, applying the above formula to each period.

Markov models are run so that movement between states within the model occurs between cycles so according to the model all patients leaving state x in year 1 leave at the end of year 1. This will naturally overestimate the time spent in state x so a half-cycle correction should be applied to each transition on the basis that, on average, each patient leaves state x after 6 months [Sonnenberg and Beck, 1993]. Also, median or mean survival estimates must be converted to probabilities based on the choice of cycle length. This involves calculating the rate of events over the whole period and then converting the rate to the correct probability, an example, assuming a monthly cycle length and an average survival of 30 months, is provided within formula 3.2.

$$Rate = {}^{1}\!/_{30} = 0.03^{*} \Rightarrow \Pr obability = 1 - e^{-0.03^{*}} = 0.0328 \qquad (3.2)$$

DES models are more flexible in their handling of time, but such models still require the specification of a minimum time period of advancement, which should also be chosen on the basis of a clinically meaningful time interval. The chosen period defines the shortest time that patients can remain in any given state within the model. A half-cycle correction is only required in a DES model if the length of time spent in a state is not sampled from a continuous distribution.

Operations research textbooks provide advice on the computer implementation stage for DES models [Pidd, 1989]. To manage the additional complexity of DES, models should be built in stages, or modules. Each module should represent a definable section that can be run separately from the remainder of the model. This approach allows the

location of any bugs in the programming to be narrowed down and identified more efficiently. If there are difficulties in the translation of the modelling assumptions into the computer-based software they should be represented openly, not buried. In all models, it is recommended that a log should be kept during the process of building a model to record the assumptions made within the model, as well as potential sources of bias (and their likely direction) [Halpern et al, 1998].

3.5.2 Verification

The processes of verification, followed by validation, are the final steps prior to the actual use of a model to generate useful results. Verification involves checking that the model is working in an internally consistent manner, i.e. that the model is free of programming bugs. Three sequential phases of verification are presented below covering the verification of logic, sensitivity testing and stress testing [Bratley et al, 1987].

Verification of logic involves running models with values for which there is a logical expected result that could easily be estimated without the model. The results produced are then compared to those expected. In a stochastic model, for example, all patients can be set to move from a state in a single specified time period, and within that state the cost and utility weights can be set at a constant value. The process of verifying model logic can be split into two categories covering the clinical parameters, and cost and utility parameters. The latter parameters are grouped together because they both involve the assignment of weights to the time spent in different health states. Finally, the logic relating to discounting both the costs and the effects must also be verified.

The second type of verification is labelled sensitivity testing, which involves varying one parameter, whilst keeping all others fixed in order to check that the behaviour of the model is sensible. The definition of sensible model behaviour is subjective and must be redefined for each parameter within the model. It may be difficult to judge the absolute magnitude of the effect of different parameters within the model, so sensitivity testing may limit the definition of sensible behaviour to a prediction of the relative magnitudes of the individual parameter effects. For example, the observed and expected relative effect of individual parameters on the variance of the model's outputs can be compared.

Such predictions may simply encompass an ordinal ranking of the expected outputs, or an attempt to incorporate cardinality between alternative parameters.

A secondary objective of this form of verification is a deeper understanding of the model and it's workings. If the results of the sensitivity testing are unexpected then the working of the model should be explored. The inspection may lead to the discovery of a programming error or to an admission that the original prediction was based on a false premise. The debasing of a false assumption can only improve the analyst's understanding of how the parameters within the model behave, which may inform the process of experimentation.

Stress testing is the final form of verification. Such testing searches for extreme errors, such as the acceptance by the model of infeasible parameter values, such as probability parameters set to values over 1. The process of stress testing involves setting parameters within the model to strange values and checking that the models 'implodes'. The likely sign of any implosion is not dramatic, an error message usually suffices.

3.5.3 Validation

Validation is 'the process by which the modeller and the client satisfy themselves that the model ... is suitable for use within its defined experimental framework' [Pidd, 1998]pg33.

In the OR field, it is often the case that the input data for a model are derived from the 'real' system that is also used to validate the model. The usual objective of an OR project is to represent the system of interest, and then undertake a series of 'what if?' experiments in order to understand their impact on the future behaviour of the system. Though problems remain around the collection of data from the real system [Kelton et al, 1998], the process is less complicated than for economic decision models used in HTA. Economic HTA decision models commonly wish to represent a system in order to make decisions on the present behaviour of the system. 'What if' analyses may be undertaken to represent the uncertainty in the main output, but they are not the primary output. OR studies generally concentrate on structural validity, whereas in economic

HTA modelling projects, the data used to populate the model are also subject to the process of validation (content validity) [Halpern et al, 1998].

In HTA possible sources of comparative observations that have been suggested include previously developed models [Halpern et al, 1998], and intermediate outputs that are more readily available [Sculpher et al, 2000]. Close comparisons with alternative models improve confidence in the validity of both models because it is less likely that two (or more) modelling teams will produce inaccurate results than just one group. Conversely, where differences are apparent comparisons of the separate models may highlight disparate assumptions that require resolution [Paul and Balmer, 1993].

One subjective approach to validation is the use of 'Turing's test'. The output from the model and from the 'real' world is shown to experts who then try to distinguish between them. If they are able to differentiate between the two sources then the explanation behind the correct definition of the outputs may inform the revision of the model [Paul and Balmer, 1993]. Such an exercise is intuitively appealing and could be conducted using the views of clinicians, though such a process could be time-consuming and not guaranteed to produce a concordant result.

There are obvious problems with all of the above validation methods, for example, an accurate prediction of 5-year survival does not necessarily validate the required output of the model, which may be the gain in QALYs over a patient's lifetime. However, the process of validating decision models of alternative health care interventions will always be an imprecise exercise. Indeed, the hypothetico-deductive approach states that a model can not be proved valid in any true sense, a valid model being defined as unrefuted for certain scientific purposes [Popper, 1965]. If the model is refuted during the process of validation then the current approximations to reality contained within the model need to be changed. Sometimes additional parameters, which were previously believed to be unimportant, will need to be included. Relationships between variables may have been ignored, either several possibly separate parameters could have been aggregated and treated as one, or independence between variables may have been assumed. Otherwise, functional relationships within the model may have been simplified, nonlinear functions may have been assumed to be linear. Finally, it may

have been assumed that the values of parameters within the model were stationary over time, whereas they are actually dynamic [Bratley et al, 1987].

Sculpher *et al* [2000] and McCabe and Dixon [2000] considered the issue of validating economic HTA decision models in detail and both sets of authors came to similar conclusions that the creation of objective tests of validity are unlikely. Instead, both papers proposed that a framework for model development is required that would enable an explicit method of assessing the quality of decision models. However, it should also be expected that the possible avenues of validation are undertaken and that the outputs of the model are judged fairly against the available data.

3.6 Experimentation

This final phase of the modelling process involves the evaluation of the model, and the analysis of the model's outputs. Experimentation with deterministic decision models is relatively straightforward, though a more subjective interpretation of the outputs is required. Most effort in the following sections, therefore, is given to describing experimentation with stochastic decision models.

The process of evaluating the model comprises the basic actions for collecting the model's outputs for a particular set of inputs. The first section below describes the methods for collecting the relevant output data from the two main types of stochastic models – Markov models and DES models. The following sections describe methods for analysing the model and the output data that relate to the objective of the evaluation. Two main objectives are possible.

1. to inform an immediate resource allocation decision, where the only economic data available are those included in the decision model.
2. to advise on the collection of further data with which a more informed resource allocation decision can be made.

3.6.1 Evaluating stochastic decision models

The usual objective of a stochastic evaluation is the estimation of the relevant outputs describing a population mean and representing population uncertainty, otherwise

referred to as second-order uncertainty (see section 3.4.3). The following steps describe the associated process for evaluating both a cohort-based Markov model and a DES model:

1. Randomly sample a set of parameter values from the input probability distributions;
2. Undertake a 'run' of the model and collect the mean value for each of the model's outputs;
   - a 'run' of a cohort-based Markov model produces an exact solution for each model output for the sampled set of input parameter values;
   - a 'run' of a DES model involves sending a large number of individual patients through the model (see section 3.6.1.1), where each patient is subject to the same set of sampled parameter values. The variation between patients is the first-order uncertainty associated with the sampled set of parameter values, which will vary according to the number of patients specified. Only the mean values of each of the model's outputs are required from a run.
3. Redo steps 1 and 2 until sufficient runs have been completed to inform a 'trial', which is defined as a set of 'runs' (see section 3.6.1.2);
4. Analyse the data within the 'trial' to describe the second-order uncertainty around the model outputs.

3.6.1.1 Estimating the required patients within a model 'run'

A great advantage of the cohort-based Markov model is that first-order uncertainty is controlled because an exact solution is produced for each set of parameter values. The stochastic evaluation of second-order uncertainty using a DES model should only be undertaken when the analyst is certain that the first-order estimates are sufficiently precise, because the aim is to remove first-order uncertainty from the model's outputs. To reduce the impact of first-order uncertainty within a DES model the corresponding model outputs should be based on the mean values of a large number of patients for each set of parameter values [Stinnett and Paltiel, 1997]. However, the running time for the second-order analysis of a DES model is influenced by the number of patients included in each run, so the required number of patients should be the minimum number that provide an adequate level of precision in the first-order estimates.

Decision models, such as DES models, that represent uncertainty around individual patient's outputs are termed first-order Monte Carlo models [Halpern et al, 2000]. To get a better idea of the necessary number of patients previously published studies that employed such models were sought. Unfortunately, the identified first-order Monte Carlo economic HTA decision models did not analyse second-order uncertainty stochastically by sampling sets of input parameter values, but rather estimated the mean values of the model's output for a specified set of parameter values [Habbema et al, 1987; Hart et al, 1997; Paltiel et al, 1998; Urban et al, 1997; Warner et al, 1996]. Sample sizes of up to 100,000 were employed in these evaluations, but such large numbers were not constrained by computing time because the models were evaluated only a few times (alternative sets of parameter values were only specified as part of deterministic sensitivity analyses).

Unlike the economic HTA literature, a number of applications of first-order Monte Carlo models employing probability distributions to represent second-order uncertainty were identified in the general clinical literature. Bagust *et al* [1999] simulated bed use in accommodating emergency admissions. Each run of the model comprised a 1000-day period with a mean daily admission rate of almost 25, so each run included 25,000 patients. Davies and Roderick [1998] used DES to aid the planning of resources for renal services. Their main analysis ran a population group one-tenth the size of the UK population over a period of 50 years. Michel *et al* [1996] modelled the planned health care facilities for neonatal extracorporeal membrane oxygenation. Their methodology is slightly unclear, but it appears that they ran the model for the period of 1 year incorporating all births in each year, replicated 1000 times. Cronin *et al* [1998], modelling cancer screening, simply chose 10,000 patients as an adequate number.

Most of the identified studies based run size on some measure of actual activity in the country of origin. Intuitively, this appears unnecessary. The purpose of sending a large number of patients within each run is to establish an accurate mean value for the outputs associated with each sampled set of input parameter values. On closer inspection, however, it is apparent that the representation of actual activity appears to be chosen on the basis of the number of patients it generates. The superficial representation of actual activity is disingenuous because it hides the real reasons for choosing the appropriate number of patients to be run through the model, which should be to reduce first-order

uncertainty to an acceptable level (however defined). However, the artificial representation of reality may be a means of selling the results of the simulation to the intended audience.

The necessary number of patients to reduce first-order uncertainty to an acceptable level is related to the complexity of the decision model. The more complex the model the more possible pathways there are for the patients to experience and the greater the extent of first-order uncertainty, which means a greater number of patients will be required to achieve an adequate level of precision. An adequate number of patients may be informed by undertaking a number of preliminary runs for a range of different sample sizes that are all subject to the same input parameter values. For example, 1000, 5000 and 10000 patients could be run through the model 50 times. If there were no first-order uncertainty the same mean values for the model's outputs would be expected for each run, in the presence of first-order uncertainty the mean values will vary. Plotting the mean values of the model's output according to the number of patients included in the individual runs provides an indication of the level of precision offered by the different sample sizes, which can be used to inform the final choice of run size. An empirical investigation of the required number of patients within a run is presented in section 6.5.

3.6.1.2 Estimating the required model 'runs' within 'trial'

As described above a 'trial' comprises a series of 'runs', each of which represents the mean values of the model's outputs for a randomly selected set of input parameter values. The adequate representation of population uncertainty (second-order uncertainty) should only be addressed after the analyst is certain that first-order uncertainty has been controlled.

Papers advising on the use of stochastic analyses in the economic HTA literature have not discussed the requisite number of runs [Parmigiani et al, 1997; Felli and Hazen, 1998]. Applied studies using probabilistic sensitivity analyses have tended not to discuss the issue, but common practice appears to have settled on the use of 1,000 runs to represent second-order uncertainty [Kuntz et al, 1999; Sisk et al, 1997; Kattan et al, 1995], though others have employed 10,000 runs [Fenwick et al, 2000]. More complex

decision models will have wider dispersions of outputs, which will require larger numbers of observations (runs) to accurately represent uncertainty. On that basis, the distributions of the model's outputs should be visually inspected during the evaluative process and additional runs should be undertaken until the definitions of the distributions are judged adequate to describe population uncertainty.

3.6.2 Analysis to inform immediate allocation decision

If resource allocation is to be based on the results of an economic evaluation it is important that the economic data is presented in a format that aids the decision making process. The methods available for presenting the results of economic evaluations have developed greatly during the last few years. The following section describes the progression of methods for representing uncertainty in economic evaluations, moving from deterministic sensitivity analyses to the estimation of confidence intervals around incremental cost-effectiveness ratios (ICERs) to the use of the net benefit statistic and cost-effectiveness acceptability (CEAc) curves.

Until recently, the results of most cost-effectiveness analyses, as opposed to cost-benefit analyses, were simply presented as the incremental cost per additional unit of effectiveness, using formula 3.3.

$$\frac{(\mu_{C1} - \mu_{C0})}{(\mu_{E1} - \mu_{E0})} \qquad (3.3)$$

Where μ_{Ci} and μ_{Ei} are the mean values of the costs and effects of intervention i, $i = 0$ and 1. In the absence of uncertainty such a ratio would be adequate – the decision-maker could judge whether the additional health benefits outweighed the additional costs and the decision would be made. Unfortunately, uncertainty is a major factor in the evaluation of any clinical intervention, and especially in economic evaluation. Briggs *et al* [1994] identified four separate sources of uncertainty, though Briggs and Gray [1999] later categorised uncertainty as relating either to the underlying methodological framework or to the data used in the evaluation. The former issue is not relevant to this discussion (the definition of a framework has been the topic of this

Chapter), but the handling and expression of uncertainty surrounding the data is central to the presentation of the output from a economic HTA decision model.

Previous evaluations that attempted to describe uncertainty around the point estimates of ICERs concentrated on specifying ranges within which the true ratio was likely to be found. Most studies were deterministic and the commonest approaches to sensitivity analysis specified alternative values (from the base case) for input parameters, either individually or jointly, and then recalculated the model outputs. Such methods - one-way, multi-way and extreme scenario analyses – result in a high and low estimate of the ICER that is presented as the plausible range. A variant on this theme is threshold analysis, which requires a value to be placed on an additional unit of effect. The value of an input parameter is varied until the allocation decision switches from one alternative to another, ie. a threshold is reached [Briggs et al, 1994].

The earliest identified method for the stochastic analysis of sensitivity was labelled probabilistic sensitivity analysis [Doubilet et al, 1985]. Probability distributions were specified around each parameter within the model and following multiple simulations of the model the mean and standard deviation of the expected utility of each strategy were recorded (costs were not included in this model). A later review of sensitivity analysis in economic evaluation found that very few had 'dealt adequately with the problem of uncertainty' [Briggs and Sculpher, 1995]pg363. Probabilistic sensitivity analysis was only identified in 1 of the 75 studies that had undertaken any form of sensitivity analysis, though even in this study the authors of the review stated that the representation of uncertainty around the ICER was unclear.

A number of approaches have been developed to represent interval estimates of the ICER using stochastic data, the alternative methods can be grouped into three categories – Taylor-series intervals, Fieller's method confidence limits, and bootstrap confidence intervals [Heitjan et al, 1999]. Several studies have compared the various approaches, and, though the Fieller theorem and bootstrap methods are preferred to the Taylor-series, general shortcomings in the estimation of interval estimates around the ICER have been noted [Heitjan et al, 1999; Briggs et al, 1999; Polsky et al. 1997]. The underlying and unresolved problem is that ratio statistics, in which the denominator can take very small values, cause problems for the estimation of confidence intervals

[Briggs, 1999]. Using bootstrapped samples the representation of negative values of differences in effects, combined with positive differences in costs reduces both upper and lower confidence intervals because such results are represented as a negative ICER and misplaced at the left of the ICER distribution [Heitjan et al, 1999].

Due to the problems surrounding interval estimation around the ICER alternative approaches have been sought, which led to the definition of the net benefits statistic in 1998 [Stinnett and Mullahy, 1998; Tambour et al, 1998]. The calculation of net benefits requires the assumed knowledge of the decision-makers maximum willingness to pay for an additional unit of effect, λ. The incremental net benefits (INBs) of a one intervention, t_1, over another, t_0, can be calculated using formula 3.4.

$$INBs = \lambda(\mu_{E1} - \mu_{E0}) - (\mu_{C1} - \mu_{C0}) \Rightarrow (\lambda\mu_{E1} - \mu_{C1}) - (\lambda\mu_{E0} - \mu_{C0}) \qquad (3.4)$$

where μ_{Ci} and μ_{Ei} are the mean costs and effects of intervention i, i = 0 and 1. The formula also illustrates that the INBs are the difference between the mean net benefits of the interventions, so the decision rule is to allocate resources to the intervention with the highest mean net benefits. Parametric and nonparametric methods for the estimation of confidence intervals for net benefits have been described. The use of the net-benefits statistic to format the output is much more convenient than the ICER statistic [Briggs, 1999], mainly because the net benefits statistic is linear in costs and effects [Stinnett and Mullahy, 1998]. The calculation of confidence intervals around a single value of λ is of limited applicability because the value of λ is likely to vary for different decision-makers. One approach to the presentation of uncertainty around the net benefits statistic that overcomes the need to specify a single value representing λ is the specification of a CEAc curve.

The initial implementation of the CEAc curve was based on the estimation of confidence intervals for ICERs, the CEAc curve presented the probability that an ICER was under λ for non-negative values of λ [van Hout et al, 1994], though the concept has since been linked to the presentation of net benefits [Stinnett and Mullahy, 1998; Briggs, 1999]. To define a CEAc curve for net benefits using the output of a stochastic

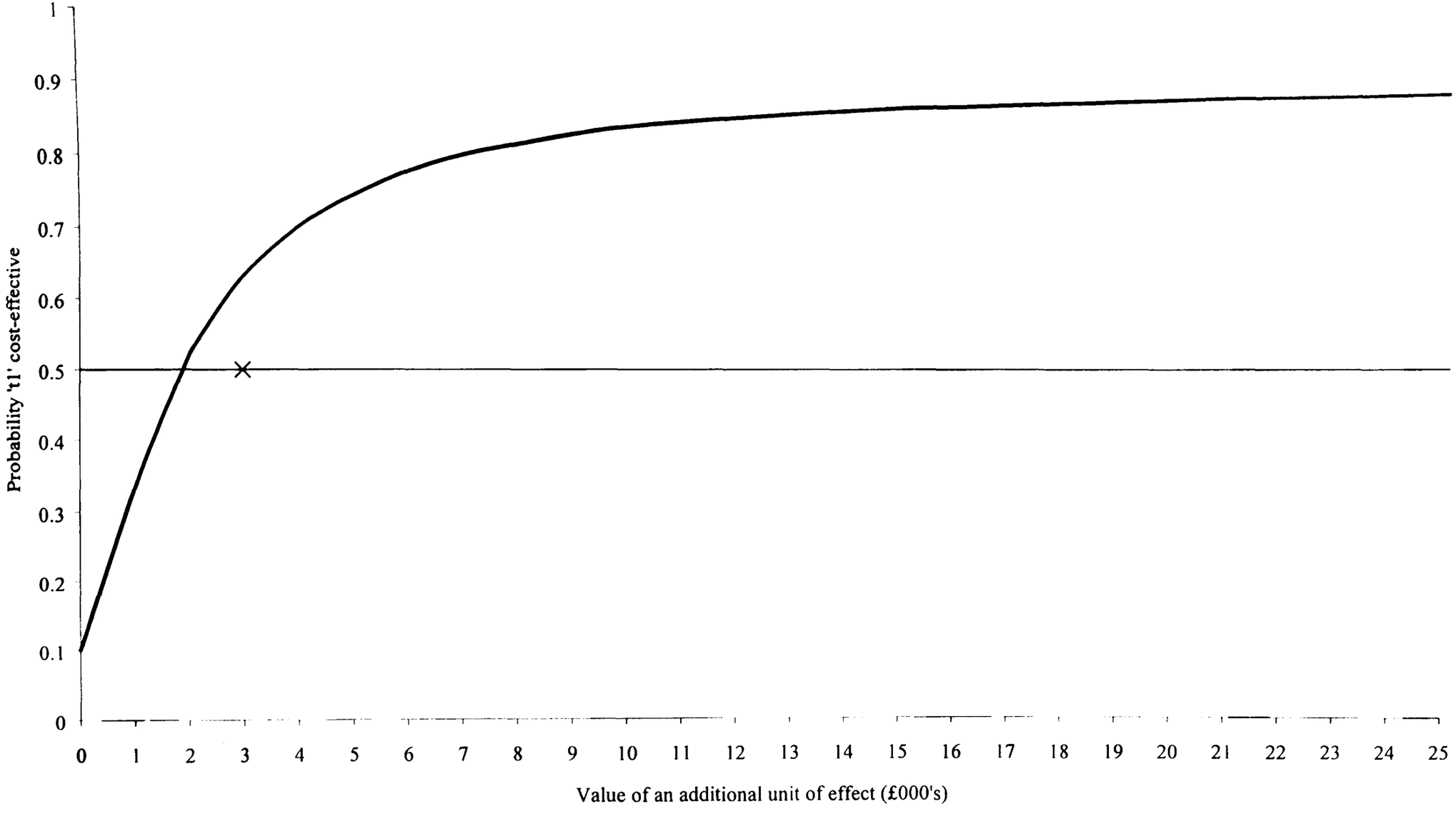
decision model, assuming multiple interventions, the following steps should be taken [Fenwick et al, 2000]:

1. Undertake a large number of model runs and calculate the mean net benefits associated with each intervention for a wide range of non-negative values of λ;
2. For each value of λ calculate the probability that each intervention has the highest mean net benefits;
3. Plot the probabilities as a function of the value of λ to obtain a family of CEAc curves.

CEAc curves represent the probability that the INBs for one intervention over another are positive as a function of the value of λ. A hypothetical CEAc curve is presented in Figure 3.1, a horizontal line has been added to the figure to highlight the point at which intervention t_1 becomes the intervention more likely to be cost-effective. From the curve, the probability of intervention t_1 being the cost-effective option reaches the 0.5 threshold at a value of λ of just under £2,000. The intuitive interpretation of this hypothetical CEAc curve, regarding the allocation of resources to intervention t_1, would state 'allocate resources to intervention t_1 if the value an additional unit of effect is at least £2,000'. Unfortunately, such a decision rule maximises the probability of choosing the optimal intervention, but it does not necessarily maximise the expected benefits of the available resources. If the distributions of net benefits are skewed then the intervention with the highest probability of net benefits will not necessarily offer the highest expected net benefits [Fenwick et al, 2000].

A slight adaptation to the conventional CEAc curve is to plot, on the pr(0.5) line, the value of λ at which the sum of the INBs becomes positive for intervention t_1. The presentation of the absolute level of the INBs alongside the probability of positive INBs draws attention to any discrepancy between the objectives of choosing the strategy with the highest probability of positive INBs and maximising the expected benefits. In Figure 3.1, X marks the value of λ at which the sum of the INBs becomes positive for intervention t_1 - £3,000. The objective of economic evaluation is to maximise health benefit, given the available resources; so the decision to allocate resources should be made on the basis of expected net benefits [Fenwick et al, 2000]. In the hypothetical

Figure 3.1 A hypothetical cost-effectiveness acceptability curve



example, therefore, the decision rule becomes 'allocate resources to intervention t_1 if the value an additional unit of effect is at least £3,000'.

An alternative method for representing uncertainty around nets benefits involves the estimation of confidence intervals for the INBs between two interventions, t_1 and t_0. The minimum value of λ at which the INBs are statistically significant from zero can then be found [Tambour et al, 1998]. Similarly, Stinnett and Mullahy suggest plotting point estimates and confidence intervals for the net benefits over a continuous range of λ [Stinnett and Mullahy, 1998].

Undertaking the analysis described above provides a good representation of the uncertainty surrounding the relative cost-effectiveness of the alternative interventions for the chosen model structure, the set of probability distributions describing the uncertainty in the values of the input parameters, and the discount rates for the costs and effects. Additional analyses may be necessary if any of these factors are also subject to uncertainty.

3.6.3 Analysis to inform further data collection

It is well documented that there is an excess of technically feasible and potentially beneficial health care interventions over the economically possible [Buxton, 1993]. The surplus of desirable health care has inevitably led to an increase in the demand for evidence to inform the enforced process of choosing between interventions. Funds for the evaluation of health care technologies are also limited and programmes of research should be developed on the basis of cost-effectiveness [Harper et al, 1998]. To date, it appears that explicit criteria for the allocation of research funds have been applied only qualitatively in the UK [Medical Research Council, 1994; NHS Executive, 1997], though a review of the literature on the preliminary economic evaluation of health care technologies identified five separate quantitative approaches to the prioritisation of research funds [Harper et al, 1998]. The primary issue in all the approaches concerned the estimation of the effect of future research on health policy or treatment practice, and hence, the potential health benefits within the relevant population. Most approaches recommended that previous research should

inform predictions of the results of future research though the analytic level of inclusion varied.

Eddy [1989] and Townsend and Buxton [1997] proposed the detailed description of possible policy scenarios, incorporating the associated costs and health benefits, resulting from a prospective trial, but provided the least detail on the estimation of the likelihood of the different scenarios occurring. Drummond *et al* [1992] demonstrated their approach by undertaking a retrospective analysis whereby the results of the 'prospective' trial were known, though they stated that the increased benefits derived from the new intervention should be accessed from the data employed in sample size calculations. Given the power to detect a stated improvement in outcomes, it was suggested that the effect on clinical practice should be derived from the views of practitioners applying the alternative therapies. Detsky [1989] proposed a more formal application of the available data, which, for a given sample size, linked the power of the prospective trial to the prior probability distribution of the different effect sizes.

The fifth study adopted a decision analytic approach to the prioritisation of research that accounted for the existing uncertainty in the values of all parameters affecting the cost-effectiveness of the alternative interventions [Claxton and Posnett, 1996], which has since been labelled Bayesian value of information (VoI) analysis [Fenwick et al, 2000]. Harper *et al* [1998] questioned the practicality of this last approach, which was based on parametric methods. More recent expositions of Bayesian VoI analyses have developed equivalent non-parametric techniques [Fenwick et al, 2000; Felli and Hazen, 1998; Felli and Hazen, 1999], which ease the analytic burden. However, the application of the full VoI methodology is rare, the only identified publication described a hypothetical parametrically derived VoI analysis [Claxton, 1999]. Work is ongoing on the conduct of the non-parametric analysis of the VoI, and the process of analysis described here is a product of the author's work through collaboration with experts in the field (Fenwick, Briggs and Claxton).

The following sections describe the methodology for a full Bayesian VoI analysis, which comprises three main stages. Firstly, the expected value of perfect information (EVPI) is generated for the evaluation as a whole, or for individual parameters.

Knowledge of the EVPI for individual parameters within the model provides a measure for assessing the value of eliminating the uncertainty associated with each parameter. Secondly, integrating the process of estimating the EVPI with assumptions about how additional sample information will affect the probability distributions for the input parameters provides an estimate of the expected value of sample information (EVSI). Finally, combining the EVSI with the expected cost of collecting additional sample information allows the estimation of the expected net benefits of sampling (ENBS) [Claxton and Posnett, 1996; Claxton, 1999]. The process for estimating the EVPI and the EVSI, which leads to the estimation of the ENBS is described below, whilst the full methodology is applied to the case study and the results reported in Chapter 8.

3.6.3.1 Estimating the expected value of perfect information (EVPI)

The EVPI is defined as the difference in the expected payoff of decisions using perfect information and the payoff using the currently available information, which is a function of the value of an additional unit of effect. To estimate the EVPI a large number of iterations of the decision model are required (as discussed in section 3.6.1.2). Each iteration provides a separate observation of the mean net benefits of treatment 1 (t_1) and treatment 0 (t_0), informed by randomly sampled sets of input parameter values. If there is uncertainty about which treatment option really is cost-effective (at the chosen value of λ) some observations will demonstrate that t_1 has larger mean net benefits, whilst others report larger mean net benefits for t_0. Resources should be allocated to the therapy option that has the highest expected net benefits over all observations at the relevant value of λ. Using the available information a single resource allocation decision is made across all observations of net benefits. Assuming perfect information individual allocation decisions can be made for each observation of relative cost-effectiveness within the distributions of net benefits (derived from the decision model).

The individual observations of incremental net benefits are categorised as being positive for either t_1 or t_0. If t_1 is the more likely cost-effective intervention, in the proportion of observations for which t_1 is found to be cost-effective the resource allocation decision informed by the available data and by perfect information are the

same. In the proportion of observations in which t_0 is the cost-effective option, without perfect information an aggregate resource decision would allocate all resources to t_1, but assuming perfect information there is no uncertainty and t_0 would be funded. The expected cost of uncertainty (the EVPI) is the sum of the positive incremental net benefits *in the proportion of observations in which t_0 displays positive incremental net benefits* (see formula 3.5).

$$EVPI_{episode} = \left(\frac{\sum_{1}^{n} NB_{t0} - NB_{t1}}{n} \right) \left(\frac{n}{N} \right) \qquad (3.5)$$

Where $EVPI_{episode}$ is the expected EVPI per patient treated [Fenwick et al, 2000], n is the number of observations in which t_0 displays positive incremental net benefits, NB_{ti} are the mean net benefits of intervention i (i = 0 and 1), and N is the total number of observations. The $EVPI_{population}$ is estimated by multiplying the $EVPI_{episode}$ by the relevant patient population over the time for which the additional research is assumed to influence the allocation decision, discounted at an appropriate rate (see formula 3.6).

$$EVPI_{population} = EVPI_{episode} . \sum_{p=1}^{P} [I_p / (1+r)^p] \qquad (3.6)$$

Where p is a period, P is the number of periods for which the research is assumed to inform decision-making, I is the incidence in a period and r is the discount rate.

The estimation of the full EVPI for the decision model provides a maximum estimate of the value of removing all uncertainty within the model. Similar techniques can also be applied to estimate the cost of uncertainty in the value of individual input parameters within the model, or sets of input parameters. These estimates can be used to focus further research on the input parameters with the most impact on the overall uncertainty in the model [Fenwick et al, 2000]. To estimate the EVPI for a set of parameters, the full EVPI is estimated and then the EVPI is re-analysed, but holding the input values of the parameters of interest constant at their mean values. The

difference between the full EVPI and the re-estimated EVPI is the partial EVPI for the parameters of interest.

3.6.3.2 Estimating the expected value of sample information (EVSI)

The basic concept driving the estimation of the EVSI is that the original probability distributions for the input parameters can be updated, reducing the variation described by the prior distribution to reflect improved precision due to the collection of more data. The key assumptions in updating the probability distributions in particular, and the VoI analysis in general, relate to the choice of data used to update the prior distributions. This aspect of VoI analysis – the hypothetical representation of additional data – is a key element in ongoing research. The methods adopted in this paper assume that the additional data will yield the same mean values as derived from the original data, because the mean represents the best estimate of the true value of the individual parameters. Bayesian methods of statistical inference were employed to estimate updated distribution parameters, whereby the properties of conjugate families of prior distributions incorporated an increased sample size on posterior distributions [Berry and Stangl, 1996; Berger, 1980]. The estimation of the updated probability distributions is straightforward for all categories of input parameters, details of the methods for updating the input distributions are provided in Appendix 5.

The EVSI must be estimated a number of times to reflect the impact of successive increments in the sample size of prospective studies. If the assumption is made that a single trial would be undertaken that will collect data on all the parameters included in the decision model, it is important to note that this does not mean that the increased sample informing the updated distributions will be the same for all parameters. Fewer data will be available for parameters within the model. For example, a study with a sample of 100 patients will not provide 100 observations regarding, say, the cost of relapse if only half the patients experience a relapse within the trial. Likewise, if the 100 patients are randomised between two treatments, only 50 observations of the cost of providing the two treatments will be available. The proportion of patients likely to provide data on the different parameters within the model should be estimated using the mean results from the baseline analysis of the original decision model.

Next, the optimal sample allocation between the treatment arms should be addressed. Claxton proposed that the optimal sample allocation between the treatment arms in a proposed trial should maximise the ENBS on the basis of the loss function, which describes the value of information associated with sample allocations between the treatment options, as well as differences in the cost of sampling between the two interventions. This would appear to require a very large amount of time to 'estimate ENBS ... for every feasible allocation of each sample considered' [Claxton, 1999]. A simpler, and quicker, method of allocating a proposed sample between the treatment options being evaluated is derived from Neyman's allocation to strata, which accounts for both differential variances and costs of sampling between different strata [Cochrane, 1977]. A ratio of the sample allocation between two treatment options is estimated by supplanting treatment options for alternative strata. Formula 3.7 estimates the allocation ratio of the sample between treatments t_0 and t_1:

$$n_0 : n_1 = \frac{\sigma_0 / \sqrt{c_0}}{\sum \sigma_i / \sqrt{c_i}} : \frac{\sigma_1 / \sqrt{c_1}}{\sum \sigma_i / \sqrt{c_i}} \qquad (3.7)$$

Where n_i is the sample allocated to treatment option i, $i = 0$ and 1, c_i is the marginal cost of investigating a sampling unit of treatment option i, and σ_i is the standard deviation of the net benefits of treatment option i. Because the hypothetical data assumed to update the prior probability distributions are drawn from the patient population, the relevant measure of variation is described by first-order uncertainty, ie. variation between patients, rather than between populations. The standard deviation for each of the treatment options is estimated by running a first-order simulation using the mean estimates of each of the input probability distributions and calculating the standard deviation among the data observed for each patient within the run. This method estimates a constant ratio that can be applied to all prospective samples that are evaluated within the decision model. The optimal allocation of the prospective samples would account for the loss function within the decision model, which may produce alternative ratios for different sample sizes. The adaptation of Neyman's allocation provides a proxy for the optimal allocation, which reduces the required period of experimentation considerably.

Applying the updated input distributions the EVPI is re-estimated and the difference between the original EVPI and the revised EVPI is the EVSI for the each prospective sample. The EVSI can be estimated for the model as a whole or for sets of input parameters within the model. Estimating the EVSI for alternative sets of input parameters allows a comparison of the incremental net benefits associated with alternative primary studies set up to collect data on different sets of input parameters.

3.6.3.3 Estimating the expected net benefits of sampling (ENBS)

The ENBS is simply the EVSI minus the cost of obtaining the additional sampling information. The cost of sampling includes the cost of setting up the study, and monitoring, analysing and reporting the data collected. The cost also includes any additional treatment cost compared to current practice, which could feasibly be negative if current practice is the more costly option. Formula 3.8 is used to estimate the $ENBS_{population}$ for an additional sample of size n.

$$ENBS_{population} = EVSI_{population} - C_{Fixed} - nC_{Variable} - \frac{n_{T1}}{n}(C_{T1} - C_{T0}) \qquad (3.8)$$

Where C_{Fixed} are the fixed costs of a trial, $C_{Variable}$ is the marginal cost per patient included in a trial, and $C_{T1} - C_{T0}$ is the difference in costs between the interventions where $T1$ is the intervention that would not be provided in the absence of a trial.

The ENBS should be estimated for a range of sample sizes, which allows the ENBS to be plotted as a function of increased sample size. If any ENBS is positive then it will be efficient to gather more information and the optimal total sample size is where the ENBS reaches a maximum.

3.7 Conclusions

This aim of this Chapter was to describe the full process of an economic evaluation comparing health care interventions that amalgamates data from disparate sources within a decision model. The modelling process was covered in chronological order moving through five main stages: specifying the theoretical model, undertaking a

literature review to obtain input data from the model, analysis of the identified data to populate the model, implementing the model, and experimentation with the model. Previous work undertaken in the field of health economics was sought, though much of the process drew on issues common to the general area of clinical research. Insights into individual stages of the modelling process were also obtained from other disciplines including the social sciences and operations research. In addition, new approaches to handling some of the issues raised in the description of the modelling process were proposed.

As well as providing a general resource informing the full methodology for the application of decision models to economic HTA evaluations, this Chapter also provides a framework for the remainder of this thesis, wherein the methods described above are applied to a case study evaluation from Chapter 4 onwards. The issues to which this Chapter has contributed, and the location in this thesis of their application, are described in chronological order below.

The initial stage of the modelling process involves the specification of a theoretical model. This Chapter established the importance that such a model is developed prior to the literature review as it provides an explicit framework to guide the review. Advice from operations research texts was adapted to formalise the process of specifying a theoretical model, whereby the problem is structured in an attempt to understand the issues that need to be addressed within the model. The specification of a theoretical model representing possible pathways for patients with early breast cancer is described in first sections of Chapter 4.

The methods employed in the literature review were mainly derived from the systematic review literature. This Chapter stated that the process of identifying data to populate an economic HTA decision model comprises a series of individual literature reviews each informing a particular input parameter, or group of parameters. If the evaluative area is blessed with plentiful data, it may be necessary to limit the length of the literature review, mainly by the specification of search related criteria that limit the search with respect to the type, year, and language of available studies. The objective of the literature review was defined as 'the collection of data sufficiently robust to engender confidence in the model inputs', which respects the

need for an unbiased selection of data whilst recognising that is may not be possible to incorporate every piece of evidence into the model. The methods employed for a literature review to inform an economic HTA decision model of alternative therapies for early breast cancer are set out in the latter sections of Chapter 4.

The next stage of the modelling process involves the analysis of the data derived from the literature review, which was the greyest in terms of established methods. In the first phase of analysing the literature review data a reappraisal of the structure of the theoretical model was proposed. The reappraisal could incorporate adjustments to the model structure derived from an improved understanding of the important events, or subtle alterations to the relationships between existing parameters could be made that better facilitate the format of the identified data. The reappraisal of the early breast cancer model is described in Chapter 5. The next phase of the input data analysis introduced the explicit adjustment of data according to observed differences in their definitions of events to improve their homogeneity, which was labelled harmonisation. No reference to such methods was identified in the existing literature, though it is unlikely that such modifications have not been made implicitly. Whilst it is not feasible to expect a full account of the steps taken to harmonise data in papers presenting the results of economic HTA decision models, details should be made available from other sources to persuade the interested reader of the quality of the model. Various sections of the case study model required elements of harmonisation, the methods for which are described in Chapter 5.

The final issue in the analysis of the gathered data covers the methodology employed to organise the data into a suitable format to populate the decision model. Four alternatives were identified. Two methods required the creation of weighted datasets, which are either inputted directly or used to fit probability distributions using statistical fitting software. A full description of the available methods for weighting the identified data was provided, including the introduction of methods for the subjective weighting of data, when no objective measures of precision are available. The third approach specified particular distribution types to different groups of input parameters and used method of moment's formulae to estimate the distribution parameters using the identified data. The comparison of these three methods is a secondary objective of this thesis. The application of these methods is described in

Chapter 5, whilst the derived results are presented in Chapter 8 and discussed in Chapter 9. The fourth approach - bootstrapping distributions from the assembled data – is not applied to the case study because it is most appropriately used to specify distributions for parameters that are informed by patient-level data.

The implementation stage of the modelling process includes the act of choosing the appropriate modelling technique, because the modelling technique should be chosen with full knowledge of the available data. To guide the building of the decision model a collection of good practice guidelines drawn from the health economic, clinical and operations research literature were presented. The applied process of building both a Markov process and a DES model is described in Chapter 6.

As more complex decision models are employed in economic evaluation so the process of verification will become more important. This Chapter adapted the process of verification from operations research texts, which provided a thorough and structured process for checking that a model is internally consistent. The process of validating decision models also drew from the operations research field, but underlying differences in the modelling objectives and data availability between the OR field and economic HTA decision models limited the relevance of such work. The validation of economic HTA decision models is extremely difficult, though intermediate forms of validation are available, the main conclusion appears to be that the analyst should aim to convince her audience that the model was developed and analysed using an explicit framework of good practice. Nevertheless, Chapter 7 presents the attempts made to objectively validate the early breast cancer model using various sources of external data. Chapter 7 also presents full details of the application of the three-stage verification process.

In the final stage - experimentation - this Chapter addressed issues relating to the stochastic analysis of economic HTA decision models, describing the conduct of such analyses in terms of 'runs' and 'trials'. Relating to first- and second-order uncertainty, respectively, the questions of 'how many patients to include in a run?', and 'how many runs to include in a trial?' were discussed with reference to applied studies undertaken in all areas of clinical research. No specific solutions to these questions were identified, though some practical suggestions were put forward that are tested within

the case study and presented at the end of Chapter 6. A review of the chronological development of methods used to present cost-effectiveness data was presented, which informed the presentation of the case study results in Chapter 8. Finally, this Chapter adapted methods for the non-parametric analysis of the value of information, a process that estimates the monetary value of future research aimed at reducing the uncertainty of the resource allocation decision. The results of a full VoI analysis of the early breast cancer evaluation are presented in Chapter 8 and discussed in Chapter 9.

Chapter 4 Case study: Specifying the theoretical model and gathering data

4.1 Introduction

The methods for the modelling process described in Chapter 3 are applied to a case study evaluation in Chapters 4 to 8. The application aims to test the appropriateness of the methods, as well as to compare alternative methods where options within the modelling process exist. This Chapter presents the preliminary stages of the modelling process. These include the definition of the clinical area to be evaluated, which leads to the development of an initial model structure. The use of the theoretical model to guide the main literature review is then presented. The final section illustrates the selection of stratified patient groups for separate analysis within a model. This latter aspect is not necessary if the study question relates to a narrowly defined, homogeneous patient group, but when the patient group is heterogeneous such selection of sub-groups of the population may improve the analysis and enhance the interpretation of the results.

4.2 Defining relevant evaluation characteristics

Cancer is a major disease area that accounts for about one in every five deaths in England and Wales [Office of National Statistics, 2000]. Breast cancer is the commonest female cancer in the United Kingdom with around 33,000 newly diagnosed cases and 15,000 deaths from the disease each year [Cancer Research Campaign, 1998]. Diagnosis of cancer

in the breast and/or axilla is defined as early breast cancer, which may be further disaggregated to stage 0, I, or II breast cancer [Fowble, 1991]. A proportion of patients with early breast cancer will be 'cured' by local treatment. However, there is a risk of micrometastatic disease, which causes systemic relapse. The aim of adjuvant therapy is to destroy this subclinical disease. The main objective of adjuvant therapy for breast cancer is to prolong survival while maintaining a high quality of life [Glick, 1991]. As a treatment area, breast cancer fulfils several criteria that make it a suitable area for modelling: there are clear definitions of the disease and its natural history, definable benefits and risks with different modes of therapy, competing alternative approaches and controversy [Gelber et al, 1991]. In addition, new types and combinations of adjuvant therapies continue to be devised meaning that conclusive clinical trial data for all possible adjuvant therapy schedules and for all patient groups are unlikely ever to be available. Two types of adjuvant therapy are used for breast cancer, either alone or in combination:

1. Chemotherapy, which are usually administered as a combination of anticancer drugs, such as CMF (cyclophosphamide, methotrexate, and fluorouracil),
2. Hormonal therapy, which deprives cancer cells of oestrogen and may influence the growth of cancer cells. The two main hormonal treatments are tamoxifen and ovarian ablation (for premenopausal patients) [National Cancer Institute, 1996].

The evaluation presented in this thesis was initiated alongside the UKCCR ABC Trial. The ABC trial is a national collaborative randomised clinical trial, the principal aim of which is to determine the value of adding chemotherapy to tamoxifen for postmenopausal patients, and chemotherapy and/or ovarian ablation to tamoxifen for pre/perimenopausal women with early breast cancer [UKCCR, 1993]. Prior to the trial, chemotherapy and ovarian ablation had both been clearly shown to prolong both relapse-free and overall survival in pre/perimenopausal women, with the benefit lasting for at least 10 years. Tamoxifen had also produced a significant improvement in disease free survival in pre/perimenopausal women [Tormey et al, 1992]. However, it was uncertain whether the benefits of these treatments were additive [EBCTCG, 1992]. In postmenopausal women, trials had demonstrated a reduction in recurrence, and suggested an improvement in survival from

tamoxifen plus chemotherapy, compared to tamoxifen alone. However, it was not clear whether any additional survival benefit was large enough to outweigh the side effects of chemotherapy.

The ABC trial is currently in progress and the data from the trial will not be available for a number of years. The case study evaluation for this thesis was undertaken using secondary data with the primary aim of estimating the relative cost-effectiveness of the alternative adjuvant therapies to inform decision-makers prior to the availability of the trial results. The following section provides a brief summary of the previous economic studies comparing adjuvant therapies for early breast cancer, which were identified during the main literature review (see section 4.4.3 for details).

4.2.1 Review of economic studies of adjuvant therapies for early breast cancer

Only one identified economic study collected prospective cost data [Legorreta et al, 1996]. Based in the US, this 4-year longitudinal study followed 16 patients with stage 0-, 99 women with stage I, and 73 patients with stage II breast cancer in a health maintenance organisation. The results were split by stage at presentation. All cost data were obtained from medical records and claims data, though only inpatient and specialist outpatient costs were captured. The estimated 4-year mean costs for stages 0, I, and II were $18,900, $23,200, and $28,800, respectively (1993 US$). A UK-based study estimated 4-year treatment costs of 102 patients diagnosed with stage I breast cancer, and 13 patients with stage II breast cancer [Wolstenholme and Whynes, 1998]. Resource use was obtained from a detailed examination of the case notes and unit costs were attached using resource use algorithms. The mean costs were estimated as £3,576 and £3,996 for stages I and II, respectively (1991 UK sterling). A Canadian group modelled the treatment costs of breast cancer therapy for the lifetime of patients using clinical data from various databases and cancer registries, as well as surveys of Canadian clinicians [Will et al, 1998]. Detailed costing methods were employed including a 10-province comparison of the cost of the most commonly used breast cancer tests and surgical procedures. The treatment costs only included direct health costs and were split by stage at presentation. Average lifetime costs

for stage I and II breast cancer were estimated as $23,098 and $25,671, respectively (1995 Cdn$)

A number of full economic evaluations were identified, though none were undertaken in the UK. The principal economic investigators in the area of adjuvant therapies for early breast cancer appear to be a group of American clinicians - Hillner, Smith and Desch. These authors have published a range of economic evaluations covering most combinations of adjuvant therapies and most patient groups [Hillner and Smith, 1991; Hillner and Smith, 1992a; Hillner and Smith, 1992b; Hillner et al, 1993; Desch et al, 1993; Smith and Hillner, 1993]. All the evaluations were based on the same methodology, which employed a Markov process that was populated using clinical data derived from published clinical trials. Only direct costs were included and these were obtained from charges and from Medicare data, utility values were derived from surveys of oncology staff. The estimated cost per quality adjusted lifeyear (QALY) gained ranged from $5,700 comparing tamoxifen versus no treatment in a 45-year old premenopausal woman with node-positive, ER-positive breast cancer to $280,000 comparing tamoxifen versus no treatment in a 45-year old premenopausal woman with node-negative, ER-negative breast cancer (1998 US$) [Earle et al, 2000].

An Australian evaluation compared tamoxifen alone with no therapy using clinical data from the early breast cancer trialists' meta-analysis [EBCTCG, 1992], though they did not appear to differentiate between node negative and node positive patients [Glasziou and Haas, 1994]. The methodology was very basic using a spreadsheet and defining patients as either 'well' or in 'recurrence', though the calculations undertaken were not clear. The only costs included were the costs of tamoxifen and a general cost of recurrence taken from the literature. The model stopped at 10 years and utility values were assumed. The estimated cost per QALY gained was $1365 (1990/1 Aus$). Messori *et al* [1996] compared 12 cycles of CMF with no adjuvant therapy in patients with node positive breast cancer based on an Italian clinical trial with more than 20 years follow-up. The outcome measure was life years gained and a lot of effort was spent extrapolating the area under the curve to estimate long term survival. Unfortunately, less effort was put into estimating

costs as only the cost of adjuvant chemotherapy was included in the evaluation. The estimated cost per lifeyear gained was $447 (1995 US$). The final identified evaluation, which was only available as a conference abstract [Selke et al. 1998], employed a stochastic Markov process to evaluate a range of adjuvant therapies, though the definition of patient groups was not reported. The estimated cost per additional lifeyear saved from using chemotherapy and hormonal therapy over hormonal therapy alone was presented as $4,071 (US$).

The above review suggests that the evidence base for allocating resources to alternative adjuvant therapies in the UK is thin with only one UK-based study providing limited information that could be used to inform such decision-making. The development of a UK-based decision model describing the costs and health effects associated with alternative adjuvant therapies for different patient groups would provide useful information for decision-makers.

4.3 Specifying the theoretical model

The preliminary structure of the model represents the main events experienced by patients diagnosed with early breast cancer. The objective of the preliminary model structure was to provide guidance for the main literature review in the form of a series of study questions. Each study question invoked a separate literature review relating to different aspects of the decision model.

The events included within a decision model should describe significant occurrences in relation to the outcomes of the model. Here, the outcomes are costs and QALYs so the structure of model was based on events that have a significant influence on either the resources allocated to breast cancer patients, their survival, or their quality of life. Information on the relevant events to be included in the decision model was sought from health professionals involved with the ABC trial as well as a preparatory review of the breast cancer literature. In addition, the Internet was searched for basic information on the impact and progression of breast cancer, sites such as www.oncolink.org provided a wide

range of information, aimed at both health professionals and patients. The use of three sources reduced the danger of incorporating biases from any one source into the theoretical structure of the model, which was informed by general agreement between the alternative sources. The preliminary structure of the model is discussed in the following sections, which split the possible events into two main categories – treatment side effects and relapses or death. The two sections describe how the relevant events to be included in the model were chosen, following which the full structure is presented.

4.3.1 Side effects of adjuvant therapies

Following the administration of the adjuvant therapies the most immediate event that patients are likely to experience is some form of treatment side effect, especially following the administration of chemotherapy. The trial clinicians advised that the therapy side effects would be more easily classified in relation to their associated resource use, and that the severity of the side effects would be correlated with the resources used to treat them. A large number of different direct toxic effects were identified and each type of effect may be subject to a range of severities, it was also reported that a small proportion of patients die as a result of chemotherapy toxicity. The preliminary model described four classes of direct toxic effects:

1. mild complications that required additional clinician time and anti-emetics,
2. moderate complications needing additional clinician time and anti-emetics, as well as referral to a specialist, for example, a dietician,
3. non-fatal major events necessitating inpatient care, and
4. fatal events.

4.3.2 Relapses and death

The only other events included in the model were relapses and death. Following the removal of the primary cancer, a relapse is the appearance of a new lesion(s) in patients, as confirmed by any relevant diagnostic procedure [Thorpe et al. 1993]. There were various possibilities for modelling the pathway of breast cancer patients from the point of relapse. Hurley *et al* [1992] investigated the cost of breast cancer relapses and identified five

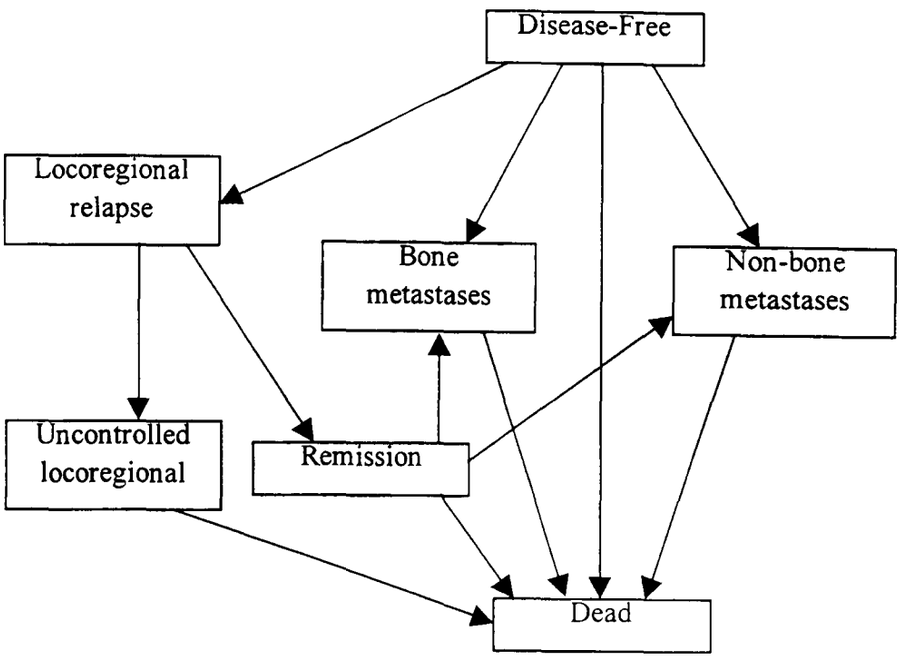
categories of sites. Two other studies also used a similar categorisation of relapses to describe the differences in prognosis relating to the site of metastases [Richards MA et al, 1993; Goldhirsch et al, 1988]: visceral, central nervous system (CNS), bone, local, and regional.

Presenting the information from the few studies identified to the trial clinicians, it was suggested that the three metastatic sites of relapse – visceral, CNS and bone – could be reduced to two – bone and non-bone – as relapses in visceral or CNS sites typically have the worst prognosis. Such relapses are not curable and the only event following a metastatic relapse is death. It was also suggested that local and regional relapses be treated as a single site within the model, but that the combined locoregional event could be either operable or inoperable. The distinction between operable and inoperable locoregional relapse was confirmed in the preparatory review. In the preliminary model a patient with an inoperable locoregional relapse entered an 'uncontrolled' state, in which the main objective of treatment is palliation. Following an operable relapse a patient entered a 'remission' state from which she could experience a further relapse or die without evidence of disease.

4.3.3 Model structure

An activity cycle diagram (ACD) was used to portray the preliminary model structure [Paul and Balmer, 1993]. The model structure is presented in Figure 4.1. The structure only includes events that are assumed to influence the progression of the disease, treatment side effects were described within the state 'disease-free'.

Figure 4.1 Preliminary model structure



Arrows indicate possible pathways between events.

4.4 The literature review

The aim of the review was to gather data to populate a decision model that described the pathways of patients with early breast cancer. Defining the structure of the model prepared the ground for targeting the literature review, five distinct data categories were defined:

- adjuvant therapies;
- treatment side effects;
- timing of relapse or death with no evidence of cancer;
- types of relapse;
- progression from relapse.

Adjuvant therapies comprised only the estimation of resource use and the unit costs associated with the various therapies, the four other categories included clinical parameters, resource use, unit costs, and utility values. The review protocol outlines the plan of the literature review, covering the study questions to be answered and the methods used to identify relevant data [NHS Centre for Reviews and Dissemination, 1996]. The following sections describe the formulated study questions, the search inclusion criteria and the

sources included in the literature search. The final sub-section presents the results of the literature review.

4.4.1 Study questions

The literature review for a modelling project comprises a series of mainly independent reviews that aim to identify data for the separate elements of the model. In the case study, separate questions were defined for each of the five data categories:

- What are the resource, and cost, requirements for the provision of the alternative adjuvant therapies?
- What proportion of patients, receiving the respective adjuvant therapies, experience the defined categories of side effects?
- What are the resource, cost and utility implications of the defined categories of side effects?
- What disease-free interval (DFI) profiles are associated with alternative adjuvant therapies?
- What are the resource, cost and utility implications of remaining disease-free?
- What proportions of patients experience locoregional relapse or metastases following DFI?
- What proportions of patients die without relapsing following DFI?
- How do patients progress from the point of relapse, either locoregional or metastatic?
- What are the resource, cost and utility implications of the defined categories of relapse?

4.4.2 Search-related inclusion criteria

The study questions did not contain any description of patient characteristics because it was intended to collect data on all patients with breast cancer. However, detailed data on the characteristics of patients included in each relevant study were collected, which facilitated the definition of more homogeneous sub-groups within the aggregate patient population (see section 4.5).

With respect to other inclusion criteria, the preliminary search used to inform the structure of the preliminary model had revealed a wealth of clinical research on most aspects of breast cancer. It was also noted that the majority of comparative studies were randomised clinical trials (RCTs). Due to resource constraints on the amount of time that could be spent appraising possible data sources, the initial literature review was limited to papers published from 1992 onwards. Another justification for limiting the review was that most clinical trials do not simply report once, rather they report at regular intervals as the length of follow-up increases and more definite judgements about the results of the trial can be made. Also, chemotherapeutic agents have evolved over time, so that some of the drugs administered in trials that reported in the 1980's (meaning their protocols may have been specified in the 1970's) may not be considered relevant today. The reduced relevance of chemotherapy regimens used in past trials may relate to the actual drugs applied, the doses, or the frequency of administration.

The identified RCTs provided a large amount of data covering the clinical parameters, so there was no reason to collect data from other primary studies of lower methodological quality. Therefore, searches relating to the clinical parameters were restricted to RCTs and reviews. Reviews were included for two reasons. Firstly, reviews present an overview of the general results within a clinical area so providing a source of primary studies. Secondly, reviews typically discuss the meaning of observed results in more detail than papers presenting primary results. It was hoped that such qualitative data would improve the understanding of the mechanisms of the disease, which could be useful during the analysis of the data or any revision of the model. Non-English written studies were excluded from the review on the basis that sufficient data to inform the required parameters could be accessed in the English language literature so the costs of translating foreign language articles could not be justified in terms of the additional data collected. Also, bias against negative studies by the main English language journals is less likely in a prominent research area such as breast cancer.

It was recognised that limiting the literature review to mainly RCTs published from 1992 onwards might identify insufficient data for some parameters. If necessary further searches

for non-RCT based studies would be made, and the search parameters could be widened to include the pre-1992 literature.

4.4.3 The literature search

The aim of the literature search was to identify a comprehensive selection of RCTs and reviews to address the study questions stated above. The search comprised both computerised and manual searches. A number of electronic databases were searched, including Medline, the Science Citation Index, the NHS database of abstracts of reviews of effectiveness, the Cochrane library, the NHS economic evaluation database and the Office of Health Economics economic evaluation database. Indexed terms were used wherever possible to search the different databases, though the use of the different databases was tailored according to the different types of parameters. For example, it was unlikely that new clinical data would be included in the economic evaluation databases so broad searches for any economic studies relating to breast cancer were undertaken. The broader medical databases contained studies reporting resource use, costs, utilities, as well as clinical parameter estimates, so more well-defined searches were used that informed specific elements of the model. Manual searches of the reference lists reported in studies identified from the databases and of selected journals were also undertaken. Full details of the literature search are provided in Appendix 1. The sources included in the literature search have been justified previously in the collection of data for decision analysis models [Murphy et al, 1994].

The titles of all the studies identified using the specified search terms were read and the abstracts downloaded if they appeared to have any connection to the parameters required for the decision model. Relevance was defined both quantitatively and qualitatively. Papers were accepted if they appeared to provide quantitative data that could be combined with data from other papers to inform parameter values within the model. Alternatively, qualitative data were sought to increase understanding of the treatment area, which might provide insights into the progress of disease and other input parameters.

Following the literature search, the data extraction process involved the collection of all the necessary data from the identified studies. It was important that the method of extraction created a database that could be searched, allowing papers holding data on similar parameters to be identified. The papers within the database were labelled with respect to the type of data reported. Furthermore, it was anticipated that the analysis of the collected data would require extensive use of spreadsheets so the easy transfer of required data into a spreadsheet package was essential. The creation of an evidential database, using an established reference database software package [Procite 5, 1999], is described in Appendix 2.

4.4.4 Literature search results

The initial Medline search resulted in the collection of 636 abstracts that were assessed for inclusion in the review. From the 636 abstracts, 230 full papers were defined as potentially relevant to the analysis. These papers were collected and read in full. Including studies from all sources, by the time of the final analysis a total of 343 full papers or documents had been reviewed. The basic results of the data extraction process are presented alongside the parameter categories in Figure 4.2 – the number of papers presenting data on the respective parameters. These results were gathered in March 2000, meaning that papers published after this date were not incorporated into the analysis. The most populous category contained clinical data on 'progression from relapse', which reflects the large number of clinical trials that have compared alternative therapies for patients with metastatic breast cancer. The next most frequently observed parameters covered data relating to treatment toxicities, DFI and overall survival, as clinical trials of adjuvant therapies tended to present data on all three issues.

Figure 4.2 Pro forma used to extract data from the literature review

Workform: ABC parameter values

Author (01):
Title (04):
Journal (09):
Date of publication (12):
Aims (13):
Location of study (14):
Time period (15):
Study design (17):
Study population (18):
Variable definition (19):
Omissions in context (22):
Comments (23):
Assessors (24):

Workform: ABC parameter values

idToxicities (25): 58
idMenopausal symptoms (26): 15
idDisease free interval (27): 63
idRates of relapse (28): 24
idTypes of recurrence (29): 73
idProgression from recurrence (31): 99
idOverall survival (33): 70
coChemotherapy (34): 21
coOvarian ablation (35): 2
coToxicities (36): 9
coMenopausal symptoms (37): 6
coTreatment of recurrence (38): 57
quToxicities (39): 10
quMenopausal symptoms (40): 3
quDisease free interval (41): 6
quRecurrence (42): 37
quGeneral (43): 11
coOverall Cost (44): 2
Parameter Keywords (45):

In the cost and resource use categories information on the 'treatment of recurrence' was most commonly identified. The data extracted for this set of parameters mostly described the range of resources associated with the different areas of treatment for metastases, i.e. therapies, surveillance and hospital visits. Fewer papers reported such data relating to chemotherapy as an adjuvant therapy for breast cancer, though more actual cost data were identified. Most quality of life data was again captured for the recurrence parameters, though this data mainly described toxicity associated with the chemotherapy regimens administered to patients with metastases.

4.5 Definition of patient sub-groups

An objective of the ABC decision model was to analyse sub-groups of the aggregate population of patients with early breast cancer, which allows the investigator to ask whether the response difference between two treatments depends on the type of patient [Pocock,

1983]. In addition, the choice of comparators within an evaluation should reflect the most relevant treatment options to the policy question [Drummond et al, 1997], which do vary according to the age and prognosis of breast cancer patients [Silva and Zurrida, 1999; Glick, 1991]. The following two sections describe the methods used to define relevant sub-groups and the resulting definition of sub-groups.

4.5.1 Methods

Subgroups can be defined with respect to either prognostic indicators or predictor markers, or both. Biological prognostic indicators are used to inform a patient's survival and disease-free survival, whilst predictor markers forecast tumour sensitivity or resistance to various therapies. The most important prognostic indicator for breast cancer is the nodal status of a patient, which is either positive or negative. In addition, widely investigated breast cancer biological markers include estrogen and progesterone receptors, p53, Bcl-2, c-erbB-2, cyclin expression, proliferative activity, DNA ploidy and the urokinase plasminogen activation system [Ravaioli et al, 1998]. Estrogen receptors have been found to be weak prognostic indicators, but good predictors of response to endocrine therapy. There are consistent data suggesting that proliferation indices are good indicators of prognosis, and that they are directly related to response to chemotherapy and closely related to response to hormonotherapy. Otherwise, there is no evidence, or conflicting data for all of the other biological markers [Ravaioli et al, 1998]. In relation to menopausal status, postmenopausal women have been found to derive less overall benefit from chemotherapy, which may be related to the biology of breast cancer in older patients [Fox, 1991].

Another factor in the definition of patient sub-groups was the need to identify sufficient data to inform the parameter values within the model for each of the relevant adjuvant therapies for the specified sub-groups. All studies reporting relevant data were carefully read to identify any prognostic information reported on the patient groups included in the study.

4.5.2 Results

The full details of the information collected on patient characteristics are presented in Appendix 6. Eligibility criteria for the identified studies varied widely, ranging from open studies that might only specify that patients must be node positive [Schumacher et al, 1994; Fisher et al, 1997], to strictly defined patient groups that controlled for nodal, receptor and menopausal status [Pritchard et al, 1997; Cummings et al, 1993]. Some of the open studies, however, presented relevant data for sub-groups of the study population [Chacon et al, 1997; Wood et al, 1994].

Data on the following indicators were recorded - nodal status, age, menopausal status, number of positive nodes and estrogen and progesterone receptor (ER) status. Unfortunately, very few studies reported proliferative activity, so this indicator could not be used to differentiate patient groups. Due to the aggregated level of reporting only relatively broad patient groups could be defined. Nodal status (primarily node positive) was the most commonly defined prognostic indicator, though menopausal status was often distinguished. Fortunately, a number of studies combined these two indicators meaning that sufficient data were available to estimate parameters for two main patient groups – node positive and postmenopausal, and node positive and premenopausal.

Fewer studies evaluated alternative adjuvant therapies for node negative patients. Therefore, less strict inclusion criteria were drawn up, whereby study results were included in sub-groups on the basis of the proportion of patients with the relevant characteristics. For example, patient groups were defined on the basis that over 80% of the patients were node negative. Some studies reported information on age but not menopausal status, and vice versa [Gelber et al, 1993; Fisher et al, 1996]. These characteristics are highly correlated, so to make the best use of the available data patient age and menopausal status were used in combination to define patient groups.

In total, 29 separate patient categorisations were identified. Tables 4.1 and 4.2 present the number of studies satisfying the chosen patient characteristics for each defined sub-group. The final column - 'minimum number of studies' - describes the least number of studies

included in any of the three treatment categories for DFI and events ending DFI. The minimum number of studies is highest for groups including older node positive patients. Younger node positive patients were adequately represented in trials of chemotherapy and/or tamoxifen+chemotherapy, but few such patients received tamoxifen alone. Even less choice was available for the analysis of node negative patients. A minimum of two trial treatment arms reported DFI for the patient categorisation '80% node negative, and median age over 50 years or 50% postmenopausal'. The final configuration of analyses is presented in Table 4.3.

4.6 Conclusions

This Chapter has applied methods informing the first stages of the modelling process to an economic HTA evaluation comparing alternative adjuvant therapies for early breast cancer. The stages covered included the specification of a theoretical model structure, the literature review to identify data to further inform the model structure and to populate the model, and the process of identifying relevant patient sub-groups.

To create an initial model structure information was sought from three distinct sources; consultations with oncology clinicians, a preparatory review of the literature, and the Internet. The use of these three sources was deemed necessary to ensure that a balanced view of the relevant events was obtained. The theoretical model highlighted five distinct categories within the model, which were used to inform the subsequent literature review. Separate literature reviews were undertaken for a range of study questions that were derived from the five parameter categories: adjuvant therapies; treatment side effects; timing of relapse or death; types of relapse; and progression from relapse. Strict inclusion criteria were defined due to the anticipated volume of data, though it was recognised that the review might be expanded if insufficient data were identified. The subsequent data analysis revealed areas where the review needed to be expanded, this is reported in Chapter 5. The explicit formulation of search-related inclusion criteria is important to avoid accusations of bias in the data collection process.

The definition of relevant patient sub-groups was necessary because data were collected on the whole population of patients with early breast cancer, which incorporates a wide range of prognoses. Possible foundations for the grouping of patients were discussed, and reasons provided for the final sub-classifications. The precision in the definition of sub-groups should be balanced against the availability of data, as demonstrated by the sub-group definitions for node-positive and node-negative patients.

The review of the modelling process presented in Chapter 3 could not provide a prescriptive methodology for all such evaluations as elements of the process are influenced by the characteristics of the treatment area being evaluated. This Chapter has illustrated the application of the first stages of the methodology and how choices specific to the treatment area can be made.

Table 4.1 Number of studies satisfying nineteen alternative categorisations for node positive patients, by adjuvant therapies

Patient categories	Disease free interval			Clinical events ending DFI			Minimum number of studies
	Tamoxifen	Chemotherapy	T+C	Tamoxifen	Chemotherapy	T+C	
100% node positive	7	64	11	6	38	10	6
80% node positive	7	66	11	7	41	10	7
100% node positive, and 100% postmenopausal	6	16	5	5	1	3	1
100% node positive, and 100% premenopausal	0	30	4	0	10	4	0
100% node positive, and median age over 50 years or 50% postmenopausal	7	32	6	6	17	5	5
100% node positive, and median age over 55 years or 55% postmenopausal	7	25	6	6	10	5	5
80% node positive, 60% postmenopausal	6	17	5	6	6	3	3
100% node positive, and median age under 50 years or 50% premenopausal	0	40	5	0	25	5	0
80% node positive, and median age under 50 years or 50% premenopausal	0	40	5	0	25	5	0
80% node positive, and median age over 50 years or 50% postmenopausal	7	32	6	7	20	5	5
100% node positive, 70% postmenopausal, and 50% 1-3+nodes	4	10	3	4	1	3	1
100% node positive, 50% 1-3 positive nodes, and median age under 50 years or 50% premenopausal	0	30	5	0	20	5	0
100% node positive, 50% 1-3 positive nodes, and median age over 55	5	14	4	5	5	5	4
100% node positive, 60% ER positive, and 60% postmenopausal	3	9	3	3	1	3	1
80% node positive, 60% ER positive, and median age under 50 years or 50% premenopausal	0	19	5	0	12	4	0
80% node positive, 90% ER positive, and 70% postmenopausal	3	1	2	3	1	2	1
80% node positive, 60% ER positive, and 50% 1-3 positive nodes	3	22	8	3	11	8	3
80% node positive, 60% ER positive, 50% 1-3 positive nodes, and median age under 50 years or 50% premenopausal	0	14	5	0	10	4	0
80% node positive, 60% ER positive, 50% 1-3 positive nodes, and median age over 50 years or 50% postmenopausal	3	9	3	3	2	4	2

Table 4.2 Number of studies satisfying nine alternative categorisations for node negative patients, by adjuvant therapies

Patient categories	Disease free interval			Clinical events ending DFI			Minimum number of studies
	Tamoxifen	Chemotherapy	T+C	Tamoxifen	Chemotherapy	T+C	
100% node negative	7	5	2	5	3	2	2
80% node negative	7	5	2	5	3	3	2
100% node negative, and 100% postmenopausal	2	0	0	2	0	0	0
100% node negative, and 100% premenopausal	1	0	0	1	0	0	0
100% node negative, and median age under 50 years or 50% premenopausal	2	3	0	1	3	0	0
80% node negative, and median age over 50 years or 50% postmenopausal	5	2	2	4	0	3	0
80% node negative, and median age under 50 years or 50% premenopausal	2	3	0	1	3	0	0
80% node negative, and 70% postmenopausal	3	0	0	3	0	0	0
80% node negative, and 60% tumour grade 1	2	0	2	2	0	2	0
80% node negative, and 45% age under 50	1	2	2	0	2	2	0

Table 4.3 Patient categorisations used for the analysis of the four generally defined patient groups

General patient group	Patient categorisation	Adjuvant therapies analysed
Younger (premenopausal) node positive	100% node positive, and 100% premenopausal	Chemotherapy and tamoxifen+chemotherapy
Older (postmenopausal) node positive	'100% node positive, and 100% postmenopausal' and '100% node positive, median age over 55 years'	Tamoxifen, chemotherapy, and tamoxifen+chemotherapy
Younger (premenopausal) node negative	100% node negative, and median age under 50 years or 50% premenopausal (80% node negative, and 45% age under 50, for tamoxifen+chemotherapy)	Tamoxifen, chemotherapy, and tamoxifen+chemotherapy
Older (postmenopausal) node negative	'80% node negative, and median age over 50 years or 50% postmenopausal'	Tamoxifen, chemotherapy, and tamoxifen+chemotherapy

Chapter 5 Case study: analysing the literature review data

5.1 Introduction

At this stage in the modelling process, data to populate the decision model have been extracted from the literature, which must now be assembled into probability distributions to represent the uncertainty about the true values of the model's input parameters. The first task addressed, however, is the reappraisal of the preliminary model structure, which may be altered due to an improved understanding of the treatment area, or to better facilitate the format of the identified data. The following sections present the harmonisation of the identified data, and the quantitative analysis of the data to specify representative probability distributions.

The harmonisation of the data makes explicit assumptions about data presented by different studies with respect to the definition of relevant events, and attempts to revise the results of outlying studies to a common definition. As discussed in Chapter 3, there are various approaches to the definition of input distributions. The applied options involved the creation of weighted datasets (that were inputted directly into the model, or used to fit probability distributions using statistical fitting software) and the specification of theoretical distribution types for different categories of parameters. The bootstrapping option was not applied to the case study evaluation because it is more suited to decision modelling studies based on patient-level data.

The application of the modelling process to the full range of sub-groups defined within the aggregate early breast cancer population is not feasible due to space restrictions. From this point onwards, therefore, the presentation of the case study evaluation is limited to a comparison of chemotherapy and tamoxifen versus tamoxifen alone in node-positive, postmenopausal women with early breast cancer. In addition, this Chapter concentrates on the methodological aspects of analysing the literature review data, presenting the reappraisal of the theoretical model, the harmonisation, and the definition of the probability distributions in three separate sections. To avoid overburdening the reader, details of the data and the presentation of the resulting distributions have not been included in this Chapter. Full details are provided in Appendix 7.

5.2 Reappraisal of the theoretical model

The process of reappraising the preliminary model structure is presented in two sections, describing the representation of treatment side effects and the depiction of events associated with the experience of relapsing and dying. The amended model structure is presented in Figure 5.1.

5.2.1 Reappraisal of the treatment side effects categories

The range of side effects associated with chemotherapy is huge, and the categorisation of the toxicities into a limited number of groups was a particularly difficult task. The theoretical model had specified four categories based on increasing orders of severity, with the most severe category representing fatalities. At the point of reappraisal, it became clear that the inclusion of a separate fatality category would cause problems for the estimation of the length of disease free interval (DFI) because the fatalities due to treatment side effects were included in the aggregate DFI survival curves. The number of toxicity categories was reduced to three. In addition, the basis for the category definitions was altered to reflect the presentation of toxicity data in the published literature. The theoretical model had specified alternative categories on the basis of differential resource use, whilst the literature tended to define side effects as major toxic events (primarily cerebrovascular events) and graded

toxicities, differentiating between grades 1 or 2, and grades 3 or 4. The identified data also specified that patients could experience multiple categories of side effects simultaneously.

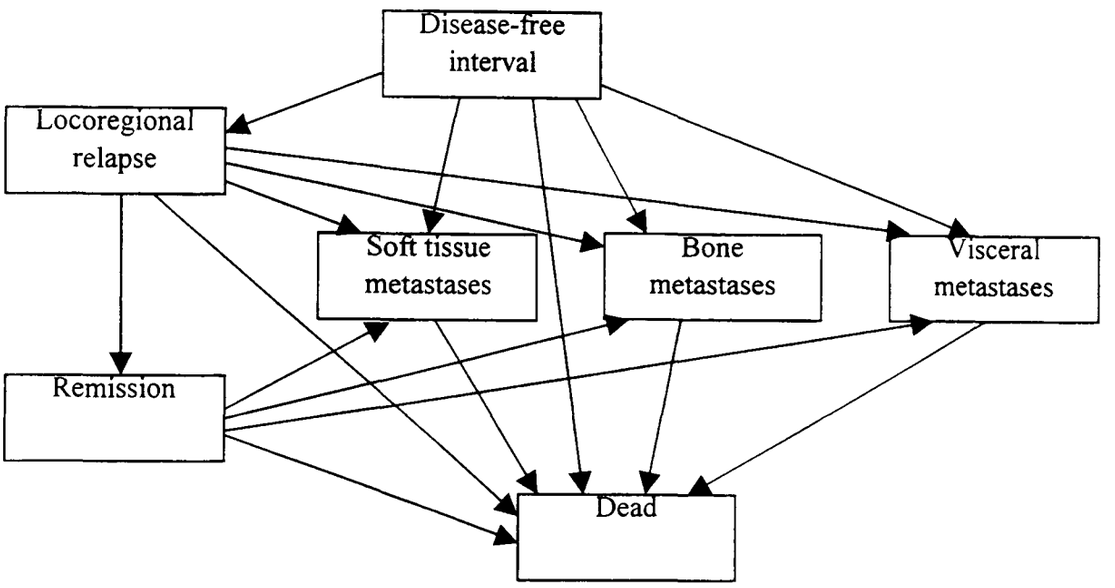
5.2.2 Reappraisal of the relapse categories

The most radical change to the model structure concerned the representation of locoregional relapses. The theoretical model had defined a locoregional relapse as operable or inoperable. Unfortunately, few data were identified describing the respective pathways of these two patient groups. An alternative structure was informed by the format of the data presented in the literature. Most studies reporting on the prognosis of patients experiencing a locoregional relapse presented survival curves illustrating the proportion of patients remaining in remission from the point of diagnosis with locoregional relapse. The revised model structure assumed, therefore, that all locoregional relapses were controlled, and all patients entered the health state 'remission', even if the period of control was instantaneous. The events following a locoregional relapse could then be described in the same way as the events following DFI, though no allowance was made for the experience of a second locoregional relapse. From the point of dissemination, the progression scenario switched to the relevant site of distant relapse. For the proportion of patients who do not progress to metastases, it was reported that survival did not differ from that of the general population [Willner et al, 1997].

It has been shown that progress from distant relapse is related to the site of metastases [Clark et al, 1987]. The most common distinction made in the literature separated visceral, bone and soft tissue sites. The model was again restructured to represent three sites of metastases, rather than the original distinction between bone and non-bone sites. It was also noted from the literature review that patients could experience metastases at two or more sites simultaneously, though additional states describing double sites of relapse were not included as few data described progression from multiple sites of relapse. In addition to survival data, events such as remission, stability and progression of the disease were investigated as well as the use of second- and third-line therapies, though few data other than survival data were identified. Due to the lack of data and the relatively short survival

time from metastases, the original representation of patients progressing only to death from diagnosis with metastases was maintained.

Figure 5.1 Reappraised structure of the ABC model



5.3 Harmonising the data

The process of harmonising the data was only required for some of the clinical parameters included in the model. The following sections illustrate some of the available methods that can be used to increase the comparability of identified data describing treatment side effects, DFI and events ending DFI, and progression from the point of relapse.

5.3.1 Treatment side effects

The decision model only describes the effect of toxicity in the DFI state because side effects associated with the treatment of a relapse were incorporated within a general description of those states. The analysis of treatment side effects did not distinguish between different patient groups, rather data were sought on the toxicity caused by alternative combinations of chemotherapeutic agents, with tamoxifen, as well as by tamoxifen alone. The amalgamation of all patient groups for the analysis of toxicity increased the available data substantially.

The analysis of the treatment side effects involved three main issues. Firstly, the aggregate proportions of patients experiencing each class of toxicity were required. Secondly, for the purpose of costing information was needed on the types of conditions, and the frequency of their occurrence, within each of the toxicity categories. Finally, compliance with the adjuvant therapies is connected to the experience of toxicity. The areas in which action was required to harmonise these data are described below.

In the estimation of aggregate incidence of the graded toxicity categories the majority of studies used the WHO classification index, although a couple of studies referred to less well-known indices, for example, the South West Oncology Group grading [Budd et al, 1995; Rivkin et al, 1996], or the ECOG index [Tormey et al, 1992]. No major differences between the grading systems were identified and they were analysed simultaneously with no adjustments. Aggregate data were reported for grade 3 or 4 toxicity, in particular, enabling an accurate estimate of the proportion of patients experiencing these grades of toxicity. Unfortunately, the grade 1 or 2 toxicity category mostly reported the proportions of patients experiencing the constituent conditions within the aggregate category. Although tending towards an underestimate, the proportions of patients experiencing the most common condition within the relevant category were taken as minimum estimates of the aggregate proportion.

Due to the number of conditions that were reported in the published literature the analysis of the type and frequency of events was difficult. In studies where specific conditions were not reported, that condition was assumed not to have occurred, i.e. to have a zero proportion. To prevent distortion due to limited reporting of toxicity conditions, studies reporting less than four toxicity conditions were excluded from the composition analysis.

Ten studies were identified that provided quantitative data on patient compliance with chemotherapy. However, four broad methods for presenting such data were established. Four studies presented the proportion of patients completing the specified cycles of chemotherapy [Scottish cancer trials breast group, 1993; Velez-Garcia et al, 1992; Zambetti et al, 1992; International breast cancer study group, 1996], and two studies each used the

following methods: the proportion completing cumulative cycles [Schumacher et al, 1994; Coombes et al, 1996], the proportion of individual components of the chemotherapy regimen completed [Misset et al, 1996; Marini et al, 1996], and the cumulative percentages of protocol dose completed [Fisher et al, 1996; Marini et al, 1996]. The clearest data covered the proportion of patients completing cumulative cycles of chemotherapy and this data was included in the analysis.

5.3.2 Disease free interval (DFI) and clinical events ending disease free interval

DFI describes the period following primary treatment of early breast cancer until the experience of a relapse or death with no evidence of disease (DNED). The length of DFI and the events ending DFI are possibly the most important in terms of the relative effectiveness of alternative therapies, but they are assessed jointly because of their natural connection. Indeed, the majority of studies used to inform DFI also provided data on the clinical events, and the issues around the harmonisation of both categories are similar.

Two discrepancies were found in the inclusion criteria for DFI. Firstly, most studies included DNED as ending DFI, though a few studies analysed such deaths as censored events [Eckman et al, 1998; Velez-Garcia et al, 1992; Gundersen et al, 1995]. If DNED data were not reported for a particular study the proportion of cases of DNED from a study with similar patient characteristics and adjuvant therapies was used as a proxy, which was added to the annual proportions of patients ending DFI.

Disparity was also noted in the reporting of second primary tumours. All studies providing data on the length of DFI were reviewed to ascertain whether second primaries had been included. Forty-seven studies provided useful data on the annual proportions ending DFI, of which 25 definitely did not include second primary tumours as an endpoint for DFI, and 12 definitely did include second primaries. Of the remaining 10 studies, six appeared likely to exclude second primaries, two appeared likely to include such events, and two gave no indication. The early breast cancer trialists excluded second primaries from the analysis of DFI. Taking this approach as a lead, second primaries were excluded from the annual

proportions of patients ending DFI in all the inclusive studies. Thus, second primaries, as a proportion of events ending DFI, were subtracted from the annual proportions ending DFI in studies that included such events.

Ipsilateral breast tumour relapses were also excluded because they can only occur after breast conserving surgery, and their prognosis is considerably better than other local relapses following mastectomy [Harrold et al, 1998]. Thus, in studies where breast conservation was an eligible option, the proportion of relapses that were in the ipsilateral breast was subtracted from the annual proportions of patients ending DFI. In studies not reporting the proportion of such relapses, the proportion reported in similar studies, with respect to treatment and the proportion of patients receiving breast conserving surgery, was substituted. Some studies included patients receiving breast conserving surgery, but did not state the proportion. In such studies it was assumed that five per cent of relapses were in the ipsilateral breast - an intermediate figure derived from all studies reporting ipsilateral relapses.

As the model did not describe the experience of multiple sites of metastases, harmonisation was also required for studies that reported the proportion of patients with different sites of metastases, but which included patients experiencing multiple sites of relapse. Visceral metastases was documented as being the most severe form of metastases, so the reported proportions of visceral metastases were not amended. However, the proportion of bone and soft tissue relapses were re-estimated using formulae 5.1 and 5.2, respectively:

$$pr(bone) = [bone/(bone + softtissue)] \times [1 - visceral] \qquad (5.1)$$

$$pr(softtissue) = [softtissue/(bone + softtissue)] \times [1 - visceral] \qquad (5.2)$$

5.3.3 Progress from relapse

Patient pathways following the diagnosis of a relapse were analysed separately for locoregional relapses and metastases. Within the locoregional category, the only divergence between the identified studies related to the assumption made in the models that

patients could not experience a second locoregional relapse. Such events are rare, but it was still necessary to ensure that the data reported by the identified studies described the time to the experience of metastases or death. The only data required from the point of diagnosis with metastases were survival times and no areas of harmonisation were identified for these data.

5.4 Pooling and formatting the data

Quantitative analysis was required to pool and format the identified data for all the parameters included in the model. The following sections describe the issues arising in the analysis undertaken for the different types of clinical parameters, resource use and cost parameters, and utility values.

5.4.1 Treatment side effects

Probability distributions were described for the proportion of patients experiencing each of the three categories of toxicity. Distributions were not assembled for the data on the type and frequency of events because such detail would increase the complexity of the model programming, but also because specifying distributions around the cost estimates could represent such uncertainty. The methods applied to create the alternative forms of probability distributions are described below.

The first two methods - empirical distributions and fitted distributions - are based on the creation of weighted datasets. As all the identified data on toxicity were collected from clinical trials meta-analytic methods for weighting the data could be employed. The analysis was based on the fixed effects model as strict definitions of toxicity had been specified, so only limited heterogeneity between the identified studies was assumed. The data from each identified study was weighted using formula 5.3.

$$\frac{p_i(1-p_i)}{n_i} \qquad (5.3)$$

Where p_i is the proportion experiencing a particular category of toxicity and n_i is the sample included in the study (see Appendix 3). After weighting the individual observations, a weighted dataset of 100 observations was created by dividing each weight by the sum of the weights and multiplying by 100 (value = x_i). The dataset comprised x_i copies of each parameter observation i (see section 3.4.3.2). The resulting datasets were inputted directly into the models as one form of representing the uncertainty in the values of the toxicity parameters, and they were also used to fit probability distributions. The fitted distributions were defined by inputting the datasets into statistical fitting software [Stat::Fit, 1996]. The software calculated goodness-of-fit statistics for each defined probability distribution so the best-fitting distribution could be selected.

The theoretically defined probability distribution assigned for proportion parameters is the beta distribution. Estimating the beta distribution parameters (α and β) did not require the creation of a weighted dataset as α and β are related to the number of events experienced (x) and the total number of patients (n) [Iverson, 1984]: $\alpha = x$ and $\beta = n - x$ (see Appendix 5).

5.4.2 Disease free interval and clinical events ending disease free interval

The DFI data were collected from survival curves presented as the results of clinical trials. Though a couple of authors were contacted for assistance in deciphering the data presented in unclear figures, or figures presented on a log scale, the majority of data were read directly off the printed page. The data initially assembled were the annual proportions of patients in the original cohort (in year 0) leaving DFI, which required transforming to the proportion of patients remaining event free at the beginning of a particular year experiencing an event during that year. Formula 5.4 was used to transform the data (see Appendix 4). Separate datasets describing the proportions of patients leaving the state DFI were created for each year following primary treatment. The weighted datasets and the theoretical distributions were specified using the same methods as for the treatment side effects data described in the previous section.

$$P[event/remainingcohort]_i = \frac{P[event/originalcohort]_i}{P[remaining/originalcohort]_i} \quad (5.4)$$

Some of the data describing the types of events ending DFI were collected from retrospective observational studies [Goldhirsch et al, 1994; Kamby et al, 1987]. No attempt was made to weight the evidence according to the type of study and such data were analysed on the same level as the clinical trial data. Both sources provided the respective sample sizes so the weighted datasets and the beta distributions were estimated as described above.

5.4.3 Progression from relapse

The quantitative analyses of the data describing progression from the two main forms of relapse - locoregional and metastases - are presented in this section. The majority of the identified data describing the time in remission following a locoregional relapse were in the form of survival curves. The DES model recorded the time at which individual patients entered the locoregional state and the separate probabilities of leaving the state in subsequent time periods could be applied individually to each patient. Thus, the data were analysed in a similar manner to the data for the length of DFI, as described above. The Markovian assumption precluded the use of differential annual probabilities of remaining in the remission state. Time in remission was described as a constant probability of experiencing a further relapse.

The definition of the input distributions describing the length of remission in a DES model was identical to the definition of the input distributions for the length of DFI. Weighted datasets were created for each year following the treatment of a locoregional relapse, which were used directly within the model, as well fitted to probability distributions using statistical fitting software. Alternatively, using theoretically defined beta distributions the numbers of patients experiencing, and not experiencing, a relapse in each year informed the relevant distribution parameters. Only a single input distribution was required to describe the remission data in the Markov model. In the initial quantitative analysis the reported

median times in remission were weighted according to the sample included in the respective studies and a weighted dataset created. This data was inputted directly into the model, as well as to fit probability distributions. The dataset was also used to estimate a mean and a standard deviation for a theoretically defined gamma distribution of the median time spent in remission. During the process of validation, however, it became apparent that the patient-level distribution of time in remission was heavily skewed and median estimates were underestimating the mean length of remission (see section 7.3). Using the presented survival curves representing time in remission, the mean remission periods were estimated on the basis of a common cut-off length of remission for those patients remaining in remission at the end of each studies follow-up (see section 9.2 for a full discussion).

Very few studies presented survival from metastases in the form of survival curves. As the survival period is relatively short most studies simply reported the median survival (usually in months). The use of median times, rather than annual proportions or rates, simplified the analysis. Three options for the analysis of the data on the progression from each of the three separate sites of metastases were possible, which would use differing amounts of the available data:

- create weighted datasets using studies presenting separate data for individual metastases categories;
- create weighted datasets using studies in which a single metastatic site is dominant in over 50% of patients;
- regress survival on three explanatory variables, defined as the proportions of patients with soft tissue, bone, and visceral metastases in each identified study (observation).

The available data were mostly taken from clinical trials of alternative therapies for metastatic cancer, of which there was a reasonable amount. 113 separate treatment arms were identified, 68 of which presented an aggregate survival estimate and details of the component proportions of the different metastatic sites. The remaining treatment arms reported survival estimates for individual sites – 11 soft tissue, 15 bone and 19 visceral. All three quantitative options were explored. Only poorly fitting regression models could be specified and as a reasonable amount of data describing survival from the separate

metastatic sites were identified, their use was preferred to the data including combinations of all sites of metastases. A number of studies also presented data on metastatic survival differentiating with respect to nodal status [Koenders et al, 1992; Venturini et al, 1996]ER status receptor [Koenders et al, 1992; Vogel et al, 1992; Alonso et al, 1995; Venturini et al, 1996], PgR receptor status [Koenders et al, 1992], menopausal status [Venturini et al, 1996], and the administration of prior adjuvant therapies [de Takats et al, 1993; Venturini et al, 1996]. These data were used to estimate a survival multiplier for alternative patients groups.

Each estimate of survival was weighted by the corresponding sample size and three weighted datasets were created. The datasets were used directly, to fit distributions using statistical fitting software, and to define the mean and standard deviation for the theoretically specified gamma distributions.

Though the model structure did not record events following the development of metastases explicitly, the time to progression (TTP) for the alternative sites and therapies were recorded in order to estimate monthly costs for the metastases states (see section 5.4.4.4). Fewer trials presented data on the TTP; such data were available for only 2 of the 19 treatment arms presenting specific data on survival from visceral metastases. The regression models were again poorly specified and so weighted datasets were created using data from studies in which a single metastatic site was dominant in over 50% of patients. The input distributions were defined in a similar manner to the survival data.

5.4.4 Cost parameters

The following sections describe the sources for the estimates of the cost parameters, as well as the methods for specifying input distributions around the baseline point estimates. The collection of data to inform the cost parameters would ideally identify UK-based estimates of the resources used to treat patients experiencing each of the events described in the model and then attach current unit costs to the recorded resource use. Unfortunately, such ideal data is rarely available and a pragmatic approach is required. For the case study

evaluation, alternative approaches were adopted for the different parameter categories depending on the data available. Priority was given to data describing disaggregated resource use rather than aggregate cost estimates because current unit costs could be attached to the former. If only aggregate cost estimates were available they were uprated to year 2000 prices using the health services pay and prices index. UK-based studies were preferred to foreign studies, though when there was no alternative the estimates provided by foreign studies were converted to UK prices using the health services purchasing power parity index for the year in which the data were recorded (and then uprated to year 2000 prices using the health services pay and prices index). The data are presented in four categories covering adjuvant therapies, surveillance, treatment side effects, and relapses. Details of the costing are provided in Appendix 7.

5.4.4.1 Costs of adjuvant therapies

The cost of chemotherapy is an integral parameter in the model and a wide search for alternative estimates of the cost of a cycle was undertaken. Six cost estimates were calculated by combining costs for the three resource elements of a cycle of chemotherapy – drugs, health professional's time, and outpatient visits. The latter two elements were varied according to alternative estimates in the literature [Lober et al, 1988; Lokich et al, 1996], and on the basis of discussion with clinicians involved in the ABC trial. Published estimates of the aggregate cost of the two most commonly administered chemotherapy regimens, CMF and CAF [Silva and Zurrida, 1999], as well as the aggregate cost of a cycle of chemotherapy from the NHS reference costs were obtained [The new NHS - 1998 Reference Costs, 1998]. An estimate of the aggregate cost of chemotherapy was also obtained from an individual NHS Trust hospital to represent the lowest end of the scale. The associated costs are presented in Table 5.1.

The data sources for the cost of a cycle of chemotherapy provided no objective measure of variability that could be used to weight the separate estimates. Subjective weights were defined with respect to the perceived relevance of each estimate to the intended audience for the evaluation. To specify a probability distribution of chemotherapy costs the ten

individual estimates were ranked according to their relevance, as judged by their source. The costs were then inversely weighted according to their rank, so that the highest-ranking cost had a weight ten times that of the lowest ranked cost. This process is presented in Table 5.1.

Table 5.1 Separate cost estimates for a cycle of chemotherapy, and associated rankings and weights

Method	Source	Rank	Weight	Cost
Separate	CMF, baseline (see Appendix 7)	1	10	284.94
	CAF, baseline (see Appendix 7)	2	9	226.65
	CMF, baseline (1 clinic visit)	3	8	198.59
	CMF, health professionals time costs[1]	6	5	141.86
	CMF, health professionals time costs[2]	7	4	700.46
	CAF, health professionals time costs[2]	8	3	467.60
Aggregate	NHS reference cost (HRG v.3)	4	7	269
	E Anglia Trust cost	5	6	67
	CMF[3]	9	2	596
	CAF[3]	10	1	648

All costs uprated to year 2000 prices, foreign costs exchanged using health services purchasing power parity index

[1] [Lober et al, 1988], [2] [Lokich et al, 1996], [3] [Silva and Zurrida, 1999]

Each cost estimate was replicated according to its weight and the resulting collection of estimates was then assembled as a dataset. The weighted dataset was used directly in the models, but also to fit probability distributions. Because cost distributions cannot include values less than zero, only four distributions were tested as feasible distribution types for all the cost distributions: lognormal, gamma, weibull and beta distributions. A gamma distribution had been specified as the theoretical distribution for cost parameters and the mean and standard deviation from the weighted dataset were used to solve for the distribution parameters (see Appendix 5).

5.4.4.2 Costs of surveillance

Surveillance is an ongoing component of the health care that breast cancer patients receive, but it is not constant. Separate estimates of the cost of monitoring patients were established for the first year following primary surgery, subsequent years in the state DFI, and from the point of diagnosis with metastases.

Estimates of the cost of surveillance following primary surgery were derived from a representative selection of published follow-up procedures in clinical trials. During DFI, the intensity of surveillance following primary surgery is known to be higher than in subsequent years. The period of increased surveillance seems to range between 9 months and 3 years [Arriagada et al, 1992; Bonadonna et al, 1995]. As the surveillance procedures reported by clinical trials are generally more intensive than normal practice, the intensity of surveillance was assumed to decrease after the first year. In order to estimate a range of costs three alternative surveillance procedures, each incorporating two levels of the intensity of follow-up, reported by separate clinical trials were itemised and costed. The main cost driver appeared to be the number of breast clinic visits so five alternative unit costs for outpatient/breast clinic visits were also included. In total, 15 estimates of the cost of surveillance for both the first year and subsequent years were established.

No data reporting surveillance of patients following locoregional relaspe were identified so patients in remission were assumed to have similar levels of health service contacts to patients in their first year of DFI. To inform follow-up after metastases a UK-based observational study comparing two options for the surveillance of patients with metastatic breast cancer was identified. The options comprised UICC assessment and Serum marker assessment, each method was costed twice, reflecting an intensive and a less intensive form of surveillance [Robertson et al, 1995]. Given the relevance of the identified study no further exploration of this cost parameter was undertaken.

To create a weighted dataset for the costs of surveillance during DFI the ranking method of weighting the estimates described in the previous section for the costs of chemotherapy was employed. The lowest level of intensity was chosen as the highest ranked pattern of follow-up, though no rank difference was employed between the respective costs. This weighting structure was chosen because surveillance reported in clinical trials is likely be higher than in normal practice, also Liberati has investigated the impact on survival of different intensities of follow-up schedules and detected no difference [Liberati, 1995]. The four estimates for the cost of surveillance for patients with metastases were assigned

equal weights as there were no grounds for increased confidence in one estimate over another.

The three methods for assigning input distributions were derived from the weighted datasets using the data directly, inputting into statistical fitting software and estimating the mean and standard deviation to calculate the parameters for a gamma distribution.

5.4.4.3 Costs of toxicity

To inform the associated cost values, it was necessary to describe the events experienced in each of the toxicity categories. A list of possible events experienced within each category of toxicity was established, and the proportion of patients experiencing each event was estimated from the literature (see section 5.3.1). The baseline cost of treatment for the three main categories - major, grade 3 or 4, and grade 1 or 2 - comprised a weighted cost of the different conditions experienced.

The events described in the major toxicity category were all well defined conditions that have been the subject of economic evaluations in their own right, such as thromboembolic and cerebrovascular events. The costs associated with treatment for such conditions were derived from published economic evaluations [Lloyd et al, 1997; Holloway et al, 1996]. Within the grade 1 or 2, and 3 or 4, categories of toxicity, up to twelve separate conditions were identified. To reduce the complexity, only conditions experienced by more than ten per cent of patients were costed. Very little information was available from the literature review on the cost of treating these less serious forms of toxicity so the treatment of each condition, at the relevant level of toxicity was discussed with a team of health professionals (clinicians and nurses) involved with the ABC trial. Costs were then attached to the resources specified by the team to calculate a baseline cost for each condition.

Due to the lack of data describing the costs of toxicity minimum and maximum costs were estimated as 50% and 150% of the baseline costs for each category of toxicity. Given a baseline estimate and a minimum and a maximum value, the statistical fitting software

estimated a triangular distribution for each level of toxicity. The triangular distributions also represented the empirical estimates of the costs of toxicity as only one real cost estimate was available for each form of toxicity. To inform the theoretically defined gamma distribution, the baseline estimate was defined as the mean and the standard deviation estimated to be one quarter of the range.

5.4.4.4 Costs of relapse

Few quantitative data were identified to cost the time spent in relapse states. It was evident that the quantitative data alone were insufficient, so the cost estimates were derived from a combination of quantitative and qualitative data. Differential costs were estimated for each of the four sites of relapse included in the structure of the model – locoregional, soft tissue, bone, and visceral.

The treatment for locoregional relapse was included as a one-off cost because patients were assumed to move from the state 'locoregional relapse' to 'remission' after one month, though parts of the treatment, such as the administration of chemotherapy and radiotherapy, may continue for a number of months. Using the available qualitative data the treatment of locoregional relapse was costed as a single event requiring surgery to remove the returning tumour, followed by radiotherapy and chemotherapy. The cost of surgery was obtained from the National Schedule of Reference Costs [The new NHS - 1998 Reference Costs, 1998]. In addition to the mean value, the minimum and maximum costs of 'Intermediate Breast Surgery' from all NHS Trusts were used to define a range of values for the costs of surgery. Little variation was found around the cost of a fraction of radiotherapy [Read, 1994], but the number of fractions undertaken appeared to be more uncertain. Different estimates of the number of fractions were taken from the literature to inform a range [Aberzik et al, 1986; Toonkel et al, 1983]. The cost of chemotherapy following locoregional relapse was described using the same estimates and distributions derived for the cost of adjuvant chemotherapy.

Using the defined ranges and the mean values for the cost of surgery and the number of fractions of radiotherapy, triangular distributions were specified by the statistical fitting software. The three elements of locoregional treatment were each assigned distributions that were incorporated in the models and the aggregate cost of treating locoregional relapse was estimated within the model.

For the model analysed using theoretical distributions a distribution of aggregate costs was established. To establish a dataset of costs a Monte Carlo simulation was undertaken. The distribution describing the number of fractions of radiotherapy was linked to the cost of radiotherapy, which was linked to the distributions describing the costs of surgery and chemotherapy. For each simulation a value from each distribution was sampled and an aggregate cost of the treatment of locoregional relapse estimated. From a total of 1,000 simulations a mean and standard deviation were estimated to inform the parameters of the theoretically defined gamma distribution.

The cost of treating patients with metastases varies considerably over time. Several studies have remarked on the moving profile of resource use over the period of relapse [Hurley et al, 1992; Will et al, 1998; Baker et al, 1991; Koopmanschap et al, 1992], with the final three to six months of life being the duration of expensive terminal care, whilst the initial three months are also relatively resource intensive. However, the model did not differentiate between the different stages of metastases because more data were available that informed aggregate costs of treatment for metastases. Protocols were again developed from the literature to estimate separate monthly costs for the treatment of the metastases with respect to systemic therapies, local treatment, and inpatient episodes.

The appropriate systemic therapeutic strategy for individual patients is influenced by a number of characteristics, including the extent of disease, DFI, ER status and age [Hortobagyi, 1998]. Treatment for both soft tissue and bone metastases was differentiated with respect to good and bad prognoses using the only available prognostic indicator - the length of DFI (good DFI>2years, bad DFI<=2 years). There was a general consensus in the literature on the ordering of different types of therapy for metastases, with respect to the use

of hormonal therapy and chemotherapy. The two schemas' presented in Figure 5.2 illustrate a similar choice of therapies for alternative patient characteristics [Coleman and Rubens, 1987; Leonard et al, 1994].

Figure 5.2a Schema for the treatment of metastases[Coleman and Rubens, 1987]

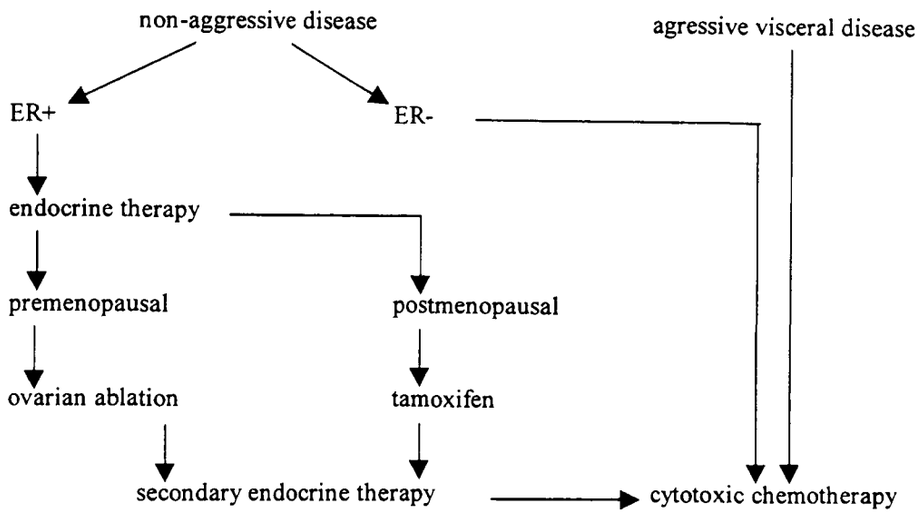
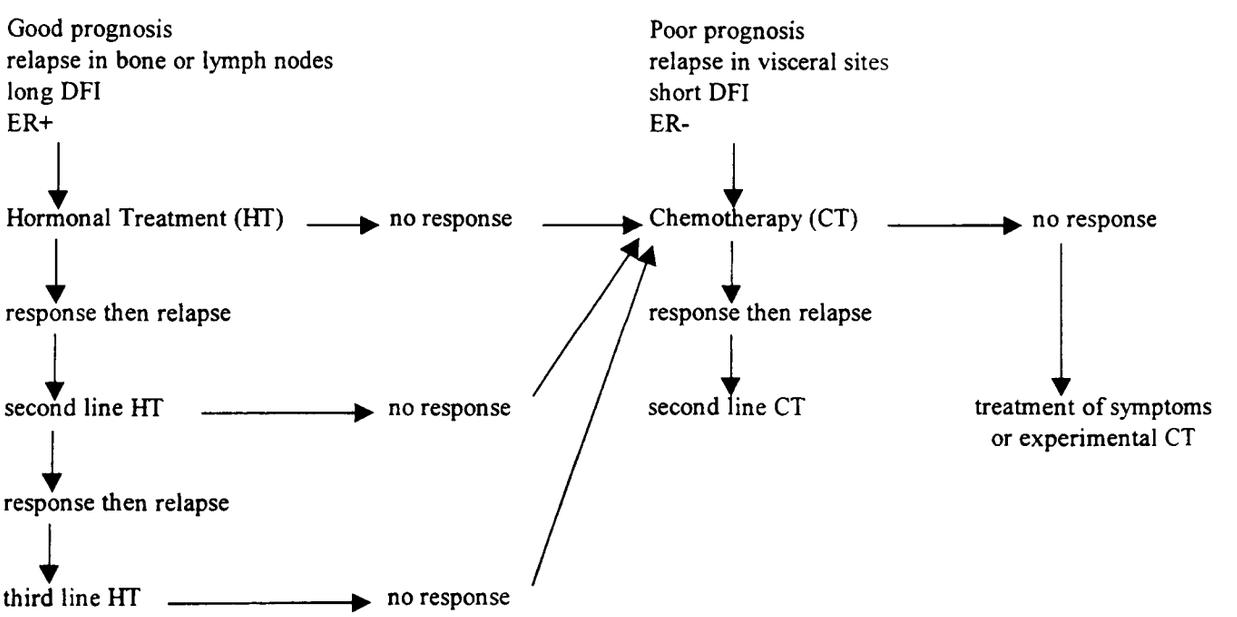


Figure 5.2b Schema for the treatment of metastases[Leonard *et al*, 1994]



The second element of resource use in treating metastases included all local treatment. The treatment for the local conditions associated with the alternative metastatic sites were mostly drug therapies and the costs were taken from the British National Formulary, the cost of radiotherapy had been estimated previously. The types of treatment administered to

patients experiencing the different forms of metastases, and their frequency, were informed by various studies identified during the literature review. The final resource category included all contacts with hospital services, from outpatient visits for monitoring to inpatient admissions, which were informed by one quantitative study that differentiated between the types of relapse with respect to hospital contacts [Hurley et al, 1992]. Data from English Trust returns including 55 Trusts were used to estimate the mean cost per patient day in medical oncology specialities and per outpatient visit.

The estimation of the monthly costs associated with each of the three metastatic sites comprised the aggregation of the various components – systemic therapies, local treatments, hospital contacts and surveillance. The monthly cost of systemic therapies was calculated on the basis of mean TTP and survival. Using the presented schema's for the progression of systemic therapy patients received their appropriate therapy until disease progression, at which point they would receive the next appropriate form of therapy. The addition of therapies continued until death, or until patients received two full courses of chemotherapy. The total cost was divided by the months of survival to estimate a monthly cost. Table 5.2 provides an example of the calculation of the cost of systemic therapies.

Table 5.2 An example of the calculation of the costs of systemic therapies for patients experiencing bone metastases, with a DFI of 22 months

Month	Therapy	Cost per month	Total cost
0 – 7.4	Anastrozole	83.16	615.38
Response to anastrozole:			
7.5 – 14.9	Megace	29.30	216.82
15 – 22	Chemotherapy*	285.00	1710.00
Cost per month			115.55
Non-response to anastrozole			
7.5 – 15.7	Chemotherapy*	285.00	1710.00
15.8 – 22	Chemotherapy*	285.00	1710.00
Cost per month			183.43

* chemotherapy is assumed to be administered for six cycles in each case.

To represent the uncertainty in the cost of treating metastases probability distributions representing the aggregate monthly costs for each metastatic site were established. Firstly, probability distributions were defined for each significant factor affecting the aggregate cost. Bootstrapping the observed estimates created distributions around the relevant categories of TTP and survival. The cost of chemotherapy was assumed to be similar to the

cost of adjuvant chemotherapy. In the local treatment of soft tissue and visceral metastases no probability distributions were assigned as the drugs administered have a negligible cost. Probability distributions were originally assigned to the number of fractions of radiotherapy received as treatment for bone metastases, but following the validation process the frequency of radiotherapy was revised downwards to a level that did not warrant inclusion as a probability distribution. There was a wide variation in the estimates of the cost of hospital visits over the country, which informed mean, minimum and maximum estimates for the three types of contact – inpatient, day case and outpatient. Using these data gamma distributions describing the cost of each element of treatment for metastases were defined. The distribution describing the cost of surveillance discussed in the previous section on surveillance was employed.

Using the defined costs and associated probability distributions a similar process to estimating probability distributions for the aggregate costs of treating locoregional relapse was employed. The individual treatment components were linked so the respective components of each form of metastases were aggregated and Monte Carlo simulations were run sampling a value from each of the specified probability distributions. 1000 observations of the monthly costs for each cost element were generated. The resulting dataset was used directly within the model, as well as inputted into statistical fitting software to estimate aggregate distributions for each site of metastases. Finally, the mean and standard deviation from the datasets were used to solve for the parameters of the theoretically defined gamma distributions.

5.4.5 Utility values

Quality of life in breast cancer patients has been investigated, but the majority of identified studies used condition- or symptom specific measures. The conversion of non-generic measures to utilities was not attempted, though the identified data could be used to inform descriptions of health states if the primary collection of health state utilities were to be considered. Few utility data associated with health states that a breast cancer patient may experience were identified. A number of previous modelling studies had assigned utility

values using either focus groups of oncology professionals [Hillner et al, 1992; Hillner and Smith, 1991; Smith and Hillner, 1993; Desch et al, 1993], or direct measurement techniques from oncology nurses [Hutton et al, 1996]. In addition, three studies were identified that reported utility values from primary studies aimed at defining quality of life in patients experiencing health states similar to those described in the ABC models [de Haes et al, 1991; Daly et al, 1993; Ashby et al, 1994].

Empirical utility weights for the separate toxicity categories were not identified in the literature review, only an aggregate weight for all patients receiving chemotherapy [de Haes et al, 1991]. Using these data the individual toxicity states were assigned utility values so that the weighted aggregate utility value (using the proportions experiencing the different toxicity categories) would equal the value quoted in the literature (0.72). The process is demonstrated in Table 5.3.

Table 5.3 Example of interpolation to estimate separate utility values for toxicity states

	Proportion of patients experiencing event	Utility value	Proportionate utility value
Aggregate utility weight		0.72	
Grade 1/2 toxicity	0.57	0.78	0.44
Grade 3/4 toxicity	0.38	0.65	0.25
Major toxicity	0.05	0.51	0.03
Sum			0.72

No utility values were found that described the period of remission following locoregional relapse, so the value for the period 2 months to one year following mastectomy was used as a proxy [Ashby et al, 1994]. Likewise, no direct weights were identified for patients experiencing visceral, bone or soft tissue metastases. Thus, the utility value reported for patients with metastases receiving chemotherapy was used as the proxy value for visceral disease. The utility value for bone or soft tissue metastases was calculated as a weighted proportion of the values for patients receiving hormonal therapy and chemotherapy for metastases, reflecting the proportion of time spent receiving the alternative therapies [de Haes et al, 1991].

From the available data, point estimates were specified for each of the health states included in the ABC models. For some health states there were sufficient data to estimate a mean and a range for the associated utility value, ie. three separate values. However, for other health states only one point estimate could be identified. In the latter cases relaxed (wide) estimates of the possible ranges of utility values for the respective health states were made after consultation with economic and clinical colleagues.

Inputting three values into statistical fitting software estimated a similar probability distribution for each utility value – the triangular distribution. Too few data were available to justify inputting the empirical data directly so the triangular distributions were used to represent the empirical data. Using the specified ranges, the parameters for the beta distribution were estimated assuming that the standard deviation of the required distribution was a quarter of the range. With the mean and variance the beta parameters were estimated by simultaneously solving the equations presented in Appendix 5.

5.5 Conclusions

This Chapter has described the process of reappraising the theoretical model structure given the identified data, as well as the analysis of the data into a suitable format to populate a decision model. The process of harmonising data for selected input parameters was introduced as a means of improving the homogeneity of the available information.

The reappraisal of the preliminary model structure was primarily based on the format of the identified data. The alterations made to the model structure were subtle changes that did not remove (add) any important events from (to) the model. The most significant difference combined operable and inoperable locoregional relapses within a single remission state because the identified data described pathways from locoregional relapse in that manner. Such an alteration simply moved the model from an explicit consideration of operable and inoperable relapses to an implicit account of these events. The reappraisal described in this Chapter demonstrated that a careful reconsideration of the preliminary

model structure can make better use of the identified data without compromising the theoretical basis of the model.

Much of the process of harmonisation involved the identification of differences in the definition of similar events presented by different studies. To facilitate the combination of such data, explicit adjustments were made to the values reported by one or more of the relevant studies so that the comparability of underlying parameter definitions were improved. The application of harmonising the data reported in this Chapter showed that the process cannot be governed by hard and fast rules as the adjustments made to the data will depend on the event described and the format of the available data. However, the harmonisation of the case study data illustrated a range of issues potentially relevant to other disease areas. For example, during the harmonisation of the DFI data, the studies that treated death as a censored event reported the number of deaths observed, which were then used to adjust the original disease-free survival data. If the relevant data to harmonise a parameter were not presented by a study, data from studies with the most similar patient and treatment characteristics were used to adjust the initial estimate.

The explicit presentation of the harmonisation of the data is necessary to enable the reader to judge the appropriateness of the parameter definitions and the alterations made to the data. The data adjustments are possible solutions to the noted divergence in the definition of events.

This Chapter also demonstrated alternative methods for the specification of probability distributions to represent the uncertainty around the values of the input parameters. When creating the weighted datasets, the preferred approach to weighting the data used the variance associated with each identified value. If the variance was unknown then the sample size informing each value was employed as the weight. Some cost parameters were informed by various primary estimates with no objective measure of variance, whereby subjective methods of weighting the data according to their perceived relevance to the objective of the study was introduced. Least satisfactorily, some cost and utility parameters were informed by only one identified value in the literature. In such cases wide ranges of

values were specified to reflect the uncertainty. The specification of theoretically derived probability distributions for alternative groups of parameters and the use of defined formulae to estimate the distribution parameters provided a simpler method of specifying probability distributions. However, this latter method did not provide an alternative to the subjective weighting of the cost and utility parameters as the distribution parameters for these variables were necessarily based on the subjectively defined weighted datasets, or the specified ranges.

The comparison of the alternative methods for defining probability distributions around the input parameters is a secondary objective of this thesis. The results of the comparison are presented in Chapter 8, the implications of which are discussed in Chapter 9.

Chapter 6 Case study: implementing the model

6.1 Introduction

The two previous Chapters described the identification and analysis of data to populate a decision model, which incorporated the specification of a preferred model structure. The next step in the modelling process is the development of an analysable model. This Chapter describes the implementation of two separate decision models, a Markov process and a discrete event simulation (DES) model, to represent the pathways of patients diagnosed with early breast cancer. The aim of this Chapter is to describe the analytic input required to develop the different modelling techniques, as well as to explicitly define the differing representations of the data used to evaluate alternative therapies for the early breast cancer.

The first two sections justify the implementation of two alternative decision models, and the application of stochastic analyses. The implementation of each model is then described, starting with the Markov model. For each modelling technique the implementation process is presented in two parts. Firstly, the process of establishing the relationships between the health states included in the structure of the model is presented (Modelling the health states). The second section describes the population of the models using the previously presented probability distributions, and the collection of the models' outputs for further analysis (Controlling inputs and outputs). A short section highlighting the differences in the modelling assumptions between the two techniques is then presented.

During the process of implementing the models two methodological issues were raised that affected the analysis of the models. These are explored in the final sections of this Chapter. The first issue related to the use of the DES model, where the question of an adequate sample size to minimise the impact of first-order uncertainty is addressed (see section 3.6.1.1). The second issue involved an assessment of alternative methods for describing probabilistic 'length of time' input parameters.

6.1.1 The choice of modelling technique

To build a decision model an appropriate modelling technique must be chosen. The economic evaluation employed as the case study for this thesis – a cost-utility analysis of alternative adjuvant therapies for early breast cancer – incorporates a relatively long time horizon. As discussed in Chapter 2 the decision tree technique can be adapted to cover patient pathways with long time horizons, but such models soon become cumbersome. The effective choice of modelling technique for the case study, therefore, was between a Markov model and a DES model. Only a partially informed choice could be made between the two alternatives as their relative merits have only been explored superficially [Chaussalet et al, 1999; Karnon and Brown, 1998]. From these earlier considerations of Markov modelling and DES it was clear that neither alternative dominated the other in every aspect of the modelling process. For example, it appears that DES allows greater flexibility in the representation of patient pathways, but a Markov model is generally easier and quicker to build. No empirical comparison of the two techniques was identified. To inform the appropriate choice of modelling technique both a Markov process and a DES model were developed to evaluate the economic impact of the relevant alternative adjuvant therapies.

A Markov process, rather than a Markov chain, was chosen because time dependent transition probabilities (dependent on the time spent in the model) were considered to be important in the modelling of early breast cancer (see section 2.3 for full definition of Markov processes and Markov chains). The Markov process was built as a cohort-based model, whereby a cohort of patients is sent through the model, rather than following single patients through the model (first-order Monte Carlo simulation). A cohort-based Markov process is simpler to build and provides a starker contrast with a DES model, which can only be analysed using first-order Monte Carlo simulation.

Moreover, first-order Monte Carlo simulation-based Markov processes (as opposed to a cohort-based Markov process) only provide additional information about first-order uncertainty, which is irrelevant to the required outputs of the model (see section 3.4.3). The cohort-based Markov process is also the most common form in the economic evaluation literature.

6.1.2 Stochastic or deterministic?

Both modelling techniques can be analysed deterministically, whereby single estimates of cost-effectiveness are generated for individually specified sets of input parameter values, or stochastically, whereby the output consists of distributions of the relevant costs and effects based on sets of input parameter values randomly sampled from specified probability distributions. For the economic evaluation of health care technologies it has been proposed that stochastic analyses provide a more realistic method of describing uncertainty in the overall results of decision models [Fenwick et al, 2000]. Stochastic models were chosen because such models facilitate the statistical analysis of model outputs for resource allocation decisions in the present, as well as enabling the statistical evaluation of the value of obtaining further primary information about the input parameters. Both of these objectives are explored in thesis.

6.2 Building the decision models

Wherever possible, the general points of good practice described in Chapter 3 were incorporated into both models. For example, no decision nodes other than at the root of the model were included and all common parameters were linked.

The specification of a minimum time period of advancement was required for the DES model, as well as a cycle length for the Markov process. Though a case could be made for differential timing on the basis of model running time, it was decided that the patient pathways could be most accurately represented in both models using a time period of one month. The choice of one month as the models' cycle length meant that the annual probabilities collected in the literature review, describing the length of DFI and remission, needed converting to monthly probabilities. Transition rates were assumed to remain constant over the year, because no data were identified that contradicted this

assumption. Formula 6.1 was used to convert the annual probabilities to monthly probabilities [Miller and Homan, 1994]:

$$P_{monthly} = 1 - [1 - P_{annual}]^{1/12} \qquad (6.1)$$

Where P is the probability of an event. A similar time horizon for both models was also adopted with patients living to a maximum of 100 years. Separate descriptions of the process of building the two models are presented in the following sections. In each section the description of the actual modelling of the health states is followed by the necessary structures for controlling the model inputs and outputs.

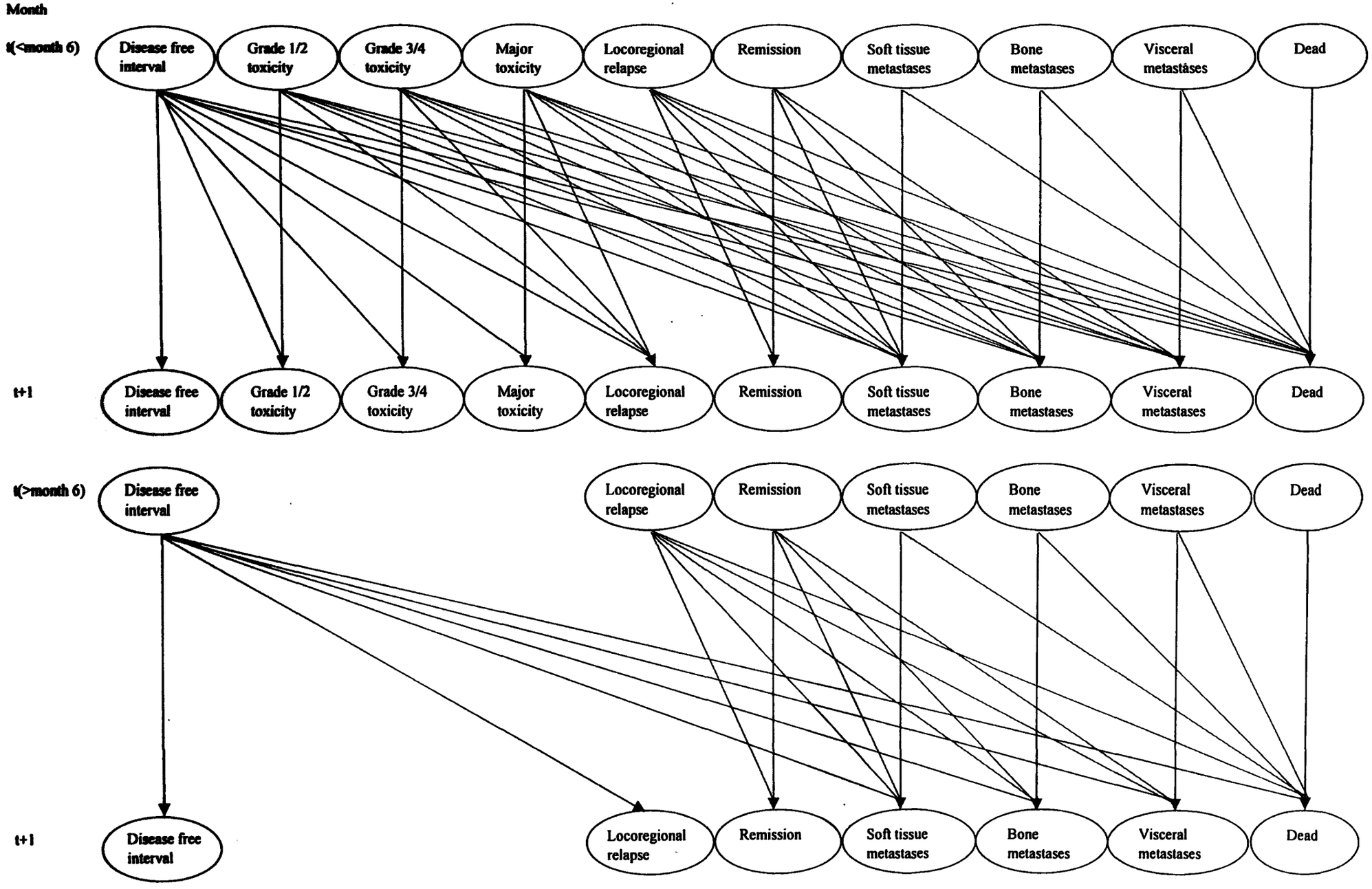
6.3 The Markov process

The Markov process was built in Excel using the Crystal Ball 2000 add-in, which is a risk analysis programme that is "easy to learn and easy to use" [Crystal Ball, 2000]. A deterministic Markov process can be built using the spreadsheet package alone, but Crystal Ball uses second-order Monte Carlo simulation to analyse the model stochastically. Sets of parameter values are randomly sampled from the input distributions, each of which is analysed using the cohort-based method. The process of building the model is described in chronological order.

6.3.1 Modelling the health states

Firstly, the model structure was transferred to the spreadsheet, which is presented in Figure 6.1, which shows the possible transitions that could be made within the model moving from period 't' to 't+1' at different phases of the model. All patients began the model in the health state 'disease free interval' (DFI). During the first six months patients in the DFI state could experience toxic effects caused by the adjuvant therapies, a relapse in one of the four specified sites of relapse – locoregional, soft tissue, bone or visceral, or they could die from a cause other than breast cancer. Patients could also experience a relapse or die from other causes from the toxicity states. In the absence of relapsing or dying patients in the toxicity states returned to the original DFI state at the end of the first six months.

Figure 6.1 Representation of the possible patient pathways within the ABC Markov process



Patients could experience one of three forms of toxicity, though the different forms of toxicity could not be experienced simultaneously. There were two reasons for this assumption. Firstly, four extra health states would be required to represent the occurrence of the joint toxicities. Secondly, further assumptions would be required to describe the probabilities of experiencing the different combinations of toxicity because no data describing the proportion of patients experiencing joint toxicities were identified. Other than the probability of experiencing toxicity, patients in the DFI state were subject to monthly probabilities of moving to a relapse state or straight to the dead state (with no evidence of disease). After the first six months, the proportion of patients remaining in the DFI state in month t was estimated using formula 6.2.

$$pr(DFI)_t = pr(DFI)_{t-1} - pr(endDFI)_t \qquad (6.2)$$

The probability of moving to a particular site of relapse incorporated the probability of leaving DFI and the probability of experiencing the different forms of relapse. For example, the proportion of patients experiencing a locoregional relapse (LR) in any given month was described using formula 6.3

$$pr(LR)_t = pr(endDFI)_t \times pr(LR / endDFI)_t \qquad (6.3)$$

Patients experiencing locoregional relapses were assumed to remain in that state for a maximum of one month, after which they entered a period of remission, experienced a more severe metastatic relapse, or died from other causes. From remission, patients could experience metastases or die from other causes, from the metastases states patients could only die. Formula 6.4 described the proportion of patients entering, and remaining in, a metastatic state of relapse (MR), because patients could enter from DFI, locoregional relapse (LR) and from remission following a locoregional relapse (RM), as well as the patients remaining in a state from one cycle to the next.

$$pr(MR)_t = pr(MR)_{t-1} - pr(endMR)_t + [pr(DFI)_{t-1} \times pr(endDFI)_t \times pr(MR / endDFI)_t] + [pr(LR)_{t-1} \times pr(endLR)_t \times pr(MR / endLR)_t] + [pr(RM)_{t-1} \times pr(endRM)_t \times pr(MR / endRM)_t] \qquad (6.4)$$

From the metastatic states the monthly probability of dying was converted from the median length of survival using the formula presented in section 3.5.1.

6.3.2 Controlling inputs and outputs

The actual Markov process ran on one spreadsheet, wherein each health state was assigned a column and each successive cycle (month) was represented by a row. From this allocation of columns and rows the basic operation of the Markov process involved describing the proportion of patients placed in each health state in each time cycle of the model. Using the formulae described in the previous section to represent the movement of patients between states the building of the model was relatively straightforward.

Separate spreadsheets were established to hold the relevant data on clinical parameters, costs and utility values. Within each of the parameter spreadsheets the relevant probability distributions were established using the Crystal Ball add-in [Crystal Ball, 2000]. Stochastic analyses consist of multiple deterministic analyses using alternative sets of parameter values that are randomly sampled from the defined probability distributions for each input parameter. Each deterministic analysis is defined as a 'run'. A single deterministic analysis (run) of a Markov process involves the following steps:

1. The relevant monthly cost and utility values are multiplied by the proportion of patients in each state in each cycle (month);
2. The respective costs and utility values are summed for each cycle;
3. Separate discount factors are applied to the cost and utility values assembled for each month;
4. Aggregate values for the model outputs are calculated by summing the costs, life years and utilities for all cycles.

The stochastic analysis consisted of 10,000 separate runs. Crystal Ball ran the necessary number of second-order Monte Carlo simulations 'solving' the model for each run. The required output data from all the runs – total costs, life years and QALYs – was collected in the form of frequency charts, which was easily extracted for further analysis. A full model was built and submitted for verification and

validation in less than 2 weeks. Using a 700MHZ PC with a pentium II processor 10,000 trials were completed in around 1 hour.

6.4 The discrete event simulation (DES) model

The first task in building a DES model was to identify an appropriate software package that could handle the characteristics of an economic HTA decision model, and which did not require a high-level of programming skills. Five software packages were identified for building simulation models – Simul8, Witness, Microsaint, ithink, and Powersim. On the basis of cost and perceived user-friendliness the Simul8 software was chosen [Simul8, 2000]. The process of learning to use the software was undertaken on a separate project that employed a simpler model comparing a hospital at home scheme with inpatient care for elderly patients [Campbell et al, 2000]. The process was gradual and continued into the time spent building the ABC DES model. The building of the model is described below.

6.4.1 Modelling the health states

The structure of the DES model is presented in Figure 6.2, which is the visual interface of the simulation software. Relationships between states, activities within states, and the collection of data associated with each state were handled by programming code. Code was implemented as patients entered a state, when they were within a state, or as they left a state. This will become clearer as the process of building the DES model unfolds.

As discussed in Chapter 3 the model was built up in modules [Pidd, 1989]. The first module described the period of time spent in the state DFI. DFI was not the first state because it was easier to assign the time spent in DFI prior to the patient entering the state. Within the first state 'trial entry' every patient passed through a loop command that sampled from a binary probability distribution for consecutive months until a 'relapse' was experienced. At that point the month was noted and transferred to the DFI state to inform the length of DFI. The binary distributions reflected the probability of experiencing a relapse in each month that the patient remained disease free. For example, if the monthly probability of experiencing a relapse was 0.01,

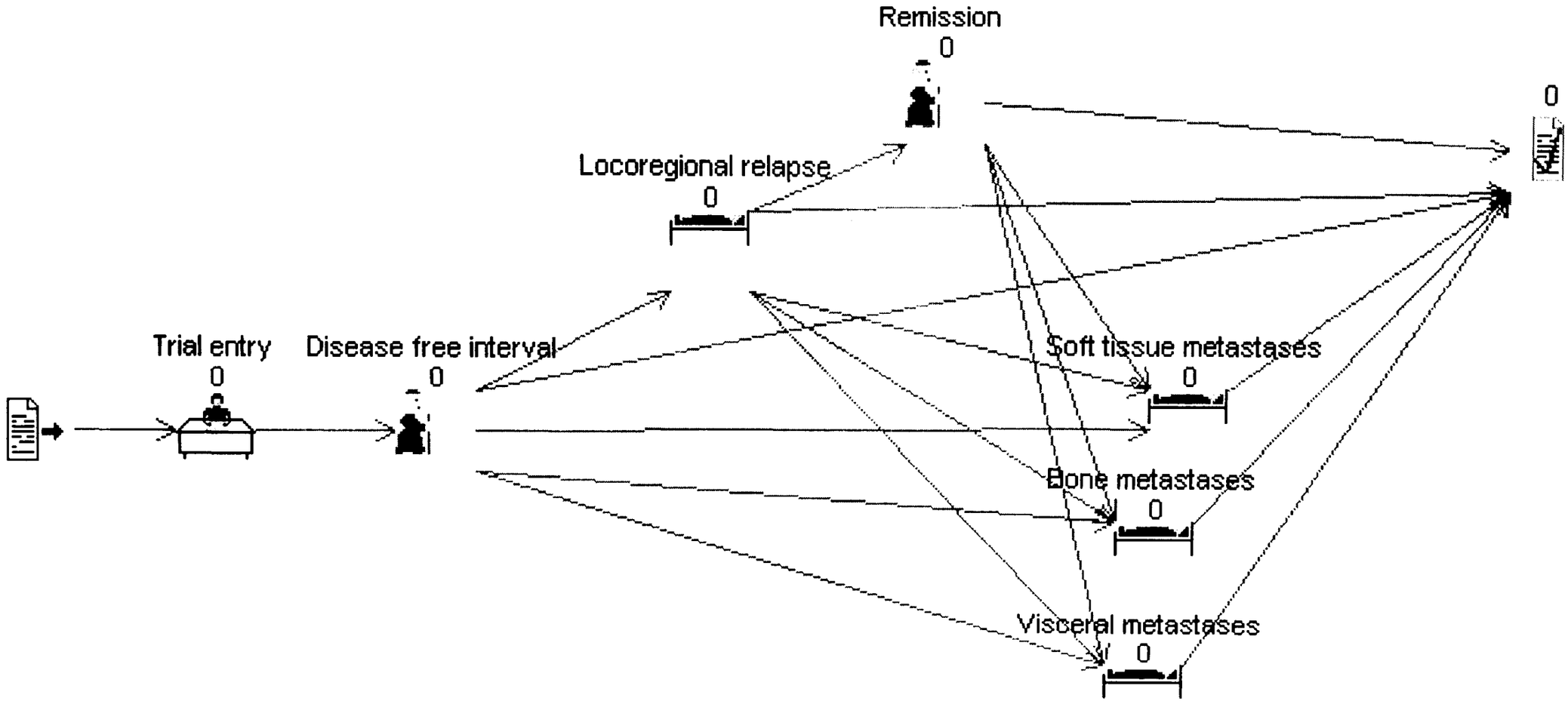
patients sample from a binary distribution with a 1% chance of experiencing a relapse in that month.

The most noticeable difference between the DES and Markov process was that fewer health states were described in the DES model. In the DES model, toxicity was modelled as an attribute of the health state DFI, rather than as individual states. Incorporating toxicity as part of DFI meant that the model could describe patients experiencing different types of toxicity simultaneously. During DFI patients sampled from a series of binary distributions to determine which, if any, forms of toxicity were experienced, and the respective durations of toxicity. The collection of the cost and utility data associated with the DFI state was especially complicated because the impact of toxicity had to be incorporated. It was possible for a patient to experience more than one type of toxicity at the same time; the associated costs of each event experienced were attached to the patient. In addition, the length of major toxicity was sampled from pre-specified distributions (the graded toxicities were assumed to last as long as the number of cycles of chemotherapy). The utility value attached to individual months spent in DFI was a function of the length of the different types of toxicity. For example, for a patient experiencing two types of toxicity, the utility value associated with the less severe form of toxicity would only be required if the less severe toxicity lasted longer than the more severe form.

Prior to leaving DFI the destination of the patient was decided. The five probabilities covering the likelihood of experiencing each of the possible events – locoregional, soft tissue, bone, or visceral relapse, or death – were combined and adjusted proportionately so that their sum equalled one. Each patient then sampled from the combined distribution to determine which health state they moved to from DFI.

The second module represented the passage of events from the point of locoregional relapse to either a metastatic relapse or death. Again, patients remained in the locoregional relapse state for exactly one month before moving onto the remission state. During the locoregional relapse state the relevant cost and utility value was attached to patients. Analogous to the role of the trial entry state, the length of remission was also determined using a similar process of looping through months until an event was experienced. At the end of patients' time within the remission

Figure 6.2 Representation of the possible patient pathways within the ABC DES model



state, a value was randomly sampled from a combined distribution representing the subsequent states that patients could enter.

The final module described the experience of patients within the three metastatic relapse states. As the only possible exit state from these states was death they were relatively easy to model. The patients simply entered the relevant state, remained in it for the allotted period, and then moved into the dead state.

6.4.2 Controlling inputs and outputs

In the DES model, a single run followed a large number of individual patients through the model. At the start of each run a set of parameters from each distribution was randomly sampled from the specified probability distributions and applied to each patient. At the end of a run only the mean values for each of the model outputs were required. Unlike the Markov process, first-order uncertainty was an issue in the analysis of DES models because no definitive estimates of the outputs for each run were obtained. Applying the same input parameter values to different sets of patients (ie. using different random number seeds) produced varying mean values for the model outputs. The only way to reduce this variation was to increase the number of patients included in each run.

Within the model the cost- and utility-based experiences within each state had to be described at the end of the state and aggregated as patients left each state. This aspect of the programming code was the most complicated due to the need to discount the outputs, which involved sectioning the time spent in each state into the corresponding years that the patient had spent in the model. Annual discount rates were applied because the monthly description of the cost and utility effects within the DFI state, which were dependent on the length of time spent in the state and the types of treatment side effects experienced, was deemed to be to complex.

Running totals of the costs and QALYs were attached to each patient as she passed through the health states. On entering the dead state the final totals for each patient were stored in an internal spreadsheet. For each run, the mean values for the model outputs were calculated and stored in a separate internal spreadsheet. When data from

a sufficient number of runs had been collected the data was exported to a spreadsheet for further analysis.

2,500 runs of 10,000 patients, running the model consecutively, took around 72 hours. Around two months were required to build a model to the point at which it could be subjected to the verification and validation process. Due to the learning curve the time required to build future models should be appreciably reduced.

6.5 Comparison of modelling assumptions

The respective assumptions incorporated in each model are presented in Table 6.1.

Table 6.1 Respective assumptions employed in the Markov process and the DES model

Area	Markov process	DES model
General	Cycle length of 1 month. Maximum age of patients 100 years old.	Minimum time period of 1 month. Maximum age of patients 100 years old.
Disease free interval	Annual probability of leaving DFI informed by survival curve. Annual probabilities converted to monthly probabilities assuming constant transition rates.	Annual probability of leaving DFI informed by survival curve. Annual probabilities converted to monthly probabilities assuming constant transition rates.
Toxicity	Patients can experience one of three categories of toxicity. Grade 1/2, 3/4 and major toxicity end after 6 months.	Patients can experience any combination of three categories of toxicity. Grade 1/2, 3/4 toxicity end after 6 months, length of major toxicity sampled from probability distribution.
Locoregional relapse	Patients remain in locoregional relapse for exactly 1 month.	Patients remain in locoregional relapse for exactly 1 month.
Remission	A mean time spent in remission employed. Patients subject to a constant monthly probability of leaving the remission state.	Annual probability of leaving remission informed by survival curve. Annual probabilities converted to monthly probabilities assuming constant transition rates.
Metastases	A median time spent in metastatic states employed. Patients subject to a constant monthly probability of dying.	A median time spent in metastatic states employed. Patients remain in metastatic states for set amount of time.

6.6 Assessing an adequate sample size to minimise the impact of first-order uncertainty on the results of the DES model

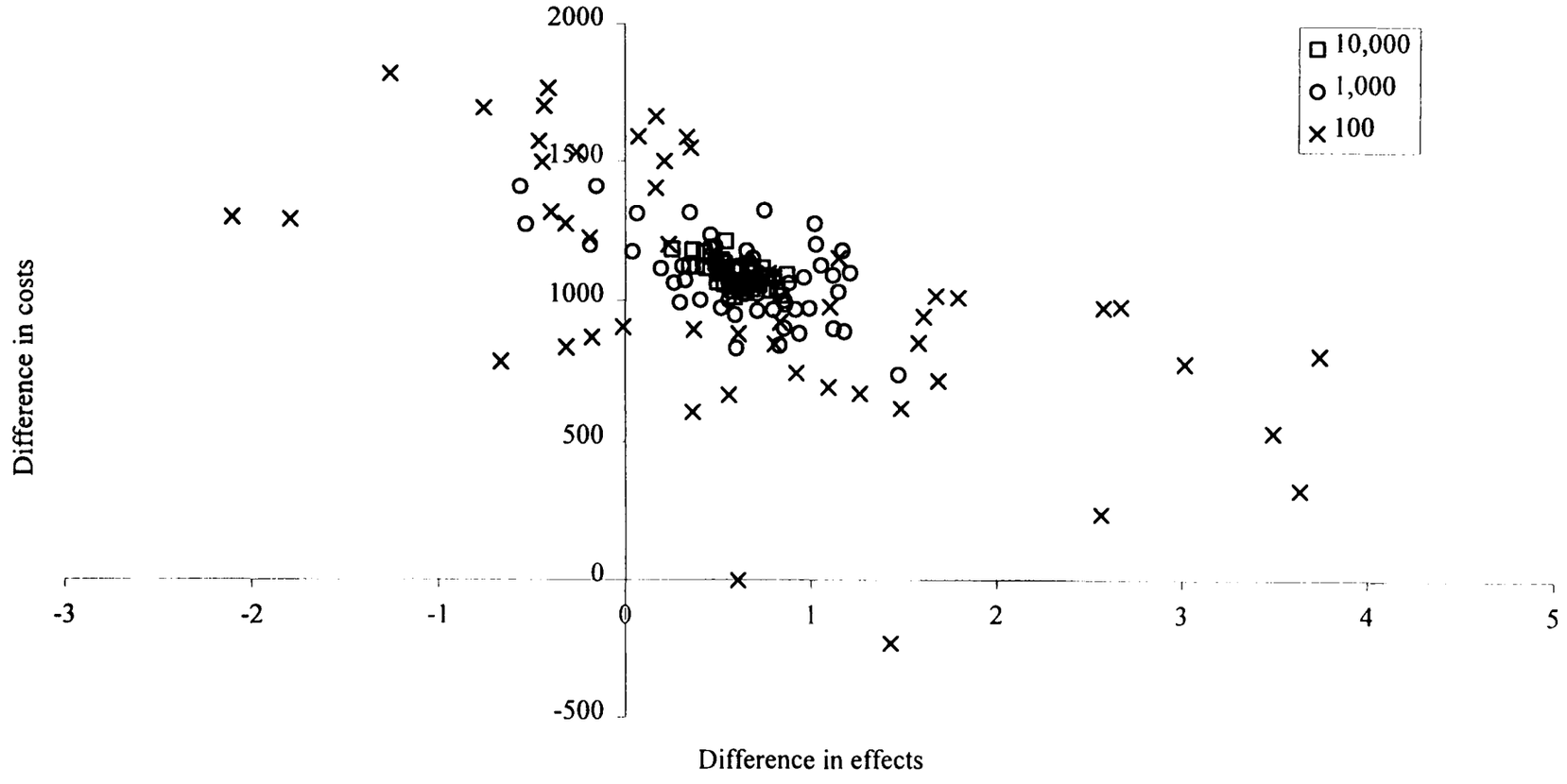
This section addresses a methodological issue that was first raised in section 3.6.1.1, which affects the analysis of the DES model. The DES model was analysed using first- and second-order Monte Carlo simulation. Sets of input parameter values were

randomly sampled from the specified probability distributions (second-order), which were analysed by sending a 'large' number of individual patients through the model (first-order). For each set of parameter values the only results of interest were the mean values for the model outputs, the variation around these means measured first-order uncertainty, which should not influence resource allocation decisions [Briggs, 2000]. The mean output values estimated for each set of parameter values were collated and analysed to inform the second-order uncertainty associated with the model, which describes the variation within the population of eligible patients.

The question addressed in this section is what is an adequately 'large' number of patients to be run through the first-order Monte Carlo simulations to be sure that the mean values obtained for the required outputs are as close as possible to the true mean value for each parameter set? To test the adequacy of alternative sample sizes the input parameter values were held constant at their mean values. Repeated replications of the model were undertaken using different random number seeds for each replication of the same sample size. The use of alternative random numbers meant that patients within alternative replications sampled different values from probability distributions within the model, but the mean values of the model outputs *should* have remained the same. Sample sizes of 100, 1,000 and 10,000 were tested. For each sample size the DES model was run fifty times, each run starting from a different random number seed.

The results of the analyses are presented in a cost-effectiveness plane in Figure 6.3. The plots show a very wide dispersion of estimates derived from runs of 100 patients, whilst the level of variation remained large when 1,000 patients were run through the model for each set of parameter values. A much tighter concentration of estimates is observed for the runs informed by 10,000 patients with the difference in costs varying between around £1,000 and £1,250 and the effects difference falling between 0.5 and 1 QALY. Though larger samples would reduce the observed random error further, runs of 10,000 patients provided the best trade-off between accuracy and the running time of the DES model.

Figure 6.3 Cost-effectiveness plane illustrating the accuracy of alternative run sizes for the analysis of the DES ABC model



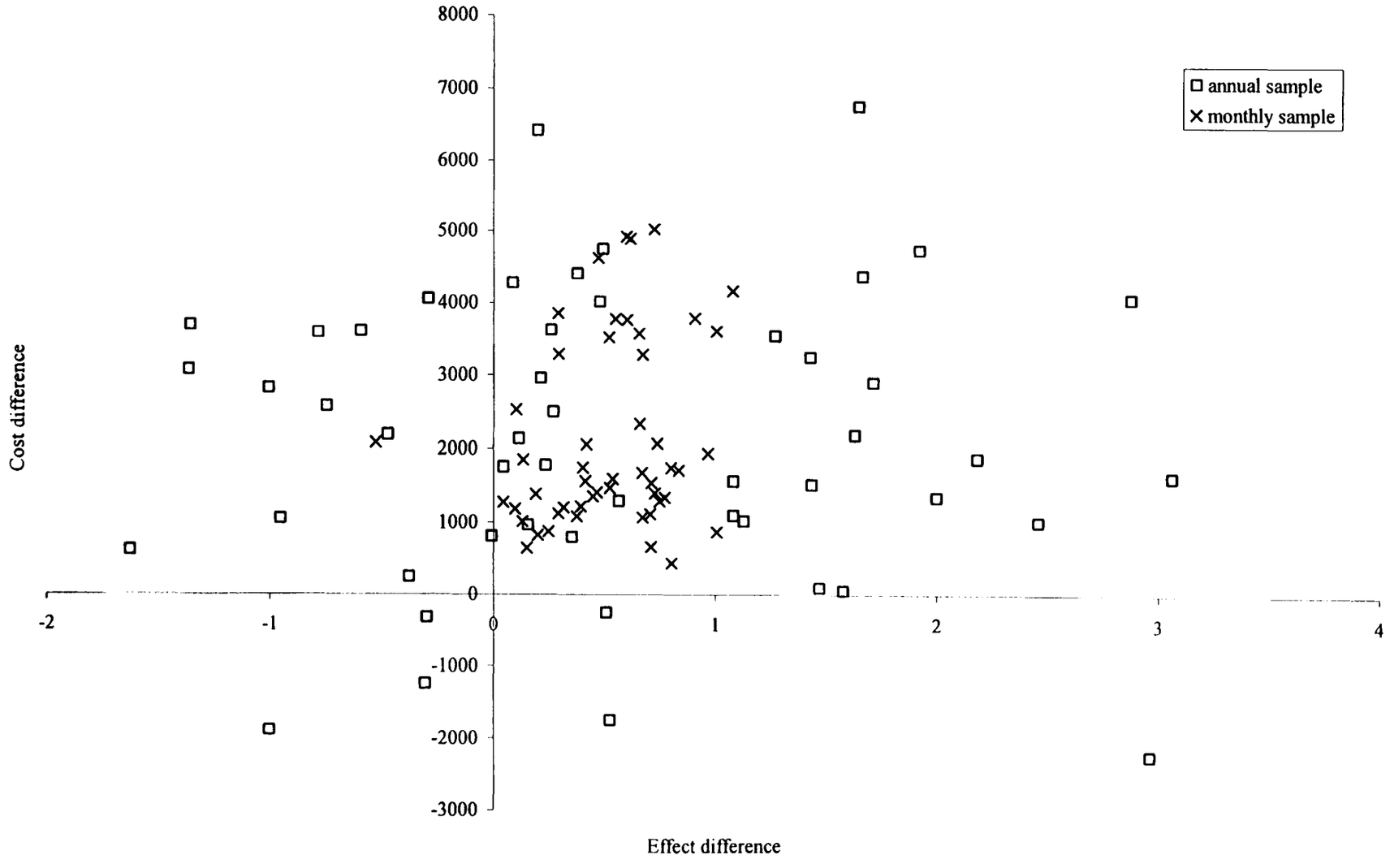
6.7 Methods of describing probabilistic 'length of time' input parameters

The second methodological issue addressed in this Chapter relates to the correct method of describing and sampling time dependent data in decision models. The parameters affected by this issue were the length of time patients remained in the state 'DFI' in both the Markov process and the DES model, and the time spent in remission following locoregional relapse in the DES model alone (this parameter was not modelled as being time dependent in the Markov process). The data informing these parameters were presented as the probability of leaving the respective states in successive years, in the form of survival curves.

Ideally, a probability distribution of individual survival curves (from separate sources) would be specified for each relevant parameter, which would be sampled as part of the second-order Monte Carlo simulation. Each survival curve could be weighted using the methods described in Chapter 3 and for each replication of the model a single survival curve could be sampled from a weighted dataset of survival curves. The advantage of this approach would be that the probabilities of experiencing events in successive years are consistent because they are derived from the same source. Unfortunately, it is rarely possible to assemble survival curves in the manner described above because different studies report differing lengths of follow-up and survival curves are rarely the same lengths.

It is normal practice to combine data from survival curves using one of the methods described in Appendix 4. The preferred method is the 'meta-analysis of failure time data' (MFD) [Earle and Wells, 2000], which involves pooling the number of patients at risk and the number of events in each time interval. Using formula 6.1 (pg120), both models interpolated the annual data describing the probability of leaving DFI to monthly probabilities. However, the DES model was set up to sample a separate probability of leaving the state every month, whereas the Markov process sampled one monthly value and applied it to all 12 months within a single year. Subsequently, in the DES model the annual probability of leaving DFI in any given run of the model was much closer to the mean of the distribution describing the probability of leaving DFI in any year. This was because the annual mean in the DES model is the mean of

Figure 6.4 Comparison of cost-effectiveness plots using alternative sampling strategies to describe the timing of an event



12 separate samples from the distribution, rather than a single sample as was the case in the Markov process. To investigate the effect of the alternative sampling strategies the programming code in the DES model was altered so that a single probability was sampled and applied to all 12 months in any given year. The model was then re-analysed and the outputs compared to original set of results. Figure 6.4 plots the results using the alternative sampling methods on a cost-effectiveness plane.

Visual inspection suggests that the choice of sampling the probabilities on a monthly or an annual basis had a substantial impact on the variation observed in the estimates of cost-effectiveness. Both models could have been constructed to sample annually or monthly, but the intuitive approach encouraged by the alternative modelling techniques led to this important divergence in the representation of the input parameter 'length of DFI'.

In order to represent the level of variation in the available data most accurately, the appropriate method of sampling should follow the intervals in the original data. The data used in the ABC models were presented as annual probabilities, which were transformed to monthly probabilities because a month was considered to be the appropriate interval within the models, which was based on clinical relevance. To be consistent with the available data, therefore, the models should sample a monthly probability and apply it to the 12 months within each year.

6.8 Conclusions

The focus of this Chapter has been the development of computer-based decision models. The construction of two alternative decision models were described - a Markov process and a DES model. Within the DES model, patients move through the model, experiencing events at any discrete time period after the previous event. The analysis of the model is triggered by the occurrence of an event, at which point the model asks what and when is the next event for this patient? This differs from a Markov process, which asks, what events are occurring within regular periods.

The procedure for building a Markov process using Excel spreadsheets, employing the Crystal Ball add-in [Crystal Ball, 2000], was described. The development of the

Markov process was generally straightforward. Formulae that described the proportion of patients moving between states were linked to cells representing the clinical parameter values, whilst the cost and utility value parameters were linked to the proportion of patients in each state in each time period.

The process of building the DES model required more complicated methods than the Markov process, though the use of specialised software provided assistance. The DES model was built on the basis of programming code that described what events occurred when. This also facilitated the collection of the required output data as patients left each state within the model. There was a steep learning curve associated with the building of the DES model. During the course of building the model many blind alleyways were encountered, and numerous improvements in the efficiency of the programming code were made over time. The time to develop future DES models will be significantly reduced, though the required time is unlikely ever to be reduced to the time to build a comparable Markov process. In addition, the analysis time for a DES model was substantially longer than the time taken to run a Markov process. Moreover, the analysis time not only included the final 'correct' experimentation with the models, but also the whole process of verification and validation, which is often more time consuming than the final process of experimentation.

Finally, two important issues that were raised during the building of the decision models were addressed. Within the experimentation of the DES model, the issue of an adequate number of patients to include in each (first-order) run of the model was investigated. After testing alternative sizes it was concluded that a sample of 10,000 patients was sufficient to provide stable estimates of the mean output values estimated for each set of input parameters. The second issue concerned the appropriate method for sampling time dependent events with decision models. Substantial differences were demonstrated between the alternative sampling strategies and it was concluded that to represent the level of variation in the available data most accurately, the appropriate method of sampling should follow the intervals in the original data. For example, if data are included in the model as monthly probabilities that have been interpolated from annual probabilities, as presented in the literature, a single monthly probability should sampled and applied to the 12 months within each year.

Chapter 7 Case study: verification and validation

7.1 Introduction

The previous three Chapters have applied the modelling process described in Chapter 3 up to the point at which the decision model has been implemented as two types of computer model, both as a Markov process and as a DES model. The final steps before experimentation can proceed, involve the verification of the models to ensure that they are internally consistent, and validation to check that the models' outputs are realistic. Respective methodologies for verification and validation were described in Chapter 3, the objective of this Chapter is the application and assessment of the proposed methods with respect to the models built to evaluate adjuvant therapies for early breast cancer.

7.2 Verification

Verification is the process of ensuring that the computer model accurately represents the conceptual model and the available data. The verification process consisted of three phases - verification of logic, sensitivity testing, and stress testing - as described in section 3.5.2. The following sections report the application of these methods to the case study evaluation.

7.2.1 Verification of logic

The verification of the models' logic ensures that the model is analysing the inputted data correctly. The tests involve entering simple combinations of input data for which

the expected output can be manually calculated, allowing the output of the model to be compared with the anticipated output. The model logic was verified in three categories. Firstly, the models' representations of the clinical parameters were assessed, followed by tests on the incorporation of the costs and utility values. Finally, the action of the model in discounting both the costs and the effects was verified. The following sections present the parameter values chosen to test model logic within these categories, and the resulting comparisons of the expected and actual outputs of the models.

7.2.1.1 Clinical parameters

The clinical parameters represent the probabilities of events occurring within the model, and their timing. The events described, in sequential order, were the:

- Incidence of toxicity (including effect on the completion of chemotherapy),
- Timing of relapse or death from the period of disease free interval (DFI),
- Type of event experienced following DFI,
- Timing of relapse or death from remission following locoregional relapse,
- Type of event following remission,
- Timing of death from metastatic relapse.

Selections of the parameter values inputted to the models and the expected and observed outputs from the models are shown in Table 7.1. The verification process for the clinical parameters produced model outputs that were broadly consistent with the expected results. Moving sequentially through the model, Test 1a describes the timing of events following the initial period of DFI. The inputted probabilities were chosen so that 25 per cent of the original patient cohort would relapse in each of four years – 1, 3, 5 and 7. A constant rate of occurrence was assumed. The results for both models show that the mean survival estimates were close to the expected output, though the results of both models slightly underestimated expected survival. The underestimated survival is due to the conversion of annual probabilities to monthly probabilities assuming a constant rate of events over the year in which an event occurs. The expected model outputs were estimated on the basis that events in a particular year occurred, on average, halfway through the year. However, assuming a constant rate of occurrence more patients experience events in the earlier months of any particular year. This issue is developed further in verification Test 1b below.

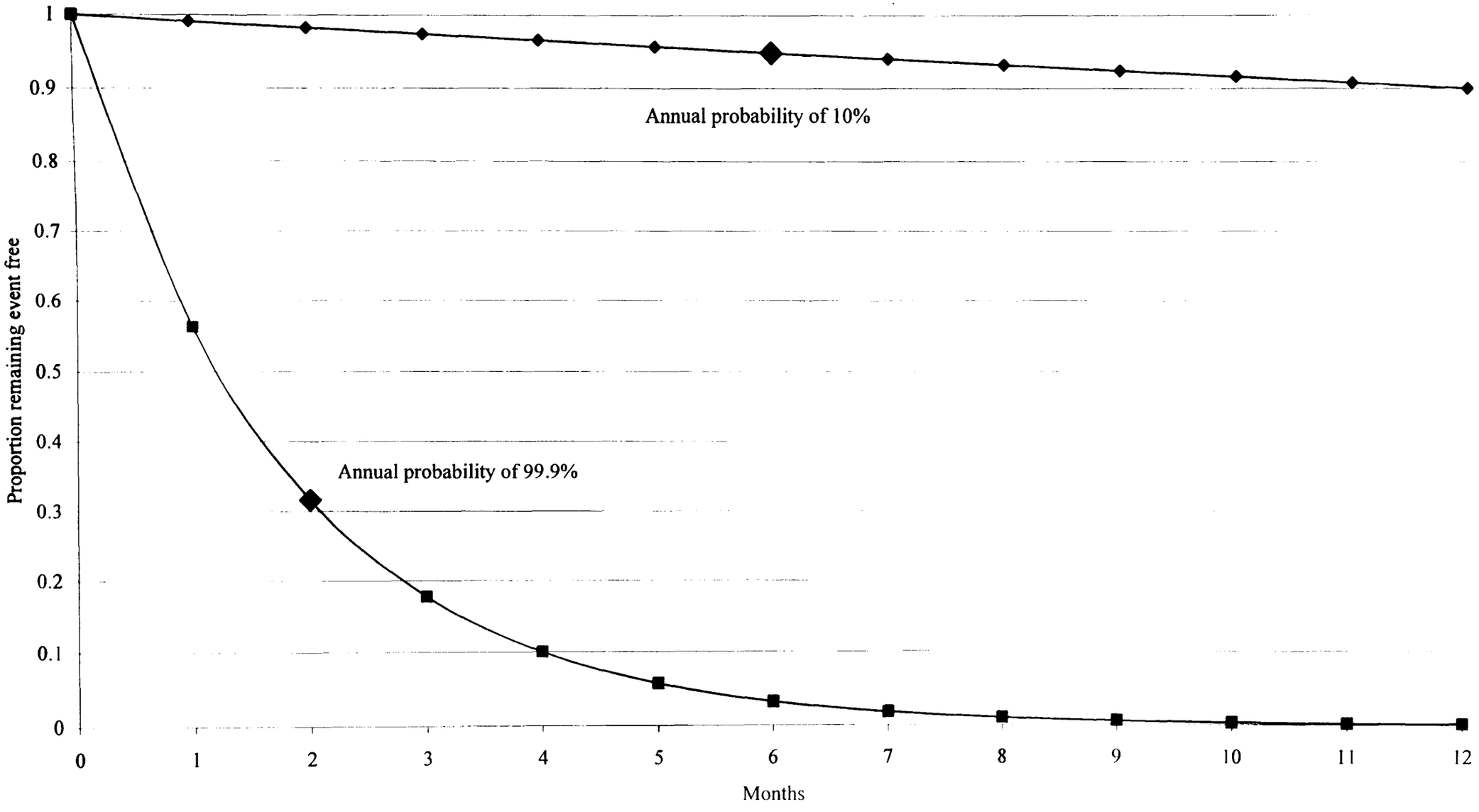
Table 7.1 Comparison of verification outputs for the DES and Markov model

Tests	Parameter categories	Parameter values	Expected Output*	Observed output*	
				DES	MP
1a	*Timing of event from DFI*				
	Pr(event following DFI in year 1)	0.25			
	Pr(event following DFI in year 3)	0.33			
	Pr(event following DFI in year 5)	0.50			
	Pr(event following DFI in year 7)	1.00			
	Pr(dned following DFI)	1.00			
	Output*		c.40.63	40.53	40.32
1b	Pr(event following DFI in year 3)	0.5			
	Pr(event following DFI in year 5)	1			
	Pr(dned following DFI)	1			
	Output*		c.39.25	38.90	37.92
2	*Type of event from DFI*				
	Pr(soft tissue metastases following DFI)	0.50			
	Pr(dned following DFI)	0.50			
	Pr(event following DFI in year 4)	1.00			
	Survival from soft tissue metastases	Empirical distribution, mean 31			
	Output*		c.52	52.17	50.83
3	*Type of event from remission*				
	Pr(bone metastases following remission)	0.50			
	Pr(dned following remission)	0.50			
	Pr(event following DFI in year 2)	1.00			
	Pr(locoregional relapse following DFI)	1.00			
	Pr(event following remission in year 8)	1.00			
	Survival from bone metastases	Normal distribution, mean 16.8			
	Output*		c.105.4	105.38	104.89

* survival in months.

Test 1b investigated further the effect of assuming a constant rate of events by stating that 50% of events following DFI occur in year 3, and the remainder in year 5. If all events occurred halfway through the year the expected survival would be 39.25 months, but both models underestimated survival by a larger amount compared with Test 1a. This discrepancy between the expected and observed output is not as important as it might appear. The illustrative examples assumed that all patients experienced an event in the same year, which maximised the impact of the constant rate assumption. In fact, the maximum annual probability of an event using the real data is approximately 10 per cent. Figure 7.1 demonstrates the reduced impact of the constant rate assumption when an annual probability of 10 per cent is converted to monthly probabilities compared to the situation when 99.9 per cent of patients experience an event within a single year. The large diamonds represent the month in which the median patient experiences an

Figure 7.1 Illustration of the impact of the assumption of a constant rate of occurrence over a year



event and shows that with an annual probability of 10% the median patient leaves DFI in month six.

Tests 2 and 3 specified the type of events patients' experienced following the DFI state and remission, respectively. The results revealed that the DES model was slightly more accurate than the Markov process. This was due to the DES model's more precise description of time on remission and survival from metastases (see section 9.2 for a full discussion)

7.2.1.2 Cost and utility parameters

The parameters describing the cost and utility effects of the various health states within the model could only be verified in connection with specified clinical parameter values. Within the decision models the cost and utility parameters were attached to the clinical parameters, so following the verification completed in the previous section the action of attaching costs and utility values to health states was relatively straightforward. However, the full benefit of such detailed verification became apparent during the testing of the utility values within the DES model. Informal testing of the model had been undertaken during the building process and prior to verification the model appeared to be producing consistent QALY outputs. However, during the verification process some subtle errors in the programming code were identified that led to the miscalculation of the associated QALYs in some of the later health states within the model (the remission and metastases states). Fortunately, the errors were rectified and both models passed this stage of the verification process.

7.2.1.3 Discount factors

The actions of the model with respect to discounting the estimated costs and effects are independent of the actions associated with moving patients through the model and attaching costs and utility values to the time spent in the model. To ensure that the discount factors were being applied properly within the models two sets of parameter values were used to compare expected and observed output. During the initial process of testing the discounting process the observed output from the DES model was consistently higher than the expected output. Within the DES model the process of

discounting costs and QALYs at the end of each state comprised the most complicated set of programming code (see section 6.3.2). Following examination of the programming code driving the discounting, an error in the programming code was discovered that discounted costs and QALYs at lower than appropriate discount rates. Once rectified the resulting estimates were extremely close to the expected outputs and the DES model was deemed to have passed the discounting stage of the verification process. The final input values, and expected and observed outputs are presented in Table 7.2.

Table 7.2 Parameter values used to verify the action of the discount factors for costs and effects*

Test	Input values†	Value	Expected output‡	Observed output	
				Markov process	DES model
1	pr(event following DFI in year 5)	1	2092,	2092,	2111,
	cost (surveillance)	£50	2.79	2.79	2.81
	utility(DFI)	0.8			
2	pr(event following DFI in year 5)	1	4098,	3894,	4059,
	pr(visceral metastases)	1	0.34	0.32	0.34
	survival from visceral metastases	1 year			
	cost(visceral metastases)	£500			
	utility(visceral metastases)	0.5			

* A discount factor of 10% was applied to both costs and effects to aid the calculation of the expected effects.
† All other cost and utility parameters are assumed to be equal to zero
‡ Outputs are presented as total costs and quality adjusted life years.

A more fundamental discovery was observed over the discounting in the Markov process. Test 1 in Table 7.2 shows that the observed output from the Markov process matched the expected output, but the observed outputs in Test 2 underestimated the expected values. The difference in Test 2 is due to the Markov processes' description of the length of time spent in states for which survival curves (in the case study model: time in remission) and set survival times (in the case study model: survival from metastases) are converted to constant probabilities. Taking the scenario described in Test 2, the input data states that patients survive 12 months from the point of being diagnosed with visceral metastases. However, the Markov process cannot remember when each patient enters the health state 'visceral metastases' and so applies a constant probability of dying to each patient remaining in the state. The problem with the conversion of survival curves and set survival times to constant probabilities is that the associated discounting will be erroneous. In Test 2, the correct procedure applies a discount rate of 0.792 $\left(1/1.10^{4}\right)$ to the cost of treating visceral metastases over the whole of the fifth year, which is the period over which all

patients experience metastases. However, in the Markov process a monthly probability of leaving the state was applied to all patients, so that some patients left the state after one month, whilst 14.7% and 5.4% of the original cohort were still in the state at the end of 24 and 36 months, respectively. In the absence of discounting the mean results would represent the data correctly, but the application of higher discount rates to those patients surviving beyond the set survival time results in underestimates of the true outputs for both costs and QALYs. This issue is raised in more detail in Chapter 9, where its' impact on the overall results of the case study is discussed.

7.2.2 Sensitivity testing

The second type of verification involves varying one parameter, whilst keeping all others fixed in order to ensure that the behaviour of the model is sensible [Bratley et al, 1987]. Sensitivity testing involved running the models stochastically using the probability distributions specified for each parameter and noting the variance around the model outputs. The model was re-analysed with point estimates describing the values of chosen parameters, rather than probability distributions. The resulting variation in the model output was then compared to the baseline run of the model.

Explicit predictions of the absolute magnitude of the effect of different parameters within the model were difficult, though it was mainly possible to predict the relative magnitudes of the individual parameter effects. Each included parameter was ranked in order of their expected relative effect on the variance of the baseline net benefits. The results were then compared to these *a priori* predictions. Table 7.3 illustrates the sensitivity testing of various parameters. It was anticipated that holding DFI probabilities constant would reduce variance by more than holding event types constant, and both would reduce variance more than holding a cost or utility value parameter constant. The comparison of the cost of bone metastases and the utility values associated with DFI was less certain and these parameters were used to check the similarity of the two modelling techniques, ie. whether both models reported the same ranking between these two parameters.

The results of the sensitivity testing are also presented in Table 7.3. The reduction in the variation around the net benefits was greatest assuming absolute certainty about the values of the 'length of DFI' parameters. Assuming certainty about the proportions of patients experiencing the alternative health states following DFI reduced overall variation by a larger amount than holding either the cost of bone metastases or the utility values associated with DFI static. No prediction was made about the ranking of the latter two parameters, but both models reported that the utility values associated with DFI had a larger impact on overall variation of the net benefits. The *a priori* assumptions about the ranking of the first two parameters were confirmed, and the Markov process and the DES model produced a similar order of ranking for the unknown rankings.

An additional result that required explanation was that the absolute values of the standard deviations between the Markov process and the DES model were very different. The reason for this difference was that only 5,000 patients were included in each run for the DES model in order to reduce the required running time for the sensitivity testing, which meant that first-order uncertainty was not adequately controlled (see section 6.5). However, the elimination of first-order uncertainty was not a necessary condition for the process of sensitivity testing, and the resulting data provided sufficient evidence for the sensitivity of the DES model to be verified.

Table 7.3 Examples of sensitivity testing

Parameters held constant	*A priori* ranking	Standard deviation of the net benefits (λ = £10,000)	
		DES	MP
Probability of event following DFI in all years in patients receiving tamoxifen+ chemotherapy	1	3718	2020
Type of event following DFI for patients receiving tamoxifen+chemotherapy	2	4333	2660
Cost of bone metastases	3/4?	5420	2800
Utility values for DFI	3/4?	5028	2759

7.2.3 Stress testing

The scope for stress testing, which tests that the model does not run if infeasible values are inputted, was limited. For example, possible areas for such testing included the assignment of negative values for probabilities. However, the models were built to transform negative values to a value of zero because input distributions

describing probabilities could include negative values, e.g. normal distributions. The only area in which stress testing was possible covered the inputting of negative cost values. Originally, no mechanisms had been incorporated to highlight such errors, but it was recognised at this stage that such a device could prevent potentially serious errors and that such mechanisms were relatively simple to create. Within the DES model a single line of code was placed after the code that sampled each cost and utility value at the beginning of each trial. The code stopped the model if a negative value was sampled from any of the input distributions. In the Markov process, using the facilities in the Crystal Ball add-in, forecast windows were set up that highlighted the occurrence of a negative value at the end of each trial. Both mechanisms captured deliberate negative values accurately.

7.3 Validation

The objective of validation is to legitimise the output of a decision model by demonstrating that the model provides a credible representation of reality. As described in Chapter 3 the precision of the validation process for most economic HTA models is not likely to be high as the process of validation is hampered by a lack of credible 'real' world data. However, a number of economic analyses involving adjuvant therapies for early breast cancer were identified. These studies provided some indication of the case study models' comparability with previous research, which is one method for validation that has been suggested (see section 3.5.3). Details of the identified studies were presented in section 4.2.1.

From the review of previous economic analyses (see section 4.2.1), only four studies were felt to provide relevant sources of comparison. Indeed, data describing survival, rather than economic outputs, were employed from one of the economic studies because the cost data was so weak [Messori et al, 1996]. The remaining studies were excluded for various reasons. One of the cost analyses was omitted from the validation process due to the large number of uncertainties around the estimated costs [Legorreta et al, 1996], whilst the Australian economic evaluation was excluded because the patient group was ill-defined [Glasziou and Haas, 1994]. The study only available as a conference abstract did not define their patient group and was also excluded from the validation process [Selke et al, 1998]. In addition to the data

derived from economic studies, survival estimates from the models were also validated against estimates from clinical papers. In the case study models, survival was determined by combinations of the length of DFI, the length of remission and survival from the three metastatic sites. Straight survival data were not inputted into the model, so any available survival data could be compared to the life years predicted by the model to validate this aspect of the model.

The most reliable data of this form was accessed from a series of meta-analyses prepared by the early breast cancer trialists group [EBCTCG, 1992; EBCTCG, 1998a; EBCTCG, 1998b]. The internal reliability of the meta-analytic data was high but several problems remained. Firstly, survival data were only presented for the first 10 years from the point of diagnosis so only the first 10 years of the ABC models could be validated. Secondly, survival curves were not presented for combination therapies, such as tamoxifen and chemotherapy and the survival data presented included all patients receiving the therapy of interest compared to all patients not receiving the therapy. For example, comparisons were stated as 'tamoxifen versus no tamoxifen', whereby trials reporting tamoxifen and chemotherapy versus chemotherapy were included alongside trials comparing tamoxifen versus no to adjuvant therapy. More usefully, the meta-analyses presented reduction ratios for annual rates of mortality for a wide range of treatment comparisons, including tamoxifen and chemotherapy versus no treatment, and tamoxifen versus no treatment. Survival for these two patient groups was estimated by applying the appropriate reduction ratios to survival estimates for patients receiving no adjuvant therapy, which was informed using data from two identified studies that reported survival up to 12 years from diagnosis [Castiglione-Gertsch et al, 1994; Cummings et al, 1993]. The mean survival times are presented in Table 7.4, the method of calculation was as follows:

1. Multiply the proportion of patients dying in each year by the year of death (-0.5);
2. Multiply the proportion of patients remaining alive at the end of the period by 12;
3. Sum the annual totals to estimate the mean number of years alive.

Table 7.4 Estimation of mean 12-year survival for alternative adjuvant therapy options

Year	No adjuvant therapy			Tamoxifen alone (annual mortality reduction 20% vs. no treatment)			Tamoxifen+chemotherapy (annual mortality reduction 30% vs. no treatment)		
	Prop. Surviving	Prop. dying	Annual survival estimate	Prop. surviving	Prop. dying	Annual survival estimate	Prop. Surviving	Prop. dying	Annual survival estimate
1	0.96	0.04	0.02	0.97	0.03	0.02	0.97	0.03	0.01
2	0.89	0.07	0.10	0.91	0.06	0.08	0.92	0.05	0.08
3	0.79	0.10	0.26	0.83	0.09	0.21	0.84	0.08	0.21
4	0.67	0.11	0.40	0.73	0.10	0.33	0.74	0.09	0.33
5	0.62	0.06	0.27	0.68	0.05	0.23	0.69	0.06	0.26
6	0.57	0.05	0.26	0.64	0.04	0.23	0.64	0.04	0.23
7	0.49	0.08	0.51	0.57	0.07	0.47	0.58	0.07	0.45
8	0.44	0.05	0.35	0.52	0.04	0.32	0.53	0.05	0.34
9	0.39	0.05	0.46	0.47	0.05	0.44	0.48	0.05	0.43
10	0.34	0.05	0.45	0.43	0.05	0.44	0.43	0.05	0.45
11	0.29	0.05	0.51	0.38	0.05	0.50	0.38	0.05	0.50
12	0.29	0.01	0.07	0.37	0.01	0.08	0.37	0.01	0.13
		0.29	3.42		0.37	4.45		0.37	4.46
Mean survival			7.08			7.81			7.88

Table 7.5 Validation data and comparison results

Test	Output	Validation estimate	MP estimates		DES estimates	
			Original	Revised	Original	Revised
1	Stage II, 4-year costs*	2417	4610	4181	4576	4161
2	Stage II, lifetime costs*	9974	10154	9058	9895	8942
3	QALYs in patients receiving tamoxifen	6.47	8.26	8.26	8.4	8.4
	QALYs in patients receiving tamoxifen+chemotherapy	6.59	8.44	8.44	8.57	8.57
	Incremental cost per life year	28537	5013	5254	5748	6012
4	Survival in patients receiving no adjuvant therapy	13.17	-	-	-	-
	Survival in patients receiving chemotherapy	16.74	-	-	-	-
	Survival in patients receiving tamoxifen	-	14.90	14.90	15.15	15.15
	Survival in patients receiving tamoxifen+chemotherapy	-	15.70	15.70	15.91	15.91
5	Survival in patients receiving tamoxifen†	7.81	8.58	8.58	8.43	8.43
	Survival in patients receiving tamoxifen+chemotherapy†	7.88	8.71	8.71	8.57	8.57

Test sources: 1 [Wolstenholme and Whynes, 1998], 2 [Will et al, 1998], 3 [Hillner and Smith, 1992], 4 [Messori et al, 1996], 5 [Castiglione-Gertsch et al, 1994; Cummings et al, 1993].
The estimates presented in the foreign studies were converted to UK sterling using the health purchasing power parity, all costs were uprated to year 2000 values using the health services index.
* revised to subtract costs associated with diagnosis and primary surgery.
† mean survival at 12 years from diagnosis.

The validation data and results are presented in Table 7.5, discount rates and length of follow-up were altered within the ABC models to match those employed in the identified studies. The similar estimates produced by the Markov process and the DES model implied that the remaining variation was an artefact of sampling variation. The original estimates refer to the estimates derived from the models without any modifications, whilst the revised estimates resulted from the changes made as a result of the validation process, as described below.

Comparing the models' original outputs with each other it was apparent that the Markov process estimated lower outputs than the DES model, which was caused by differences in the estimated time spent in remission following a locoregional relapse. The identified data describing this parameter were in the form of disease free survival (DFS) curves, in which a proportion of patients remained alive at the end of the follow-up periods. It was not possible, therefore, to estimate mean DFS precisely. The DES model facilitated a more accurate representation as the survival curve data could be applied to each patient over the initial 11 years in the state, after which time age-specific mortality rates in the general population were applied to those patients remaining disease-free. Such methods could not be applied to patients within the Markov process because the model could not note when individual patients entered the state and a constant probability rate had to be applied to all patients within the state. Time dependency within a Markov process may only be applied from the starting point in the model and remission following a locoregional relapse was not a starting state. Initially, the median DFS was employed as a proxy for mean DFS but the distribution of DFS was highly skewed and the median greatly underestimated the mean. On the basis of the validation process mean DFS were estimated on the basis of a common cut-off length of remission for those patients remaining disease-free at the end of each studies follow-up.

In the comparisons of the models' outputs with published sources Test 1 showed that the ABC models estimated far higher costs than the retrospective UK-based study [Wolstenholme and Whynes, 1998]. No obvious data-related differences were apparent, no data on the resource use or unit costs associated with different events were presented, or the proportion of patients undergoing any of the included events, though it was noted that the cost estimates were based on only 13 patients. The

immediate conclusion of exaggerated cost estimates was tempered by Test 2, in which a Canadian study predicted lifetime treatment costs that were roughly similar to those produced by both case study models [Will et al, 1998]. However, the Canadian study included costs associated with the administration of radiotherapy following primary surgery in the Canadian study. Without such costs the Canadian study would have produced substantially lower cost estimates.

The conclusion drawn from this section of the validation process was that some of the costs assumed in the case study models were too high. It appeared that the costs associated with both interventions might be overestimated, so the input data describing the cost of relapses were critically reviewed. The initial assumptions regarding the treatment of relapse are described in section 5.4.4.4. These costs had been estimated using mainly published guidelines on the course of treatment for such patients. Empirical data were only used to inform the frequency of hospital contact [Hurley SF et al, 1992]. A major component of the assumed cost of treating metastatic relapses covered the cost of chemotherapy, which had been assumed to continue from the point of diagnosis to death. It was felt that this area was the most likely to be exaggerated. To improve the understanding of the administration of chemotherapy in practice, data from eight sets of deceased patient notes on treatment following diagnosis with recurrent breast cancer were inspected. The review of the notes indicated that the assumptions made with respect to the treatment of locoregional relapse were broadly correct, but that the assumed use of chemotherapy in patients with metastatic relapse was certainly over-estimated. The estimation of the monthly costs associated with metastatic disease was revised in line with reduced use of chemotherapy. The details, and the resulting cost estimates were described in Chapter 5. The revised validation estimates are presented alongside the original estimates in Table 7.4. The revised analysis did not aim to match the validation estimates due to the inconsistencies in these data discussed above, but rather to make the ABC estimates appear more realistic.

In Test 3, the ABC models estimated higher estimates of life years and QALYs gained than estimated by Hillner and Smith [1992]. This difference was possibly due to the assumptions made with respect to the continued occurrence of relapses in both patient groups. The current study used survival curves derived from the literature

review in both models that tailed off to a relapse rate of 0 from year 15 onwards, whereas the US group assumed a constant probability of recurrence for the duration of the model (4% in the baseline). In addition, the incremental cost per lifeyear gained was significantly lower in the ABC models than in the Markov process used by Hillner and Smith. The QALY differences between the treatment groups were similar, which indicated that the difference in costs between the treatment groups was larger in the US-based study. Looking at the inputted costs it was clear that the costs associated with the tamoxifen+chemotherapy treatment group were higher in the published study. For example, the average cost per cycle of chemotherapy was estimated at £500 (£285 in the case study evaluation), whilst the average cost of major toxicity was inputted as almost £5000 (c.£2000 in the case study evaluation). No toxicity was assumed for the tamoxifen alone arm in the US study. Allowing for the differences between the two studies it was difficult to refute the validity of the ABC models on the basis of these economic comparisons.

Test 4 compared the ABC models' outputs with the full survival estimated by Messori *et al* [1996], in which the ABC model appears to slightly underestimate the published survival times. However, the comparisons undertaken in Test 5 showed the ABC models overestimating 12-year survival [Castiglione-Gertsch et al, 1994; Cummings et al, 1993]. The mid-placing of the ABC estimates is an encouraging validation result.

An additional issue raised in the validation process concerned the need to assess model outputs for periods less than the full running time of the model, i.e. for 4 and 12 years in order to provide an appropriate basis for comparison with the identified data used to validate the model. After specifying the possible avenues of validation, it soon became apparent that significant modifications to the programming code within the DES model were necessary to facilitate the collection of output data for periods less than the full lifetime of patients. This process is described in more detail in the comparison of Markov process and DES models (Chapter 9).

7.4 Conclusions

This Chapter described the final stages in the implementation of the decision models prior to actual experimentation. Verification and validation are necessary elements of the modelling process as they give the model user confidence in the results emanating from the experimentation phase.

The process of verification applied to the case study evaluation was adapted from methods described in the general operations research literature [Bratley et al, 1987]. Three categories of verification covered different aspects of the model's operations. Firstly, the verification of logic was dependent on the specification of simple input data for which the expected output could be determined manually, which ensured that the internal mechanisms of the models were working correctly. The explicit presentation of three classes of logic testing - clinical parameters, costs and utility values, and discounting - provided ample evidence of the analyst's attention to detail.

The verification of logic highlighted several issues of interest relating to the implications of formatting data to populate the models and to differences in the mechanisms between Markov processes and DES models. Firstly, the implications of assuming a constant rate of occurrence when converting data describing annual probabilities of experiencing an event to monthly probabilities were identified. To test the logic of the models large annual probabilities were inputted and the models' outputs were consistently lower than the expected results, which had been estimated assuming that the mean event within a year would occur at six months. On closer inspection it transpired that this effect was not significant for the annual probabilities inputted to the ABC models. However, such effects may be important in other treatment areas. Secondly, during the logic testing of discounting within the two models the Markov process produced lower than expected estimates for scenarios that included time within a metastatic site. This result was due to the necessary conversion of set survival times to constant probabilities that spread the occurrence of events over a longer period.

The second form of verification tested the sensitivity of the models by comparing expected and observed effects on the variation in the models' outputs caused by the

uncertainty in the values of different input parameters. The sensitivity tests provided an alternative method of verification that complimented the verification of logic. A particular advantage of sensitivity testing was the improved understanding about the functioning of the models that it provided. In the context of building two separate decision models to evaluate the same interventions, it also confirmed that the two decision models were producing consistent outputs. The final form of verification, stress testing, provided assurance that the models would recognise nonsensical input data, alerting the analyst to data entry errors.

The technique adopted in this thesis to validate the case study models identified a wide range of sources of data that could be compared to any of the outputs from the models. The corresponding outputs from the models were compared to the identified data and reasons for any differences between the compared outputs were sought in the context of methodological and data differences between the case study evaluation and the identified studies. The process proved to be beneficial as it highlighted an area of the model that appeared to be inconsistent with the published data. Further data were collected on the costs associated with breast cancer relapses and revisions were made to some of the cost input data, which improved the comparability of the estimated data with the published sources.

The process of validating separate models that were built to evaluate the same interventions also produced benefits, as demonstrated by the identification of the inaccurate use of median estimates of disease-free survival (DFS) as proxies for mean DFS in the Markov process. Better estimates of mean DFS were applied to the model, which improved the comparability of the different models. The methodological issues raised in this Chapter are potentially important and are discussed in more detail in Chapter 9, in the context of the experimentation with the case study models. The full and explicit process for verifying and validating the case study models identified a range of errors in the construction of the models and inaccuracies in the representation of the identified data, it also provided great insight into the functioning of the models.

Chapter 8 Case study: results of experimentation

8.1 Introduction

The results of the experimentation undertaken with the two decision models are presented in this Chapter. There are two main sections. Firstly, cost-effectiveness results are presented in order to inform the allocation of resources between the two therapy options on the basis of the available information. For each decision model, three sets of results were derived relating to the alternative methods for assembling input distributions for stochastic analyses (see section 3.4.3.2). To recap, the three methods specified theoretical distributions for alternative categories of input parameters (using the empirical data to inform the distribution parameters), used weighted empirical data directly, and specified fitted distributions based on the weighted empirical data. The results from the two models are presented in separate sections, within each section the results derived from the three methods for specifying probability distributions are presented.

The second method for experimentation analyses the value of information (VoI), where the objective was the estimation of the value of securing more accurate data on the values of the input parameters. The presentation of the results includes a commentary on the stages of experimentation, and the respective results from the two models are presented together.

The policy implications of the differential results are discussed in the final section of this Chapter, whilst the explanation for the differences is presented in the following

Chapter. Discount rates of 6% for resources and 1.5% for QALYs were applied to all the analyses undertaken, in accordance with Treasury guidelines [HM Treasury, 1997]. Life years were not discounted as they were not incorporated into a cost-effectiveness ratio.

8.2 Results from alternative methods for assembling input distributions

In Chapter 3, three alternative methods for specifying input distributions were described. Two methods involved assembling weighted datasets to describe the available data, the datasets were either inputted into the model directly or inputted into statistical fitting software that fitted the data to the best fitting distributions, which were then entered into the model. The third method employed distributions that were theoretically derived according to the characteristics of the alternative categories of parameters within the model. The available data were used to estimate the appropriate parameters for the chosen distributions. The results produced by the two models using the three methods for assembling the input parameter distributions are presented in the following sections.

8.2.1 Results from the case study Markov process

The analysis for each method for specifying input distributions involved a total of 10,000 runs, each sampling an alternative set of parameter values from the relevant input distributions. Table 8.1 presents the mean estimates of costs, QALYs and life years for each approach to assigning input distributions. Employing the mean of the three data methods as the baseline point estimate for each output parameter, the Markov process estimated that the lifetime treatment costs between the two therapies differed by £2,130 and the difference in QALYs was 0.53. The mean life expectancy for a patient receiving tamoxifen and chemotherapy was 15.70 years, and for tamoxifen alone it was 14.90 years. The mean incremental cost-effectiveness ratio (ICER) was calculated using the 'ratio of the means', rather than the 'mean ratios' [Stinnett and Paltiel, 1997]. Based on these estimates the ICER was £3,988.

Table 8.1 Mean costs, QALYs and life years associated with the two adjuvant therapies estimated using three alternative methods of specifying input distributions

	Costs	QALYs	Life years
Tamoxifen and chemotherapy			
Theoretical	£8,718	12.01	15.74
Fitted	£8,893	11.56	15.63
Empirical	£8,862	11.62	15.72
Mean*	£8,824	11.73	15.70
Tamoxifen alone			
Theoretical	£6,709	11.41	14.86
Fitted	£6,721	11.07	14.90
Empirical	£6,653	11.11	14.94
Mean*	£6,694	11.20	14.90

* Mean values estimated by dividing sum of theoretical, fitted and empirical estimates by three.

The extreme closeness of the models' outputs using the fitted distributions and the empirical data directly was as expected and provided further proof of the correct workings of the models, though the results varied little between all three input data analysis methods. The mean QALYs associated with tamoxifen and chemotherapy displayed the largest percentage range between the three (taking the mean as the base) of just under 4%, the corresponding range for tamoxifen alone was just over 3%. None of the life year estimates varied by more than 1%, whilst the estimates for costs varied by 1% and 2% for tamoxifen alone and tamoxifen and chemotherapy, respectively. Table 8.2 presents the mean estimates of the ICERs for tamoxifen and chemotherapy versus tamoxifen alone, as well as details of the percentiles derived from the different methods of specifying probability distributions. The largest ICER was derived from the fitted distributions data, which estimates an ICER of £4,473, whilst the data based on theoretical distributions estimated the lowest ICER of £3,334 - a range of almost 30% around the mean ICER.

Table 8.2 Incremental cost-effectiveness ratios estimated using three alternative methods of specifying input distributions

Method	Mean ICER	2.5th percentile	97.5th percentile	95th percentile
Theoretical	£3,334	£452	Tamoxifen dominates	Tamoxifen dominates
Fitted	£4,473	£248	Tamoxifen dominates	Tamoxifen dominates
Empirical	£4,295	£314	Tamoxifen dominates	Tamoxifen dominates

Figure 8.1a Observations of the cost and QALY differences plotted on the cost-effectiveness plane derived from the case study Markov process using theoretically specified input distributions

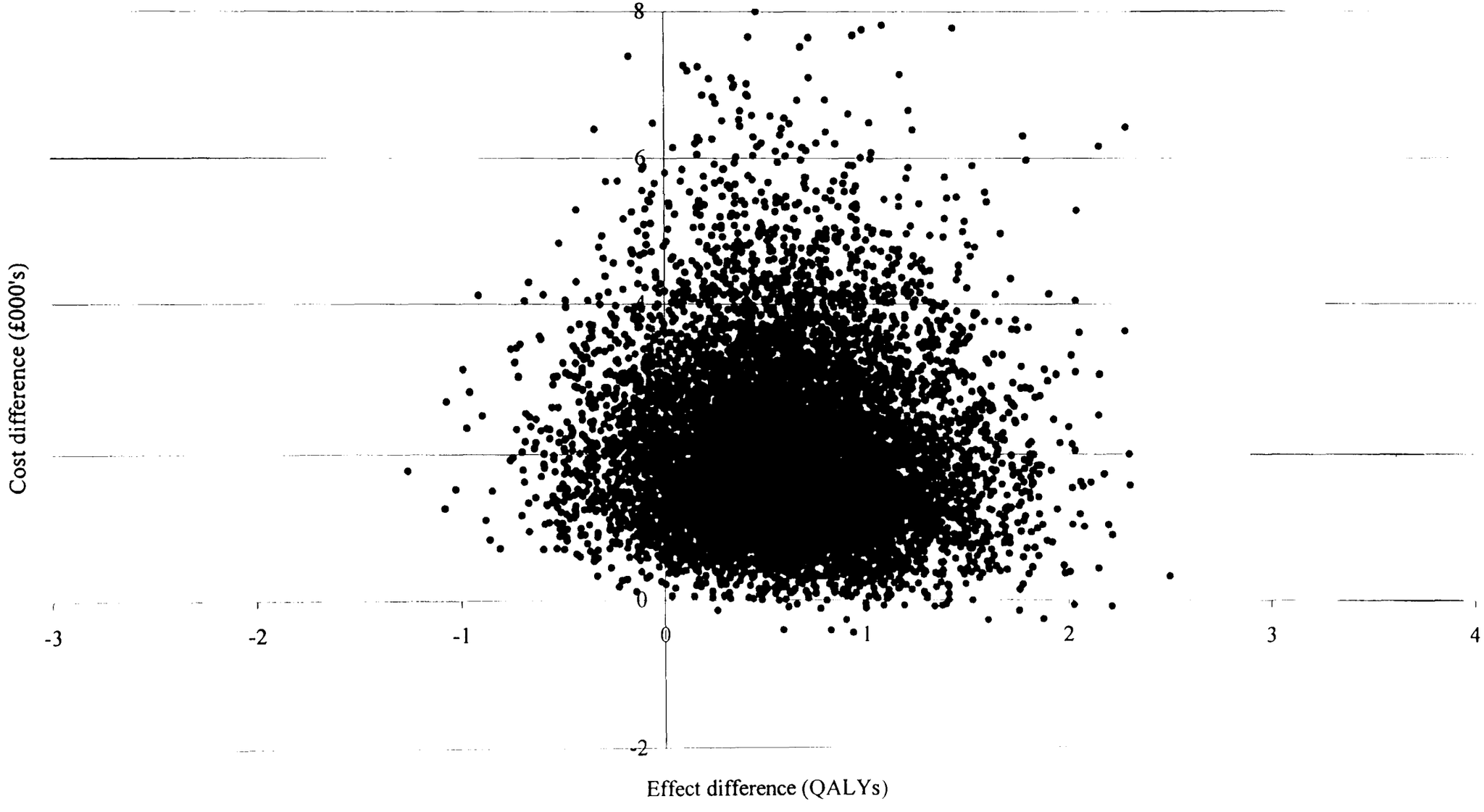


Figure 8.1b Observations of the cost and QALY differences plotted on the cost-effectiveness plane derived from the case study Markov process using distributions fitted to the weighted empirical data

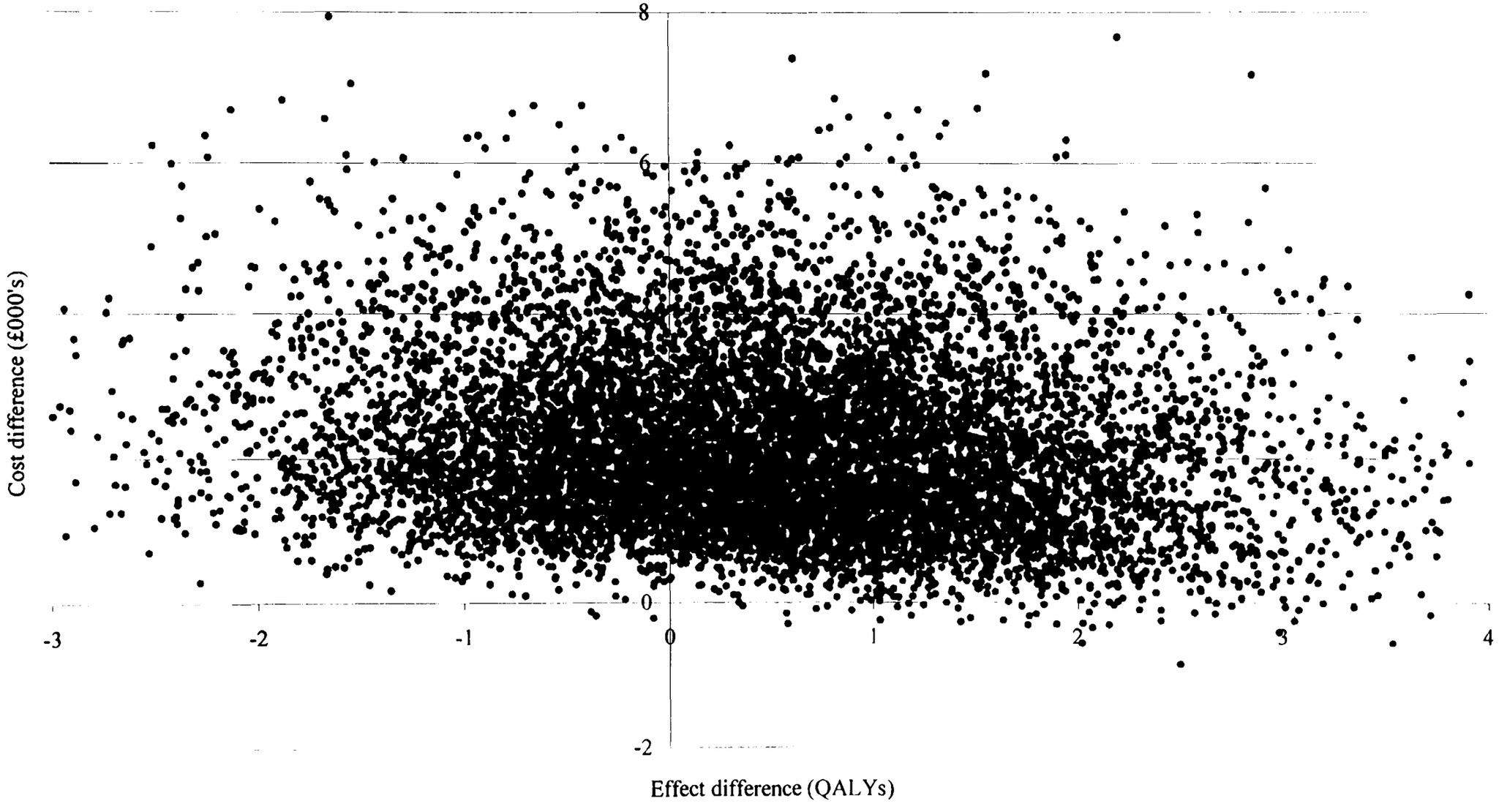


Figure 8.1c Observations of the cost and QALY differences plotted on the cost-effectiveness plane derived from the case study Markov process using weighted empirical data directly

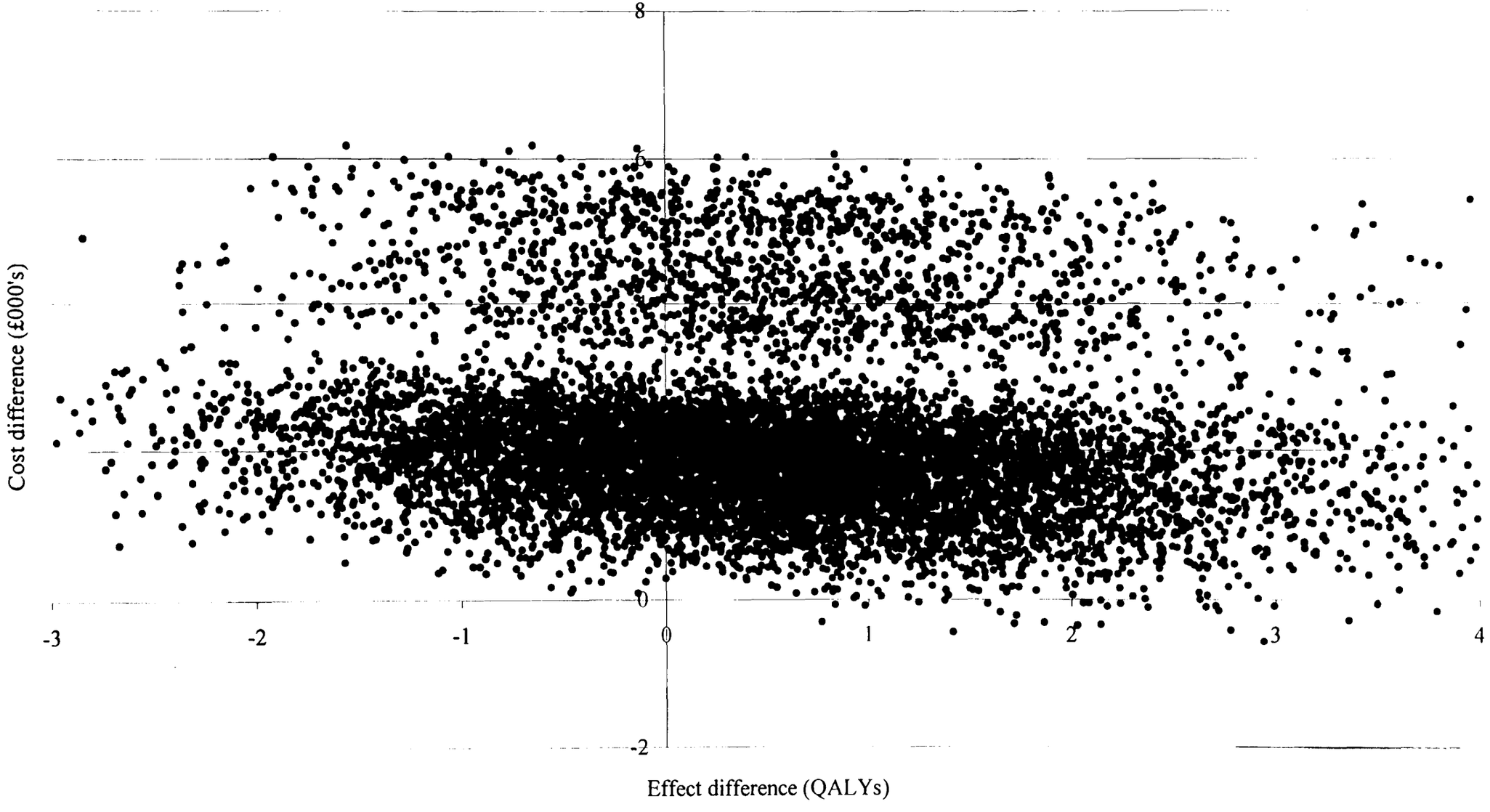
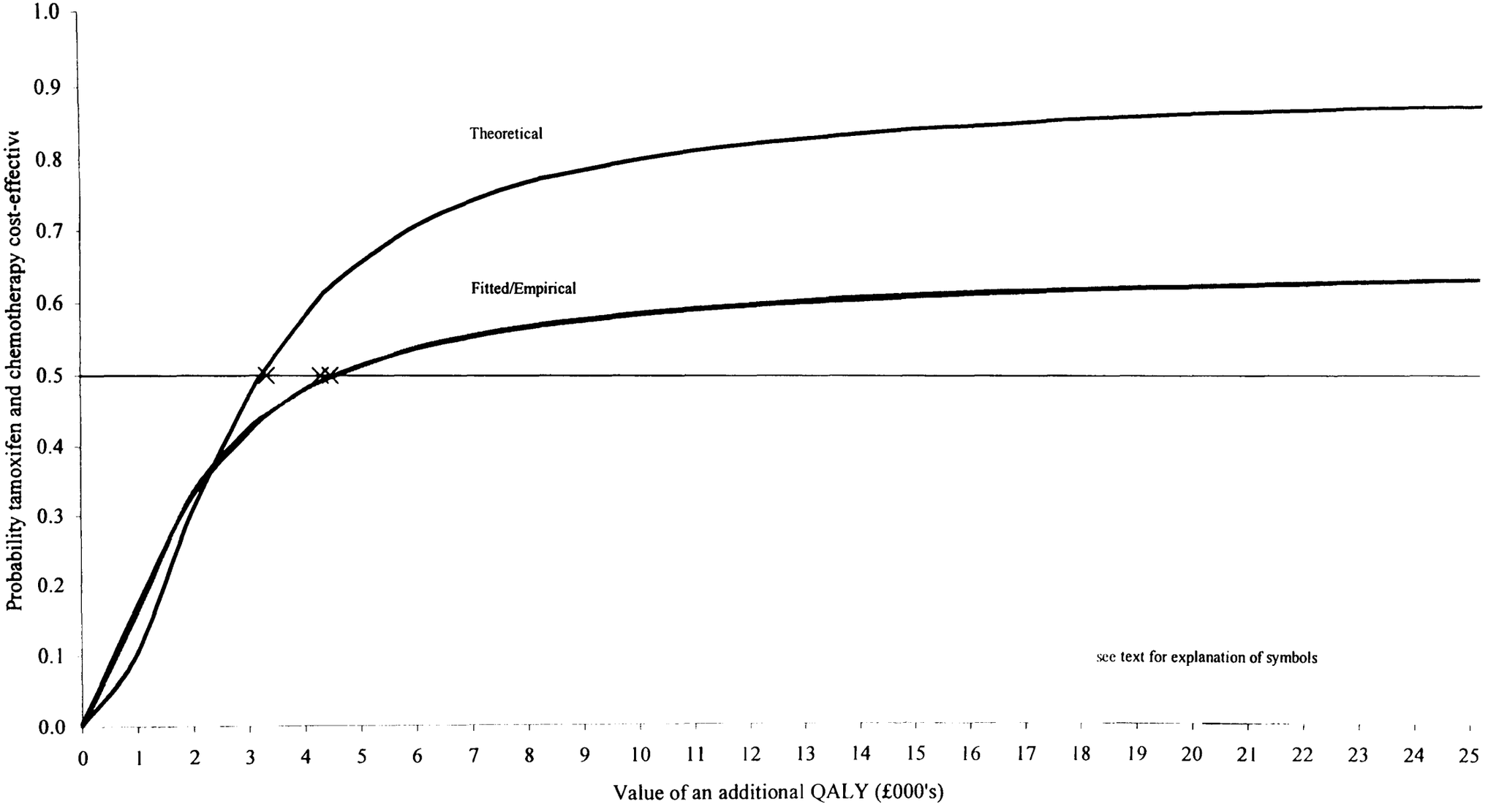


Figure 8.2 Cost-effectiveness acceptability curves derived from the case study Markov process, using the alternative methods for specifying input distributions



The observations of costs and QALYs for each data method have been plotted separately on cost-effectiveness planes in Figures 8.1a, 8.1b and 8.1c. These plots show that the majority of observations were located in the north east quadrant, which indicates increased costs and effects associated with tamoxifen and chemotherapy. A sizeable minority of the plots derived from specifying fitted distributions and using the empirical data directly indicated increased costs and lower effects, whilst very few observations revealed lower costs associated with tamoxifen and chemotherapy. The 'credible interval' percentiles, presented in Table 8.2, show that all the 2.5th percentiles were under £500. At the upper end, the 95th percentiles showed tamoxifen alone dominating tamoxifen and chemotherapy according to all three data methods. The threshold percentiles at which tamoxifen alone no longer dominated were 9.2% for the theoretical distributions, and 34% for both the fitted distributions and empirical data.

From the vertical axis, the first, second, and third crosses presented in Figure 8.2 represent the values of a QALY at which the sum of the incremental net benefits became positive for the theoretical, empirical, and fitted distribution data sources, respectively (see section 3.6.2, pg60). From Figure 8.2 it can be seen that the sum of the net benefits became positive at values of an additional QALY that were close to the point at which the probability of tamoxifen and chemotherapy being the cost-effective therapy reached 0.5. However, the slight skewness apparent in each of the distributions was in opposite directions for the data derived from weighted datasets compared to the theoretical distributions. For the theoretical distributions-based data the probability of being cost-effective reached 0.5 (median) at a lower value of an additional QALY than the sum of the net benefits reached zero (mean), which meant that the distribution was left-skewed. Basing the resource allocation decision on the probability of tamoxifen and chemotherapy being cost-effective may lead to the inappropriate allocation of resources to chemotherapy. Conversely, using the other methods of specifying input distributions, which produced right-skewed distributions of net benefits, resource allocations based on the probability of tamoxifen and chemotherapy being cost-effective may wrongly withhold resources from chemotherapy.

The remaining portion of the CEAc curve shows that the probability of positive net benefits associated with the empirical data directly and the fitted distributions did not rise above 0.67. The corresponding probability for the theoretical probability

distributions was around 0.9. These results follow from the proportion of observations in which tamoxifen alone was dominant.

8.2.2 Results from the case study DES model

The results produced by the DES model using the three methods of specifying probability distributions are presented in this section. The following section compares the results from the Markov process and the DES model.

2,500 runs for each method of specifying input distributions informed the analysis of the DES model. For each run 10,000 patients receiving tamoxifen and chemotherapy, and tamoxifen alone were sent through the model. Table 8.3 presents the mean results for each approach to assigning input distributions. Using the mean of the three data methods to represent the baseline point estimate of cost-effectiveness the lifetime treatment costs differed between the two therapies by £2,176 and the difference in QALYs was 0.51. Based on these estimates the 'ratio of means' ICER was £4,226. The mean life expectancy for a patient receiving tamoxifen and chemotherapy was 15.91 years, whilst a patient receiving tamoxifen alone could expect to live for 15.15 years. The aggregate costs, life years and QALYs, estimated by the three data methods, for the alternative therapies varied by less than 5% from the mean.

Table 8.3 Mean costs, QALYs and life years from the DES model associated with the two adjuvant therapies estimated using three alternative methods of specifying input distributions

	Costs	QALYs	Life years
Tamoxifen and chemotherapy			
Theoretical	£9,146	12.14	16.01
Fitted	£9,439	11.63	15.82
Empirical	£9,473	11.71	15.91
Mean*	£9,353	11.83	15.91
Tamoxifen alone			
Theoretical	£7,115	11.56	15.16
Fitted	£7,182	11.17	15.13
Empirical	£7,233	11.22	15.18
Mean*	£7,177	11.31	15.15

* Mean values estimated by dividing sum of theoretical, fitted and empirical estimates by three.

The mean ICERs estimated using the different data input analysis methods are presented in Table 8.4. The largest ICER was derived using the fitted distributions data (£4,869),

whilst the data based on theoretical distributions estimated the lowest ICER (£3,483) - a range of almost 33% around the mean ICER.

Table 8.4 Incremental cost-effectiveness ratios from the DES model estimated using three alternative methods of specifying input distributions

Method	Mean ICER	2.5th percentile	97.5th percentile	95th percentile
Theoretical	£3,483	£584	Tamoxifen dominates	Tamoxifen dominates
Fitted	£4,869	£288	Tamoxifen dominates	Tamoxifen dominates
Empirical	£4,498	£325	Tamoxifen dominates	Tamoxifen dominates

The differences in the observations of costs and QALYs for each data method are plotted on cost-effectiveness planes in Figures 8.3a, 8.3b, and 8.3c. The majority of observations were located in the north east quadrant for each data method, though a significant minority of observations derived from the fitted distributions and the empirical data direct were located in the north west quadrant, which indicated tamoxifen alone dominating tamoxifen and chemotherapy. Only a few observations portrayed tamoxifen and chemotherapy as the lower cost option. The 'credible interval' percentiles presented in Table 8.4, show that all the 2.5th percentiles were under £600. At the upper end, the 95th percentiles showed tamoxifen alone dominating tamoxifen and chemotherapy according to all three data methods. The threshold percentiles, at which tamoxifen alone no longer dominated were 10.8%, 33.6% and 34.6% for the theoretical distributions, fitted distributions and empirical data, respectively.

The CEAc curves derived from the DES model are presented in Figure 8.4. The probability of tamoxifen and chemotherapy being cost-effectiveness passed the 0.5 probability threshold at a value of a QALY of around £3,400, using the theoretically specified distributions, and between £4,500 and £5,000 using the alternative data methods. The distributions of the net benefits appeared to be slightly skewed for two of the data methods - left skewed for the theoretical distributions, and right skewed for the fitted distributions. The distribution of net benefits derived from the direct use of the empirical data displayed an extremely close fit to a normal distribution. The remaining portion of the CEAc curve shows that the probability of positive net benefits associated with all the data methods levelled out at around 0.88 based on the theoretical distributions, and 0.65 for the results derived from the fitted distributions and the direct use of the empirical data.

Figure 3a Cost and QALY differences plotted on the cost-effectiveness plane derived from the case study DES model using theoretically specified input distributions

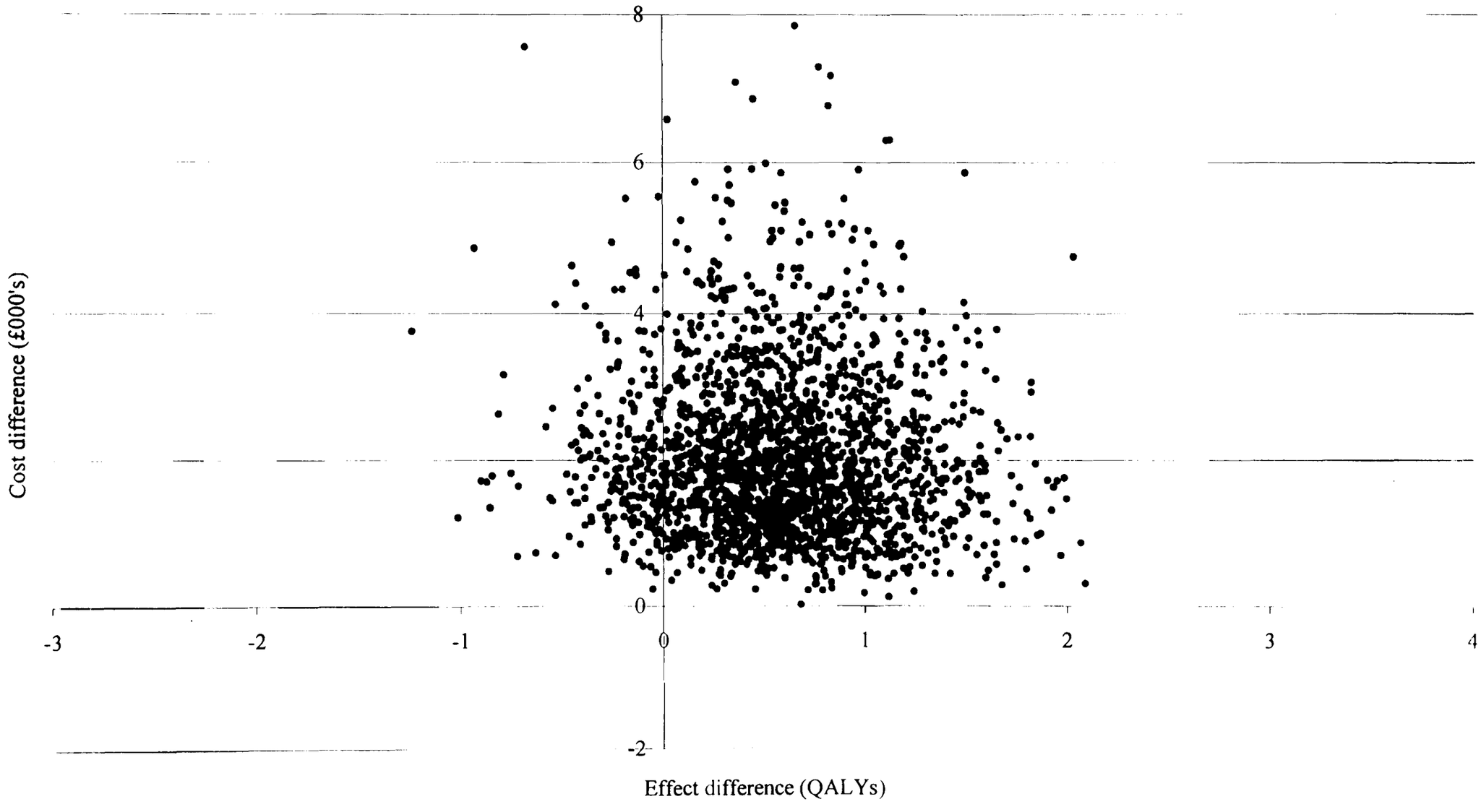


Figure 3b Observations of the cost and QALY differences plotted on the cost-effectiveness plane derived from the DES model using distributions fitted to the weighted empirical data

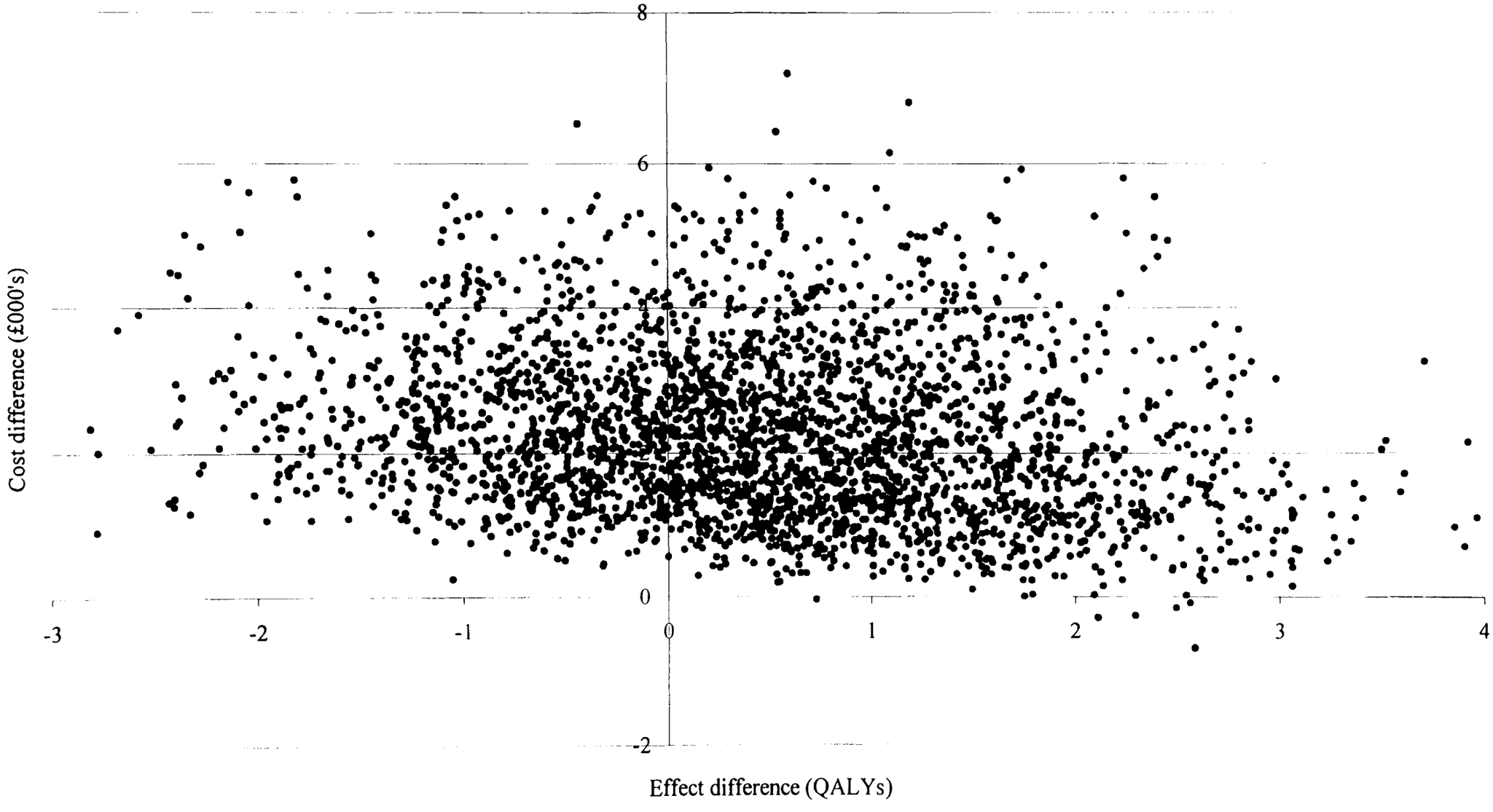


Figure 3c Observations of the cost and QALY differences plotted on the cost-effectiveness plane derived from the DES model using weighted empirical data directly

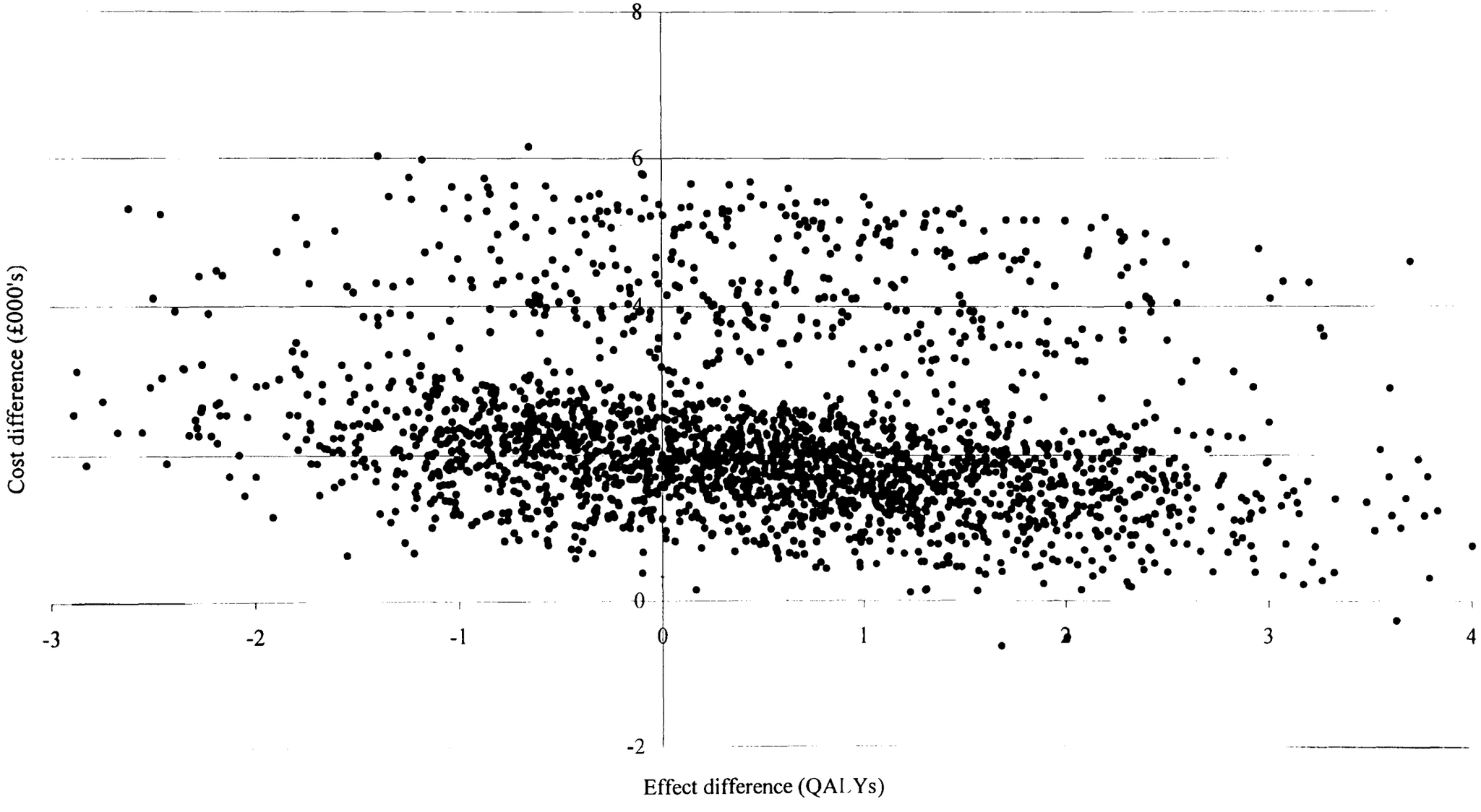
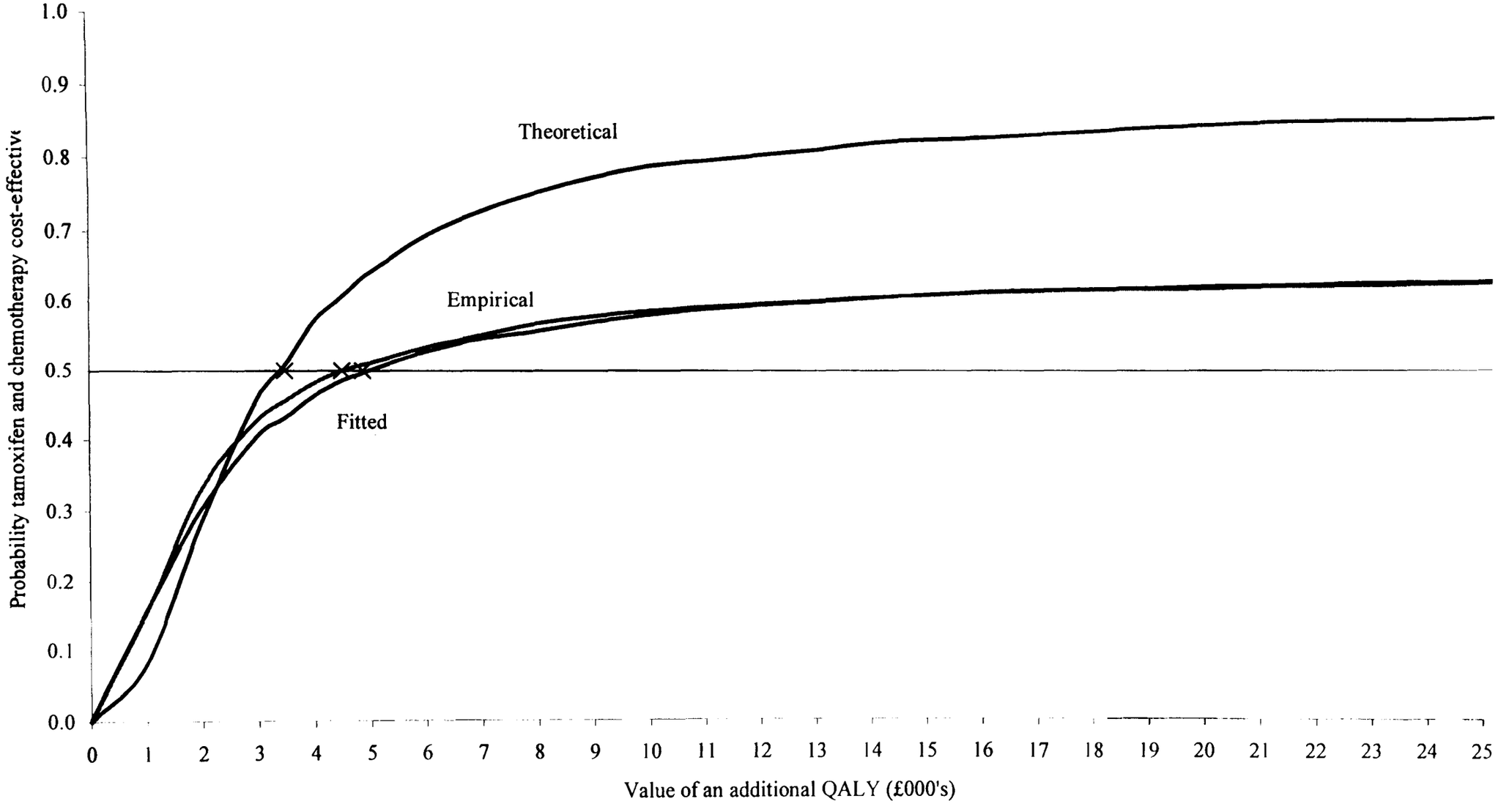


Figure 4 **Cost-effectiveness acceptability curves derived from the case study DES model using the alternative methods for specifying input distributions**



8.3 Comparison of Markov process and DES results to inform immediate resource allocation

The model outputs derived from the two modelling techniques, using all three methods of specifying probability distributions, are brought together in Table 8.5.

Table 8.5 Comparison of the model outputs from the case study Markov process and DES model

	Costs		QALYs		Life years		ICERs*	
	DES model	Markov process	DES model	Markov process	DES model	Markov process	DES model	Markov process
Tamoxifen and chemotherapy								
Theoretical	£9,146	£8,718	12.14	12.01	16.01	15.74	£3,483	£3,334
Fitted	£9,439	£8,893	11.63	11.56	15.82	15.63	£4,869	£4,473
Empirical	£9,473	£8,862	11.71	11.62	15.91	15.72	£4,498	£4,295
Mean	£9,353	£8,824	11.83	11.73	15.91	15.70	£4,226	£3,988
Tamoxifen alone								
Theoretical	£7,115	£6,709	11.56	11.41	15.16	14.86		
Fitted	£7,182	£6,721	11.17	11.07	15.13	14.90		
Empirical	£7,233	£6,653	11.22	11.11	15.18	14.94		
Mean	£7,177	£6,694	11.31	11.20	15.15	14.90		

* The mean ICER is calculated as the ratio of the difference in mean costs and the difference in mean QALYs, rather than the mean ratio of the ICERs.

The mean cost, QALY, and life year estimates varied between the two modelling techniques, but in the same direction - all three outputs were higher in the DES model. The differences in the cost estimates were between £406 and £611. The QALY estimates varied by between 0.07 and 0.15 QALYs, though for this output tamoxifen alone was subject to the greater differences. These results led to small divergences in the ICERs reported by the two models, though the maximum difference was under £400. The credible intervals derived from the two models, presented in the sections above, were also similar. None of the 2.5th percentiles exceeded £600 and the 95th percentiles all reported tamoxifen alone dominating tamoxifen and chemotherapy. The reasons for these differences are explored in section 9.2.

8.4 Value of information analysis

The results presented in the previous sections informed the decision regarding the provision of the alternative therapies given the available data. An additional objective of an economic HTA decision model may involve the valuation of securing more data on the values of the input parameters. The following sections present the results of both the Markov process and the DES model with respect to valuing the collection of further information to inform the resource allocation decision. Differences in the respective results are highlighted, though their implications are discussed in the conclusions section of this Chapter, and the causes of variation are explained in the following Chapter.

The VoI is partly determined by the value to the decision-maker of an additional unit of effect - in this evaluation the value of a QALY. If the relevant value is unknown the VoI should be re-estimated for a range of non-negative values. For the purpose of the case study evaluation a VoI analysis was undertaken for a single baseline value of a QALY, which enabled an understanding of the approach, as well as facilitating a comparison of the two modelling techniques. The value of an additional QALY was assumed to be £5,000 - a value chosen to demonstrate different facets of the technique, rather than an attempt to accurately reflect decision-makers beliefs. Chapter 3 contains a full description of techniques adopted to analyse the value of information (VoI). Summaries of the three phases in the estimation of the VoI - estimating the expected value of perfect information (EVPI), the expected value of sample information (EVSI) and the expected net benefit of sampling (ENBS) - are provided alongside the results presented in the following sections.

8.4.1 Estimating the expected value of perfect information (EVPI)

From the baseline analysis presented in the previous sections, at a value of an additional QALY of £5,000 tamoxifen and chemotherapy had the highest probability of positive incremental net benefits, but there was a probability that tamoxifen alone was the cost-effective therapy option. The costs of uncertainty (the EVPI) are the expected benefits of providing tamoxifen alone *in the proportion of observations in which tamoxifen alone displayed positive incremental net benefits.*

The EVPI per patient, also known as the $EVPI_{episode}$, describes the costs of uncertainty in applying the resource decision to a single patient. The process for calculating the $EVPI_{episode}$ is portrayed in Table 8.6. For example, the results derived from the Markov process for the theoretically defined probability distributions showed that there was a 0.346 probability that tamoxifen alone was the cost effective therapy option at a value of an additional QALY of £5,000. The mean net benefits per patient receiving tamoxifen alone minus the mean net benefits associated with tamoxifen and chemotherapy was £619.59 for the observations in which tamoxifen alone displayed higher net benefits. Without perfect information, tamoxifen and chemotherapy would be provided to all patients, but with perfect information the decision-maker could identify the observations in which tamoxifen alone was the cost-effective option. The $EVPI_{episode}$ was estimated as the net benefits lost in the proportion of cases in which tamoxifen alone was observed to be the cost-effective therapy option (£619.59) multiplied by the proportion of cases in which tamoxifen alone was the cost-effective therapy option (0.346), which was £214.19.

Table 8.6 Calculating the $EVPI_{episode}$ for both ABC models*

	Observations reporting positive incremental net benefits for Tamoxifen alone		
Markov process	Theoretical distributions	Fitted distributions	Empirical data direct
A. Tamoxifen alone mean net benefits	£17,722.76	£24,774.89	£24,876.76
B. Tamoxifen and chemotherapy mean net benefits	£17,103.35	£22,484.60	£22,607.40
C. Incremental net benefits (A-B)	£619.59	£2,290.29	£2,269.36
D. Probability tamoxifen alone cost effective	0.346	0.486	0.487
$EVPI_{episode}$ (CxD)	£214.19	£1,113.77	£1,106.09
DES model	Theoretical distributions	Fitted distributions	Empirical data direct
A. Tamoxifen alone mean net benefits	£18,542.87	£25,350.01	£24,957.33
B. Tamoxifen and chemotherapy mean net benefits	£17,877.73	£23,019.75	£22,589.88
C. Incremental net benefits (A-B)	£665.13	£2,330.25	£2,367.45
D. Probability tamoxifen alone cost effective	0.3594	0.4998	0.4891
$EVPI_{episode}$ (CxD)	£239.08	£1,164.74	£1,157.92

* The value of an additional QALY is assumed to be £5,000.

The estimates of the $EVPI_{episode}$ emanating from the DES model are very similar to the results from the Markov process, not only in the relationship between the different methods of specifying probability distributions, but also in terms of the magnitude of

the results. However, the $EVPI_{episode}$ estimated using the input data analysis methods based on the creation of weighted datasets was considerably larger than the $EVPI_{episode}$ estimated using the theoretical distributions. At a value of an additional QALY of £5,000, the probability of tamoxifen and chemotherapy being the cost-effective therapy option is around 0.51 using the weighted datasets, which maximises the costs of uncertainty.

The prospective decision informed by the collection of additional data will not be limited to a single patient, but the whole population of eligible patients [Claxton, 1999]. The $EVPI_{population}$ is the EVPI for the relevant patient population, which can be estimated by multiplying the $EVPI_{episode}$ by the eligible patient population over the period for which the allocation decision is expected to remain, discounted at an appropriate rate (see section 3.6.3.1). There are around 33,000 new cases of breast cancer diagnosed annually in the UK [Cancer Research Campaign, 1998]. Roughly 20% of the new cases involve postmenopausal women with node positive (stage II) breast cancer, so the annual incidence for the patient group evaluated in this evaluation was estimated as $33{,}000 \times 0.2 = 6{,}600$. The specification of the number of years over which the research will inform the decision-making process was primarily subjective. In this treatment area, five years appeared to be a reasonable estimate of the time between the availability of new therapy options and was employed as the baseline value. Sensitivity analyses were undertaken to test the impact of the assumed length of usefulness of the results of the prospective research.

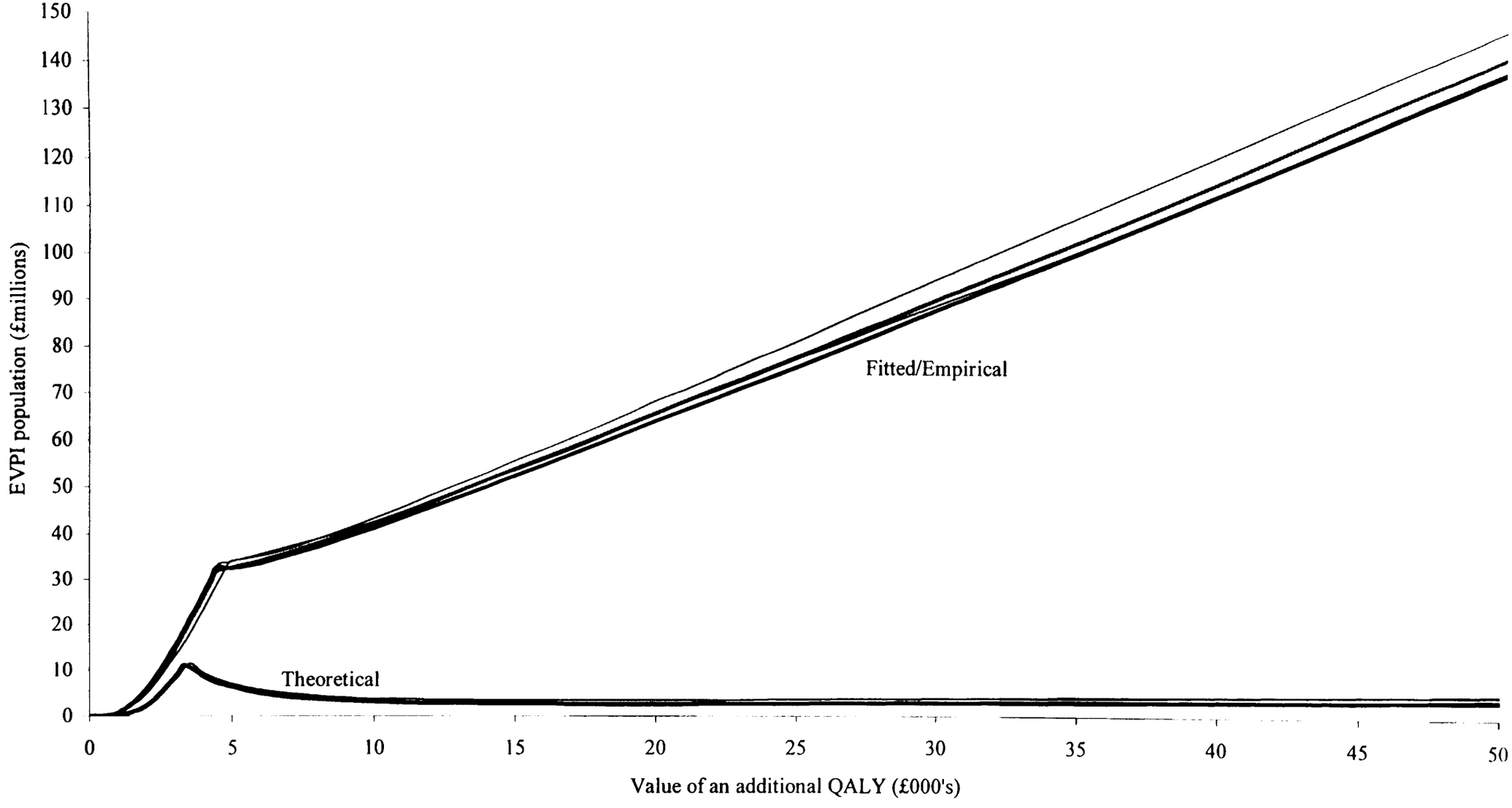
The relationships between the value of a QALY and the $EVPI_{population}$ are presented in Figure 8.5 for the alternative methods of specifying probability distributions for both ABC models. The first result to note is that the two models produced very similar curves for the respective methods for specifying probability distributions. Looking at the alternative input data analysis methods the curves derived from the theoretical distributions show that the EVPI increased steeply from a value of a QALY of £0 to £4,000, which was around the value at which tamoxifen alone had the highest probability of being cost-effective. At the point at which tamoxifen and chemotherapy became the more likely cost-effective option the EVPI fell successively until the value

of a QALY was £19,000, at which point the EVPI resumed an upward path. This relationship demonstrates that the EVPI is not necessarily higher with larger values of an additional QALY because the probability of choosing the 'wrong' therapy option is likely to decline at higher values, which may be a stronger factor on the EVPI than the value of an additional QALY. At extreme values of an additional QALY, where the probability of tamoxifen and chemotherapy being cost-effective levels out, the EVPI rose continuously.

The $EVPI_{population}$ estimated using the fitted distributions and the empirical data directly were substantially larger than the $EVPI_{population}$ estimated using the theoretically specified probability distributions. The increased magnitudes were due to differences in the level of certainty that tamoxifen and chemotherapy was the cost-effective therapy option. Using the theoretical distributions it was more likely that tamoxifen and chemotherapy was cost-effective. Using the fitted distributions and the empirical data direct the probability of tamoxifen and chemotherapy being cost-effective never rose above 0.7. This meant there was always a sizeable probability that tamoxifen alone was the cost-effective option, whereas the probability of tamoxifen and chemotherapy being cost-effective using the theoretical distributions reached 0.8 by a value of a QALY of £7,000.

Similar methods for the analysis of the EVPI for specific input parameters, or groups of parameters, can be applied, whereby the model is analysed with only the parameters of interest being described stochastically. The estimated EVPI is then attributed to the stochastic parameters. The segregated analysis of particular parameters can be extended to the full VoI methodology, but only an aggregate VoI analysis is presented in this Chapter because it provides a sufficient basis for understanding the methodology, as well as for the comparison of the two modelling techniques.

Figure 8.5 Plot of expected value of perfect information against the value of a QALY derived from both case study models using the alternative methods for specifying input distributions



8.4.2 Estimating the expected value of sample information (EVSI)

The EVSI is the difference in the EVPI estimated using the baseline input data and the EVPI estimated using alternative input distributions that reflect the assumed impact of additional data. The additional information reduces the variation described in the input distributions, which in turn reduces the EVPI because the variation in the distribution of the net benefits is also decreased.

Updating the input probability distributions to reflect an assumed decrease in the uncertainty around the values of the input parameters is a key element in the estimation of the EVSI. Any form of distribution may be updated using the Bayesian statistical software - WinBUGS - but the operation of updating data using this method is not fully developed, though work is ongoing (personal communication: Liz Fenwick). The development of methods for updating input probability distributions would shift the focus of this thesis away from the central aim, which is the comparison of alternative modelling techniques. Therefore, the estimation of the EVSI presented in this thesis was undertaken using only the theoretically defined input distributions. Simple formulae for updating these probability distributions were available that used established knowledge about the relationships between conjugate families of probability distributions [Barnett, 1999]. As discussed in Chapter 3, a reasonable assumption about the mean value of the additional data for any parameter is that they are most likely to have the same mean values as the data used to populate the baseline model. Details of the process for updating the probability distributions are provided in Appendix 5.

To establish the relationship between the size of a prospective study and the reduction in the costs of uncertainty the EVPI was re-estimated for successively increasing prospective samples until the optimal size was determined, i.e. the net benefits of sampling started to decline. Each patient entered a prospective trial as a node positive, postmenopausal patient with early breast cancer. Within each therapy group the proportions of patients who would provide data on each of the health states within the model was estimated using the baseline point estimates describing the probabilities of experiencing events.

The allocation of a prospective sample of patients between the two therapies may not be equal. Ideally, the optimal sample allocation between therapy options in a prospective trial would be established by estimating the ENBS 'for every feasible allocation of each sample considered' [Claxton, 1999]pg355. However, to reduce the analytic burden in this evaluation a constant ratio of patients allocated to the alternative therapy options was estimated using Neyman's allocation to strata formula, which accounts for differential standard deviations and costs within separate strata (or in this case - treatment arms). In this formula the standard deviations of the net benefits describe the variation between patients, which required the estimation of the first-order uncertainty associated with the point estimates for each input parameter. The estimation of first-order uncertainty was possible using the DES model, but the Markov process was cohort based, which precluded the analysis of first-order uncertainty so the deviations from the net benefits associated with the two therapy options were assumed to be equal.

The marginal sampling costs were obtained from the ABC trial and included the average cost of recruiting each patient into the trial, which was based on a grade F nurse spending an hour explaining the trial to each patient and a successful recruitment rate of 1 in 3 patients (£150). Additional treatment costs were attached to patients receiving tamoxifen and chemotherapy, which were estimated as the baseline difference between the two therapies (£1,225). The above data were applied to Neyman's allocation to strata formula (see section 3.6.3.2) to estimate a constant ratio of sample allocation between tamoxifen alone and tamoxifen and chemotherapy for the Markov process:

$$n_1 : n_2 = \frac{1/\sqrt{150}}{(1/\sqrt{150}) + (1/\sqrt{1{,}375})} : \frac{1/\sqrt{1{,}375}}{(1/\sqrt{150}) + (1/\sqrt{1{,}375})} = 0.752 : 0.248$$

and for the DES model:

$$n_1 : n_2 = \frac{47656/\sqrt{150}}{(47656/\sqrt{150}) + (48586/\sqrt{1{,}375})} : \frac{48586/\sqrt{1{,}375}}{(47656/\sqrt{150}) + (48586/\sqrt{1{,}375})} = 0.748 : 0.252$$

Table 8.7 presents the re-estimated EVPI associated with alternative prospective samples, and the related EVSI. Each re-analysis of the Markov process was informed by 10,000 runs, whilst each analysis of the DES model comprised 1,000 runs. The

results show that 65% and 69% of the $EVSI_{population}$ derived from a prospective sample of 2,000 patients were provided by a sample of only 500 patients using the Markov process and the DES model, respectively. The magnitude of the $EVSI_{population}$ were slightly higher using the DES model.

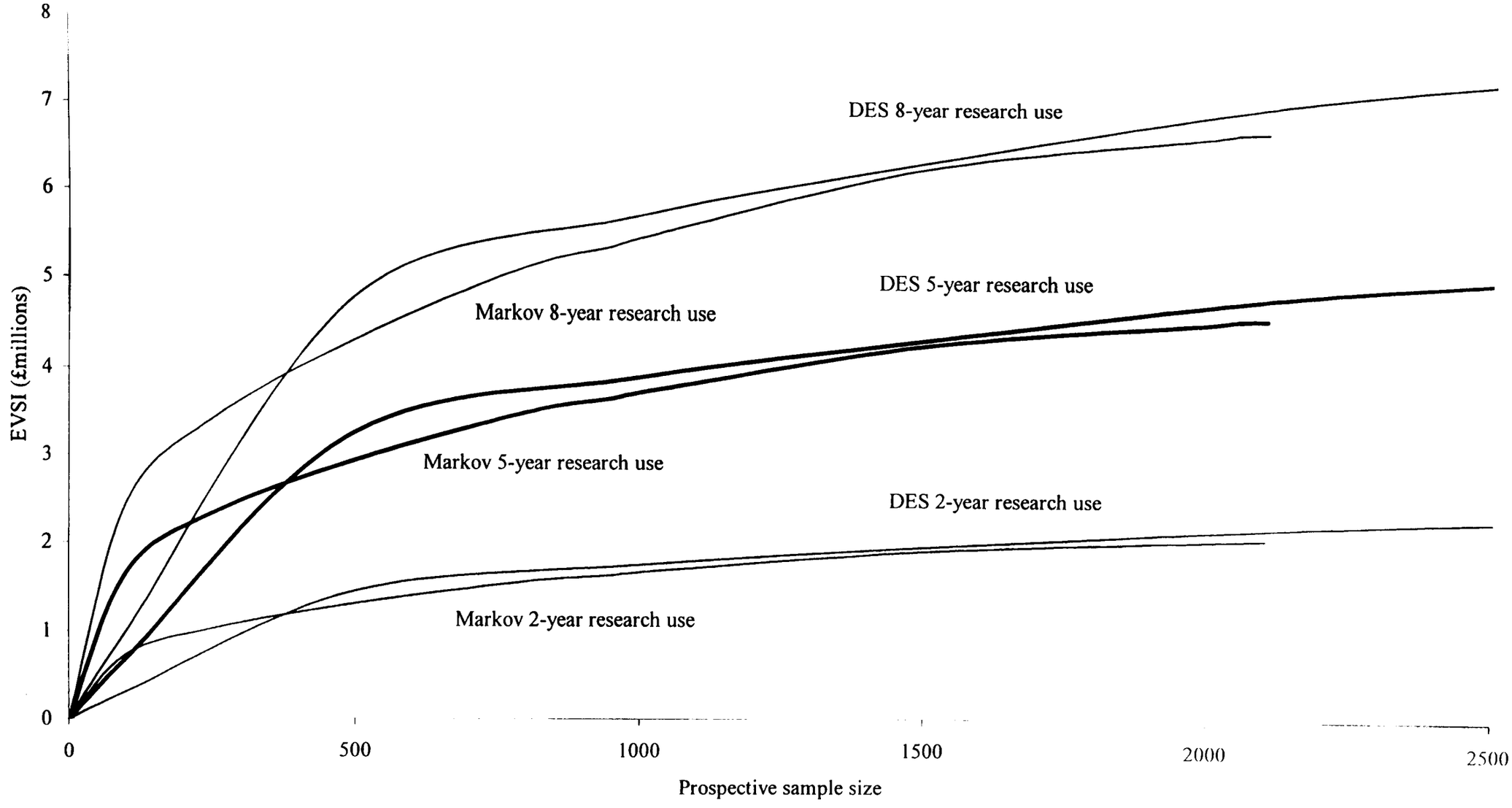
Table 8.7 Calculating the $EVSI_{population}$ for both case study models

	Markov process			DES model		
Prospective sample	$EVPI_{episode}$	$EVSI_{episode}$	$EVSI_{population}$	$EVPI_{episode}$	$EVSI_{episode}$	$EVSI_{population}$
0 (baseline)	£214.2			£239.08		
100	£158.2	£55.9	£1,664,626	£216.5	£22.6	£686,394
500	£116.2	£98.0	£2,946,781	£130.4	£108.7	£3,268,146
1000	£90.4	£123.8	£3,737,909	£110.0	£129.1	£3,913,835
1500	£72.6	£141.6	£4,282,003	£96.6	£142.5	£4,343,201
2000	£64.7	£149.4	£4,533,611	£84.6	£154.5	£4,722,612
2050	£63.6	£150.6	£4,568,562	£79.8	£159.3	£4,874,716
2100	£63.5	£150.7	£4,575,442	£76.5	£162.6	£4,982,320

$EVSI_{episode}$ = Baseline$EVPI_{episode}$ - Re-estimated $EVPI_{episode}$, $EVSI_{population}$ = $EVSI_{episode}$.Patient population (29470), assumed length of usefulness of research is 5 years.

Figure 8.6 illustrates the data graphically, plotting the EVSI against prospective sample sizes for both modelling techniques. Three curves are displayed to illustrate the effect of the assumed duration of applicability of the proposed research. The $EVSI_{population}$ rose steeply at lower samples before starting to plateau towards prospective samples of around 2,000. The assumed use of the research clearly made a significant impact on the $EVSI_{population}$, which varies by almost £5 million for a sample of 2,000 between an assumed period of research use of 2 and 8 years. The curves reinforce the similarity between the results produced by the alternative models.

Figure 8.6 Expected value of sampling information for alternative lengths of applicability of the proposed research



8.4.3 Estimating the expected net benefit of sampling (ENBS)

The ENBS for each prospective sample size is the EVSI minus the cost of obtaining the additional sampling information. The cost of sampling includes the cost of setting up the study, and monitoring, analysing and disseminating the data collected. To specify increasing costs for larger sample sizes it was necessary to make an assessment of the fixed and variable costs associated with the trial. In addition, the cost of chemotherapy was subtracted from the EVSI on the assumption that chemotherapy would not be administered to this patient group in the absence of the trial. Fortunately, estimates of the relevant costs were available from the ABC trial, which is currently ongoing [UKCCR, 1993]. The Medical Research Council has funded this trial for a total of 10 years and the fixed costs have been estimated at around £1.125 million. Applying these fixed costs, in addition to the marginal costs presented above, to the formula presented in section 3.6.3.3 estimates the total costs associated with any prospective samples:

$$ENBS_{population} = EVSI_{population} - £1{,}125{,}000 - £150n - £304n$$

Table 8.8 presents the previously estimated $EVSI_{population}$ alongside the additional sampling costs for alternative prospective samples in order to estimate the $ENBS_{population}$. The DES model reported that net benefits were actually negative for a prospective sample of 100 patients. Figure 8.7 plots the $ENBS_{population}$ against the respective sample sizes, again plotting curves for three assumed lengths of application for the proposed research. The assumed use of research had a major impact on the value of collecting further information. The baseline assumption of 5 years applicability indicated that the $ENBS_{population}$ was maximised at a prospective sample of around 2,050 patients using the Markov process and between 2,250 and 2,500 patients using the DES model. A similar result was observed for a length of research use of 8 years. However, if the research influenced policy for only 2 years, the optimal sample size would be between 1,500 and 2,000 using the Markov process, and between 1,000 and 1,500 using the DES model.

Figure 8.7 Expected net benefits of sampling for alternative lengths of applicability of the proposed research

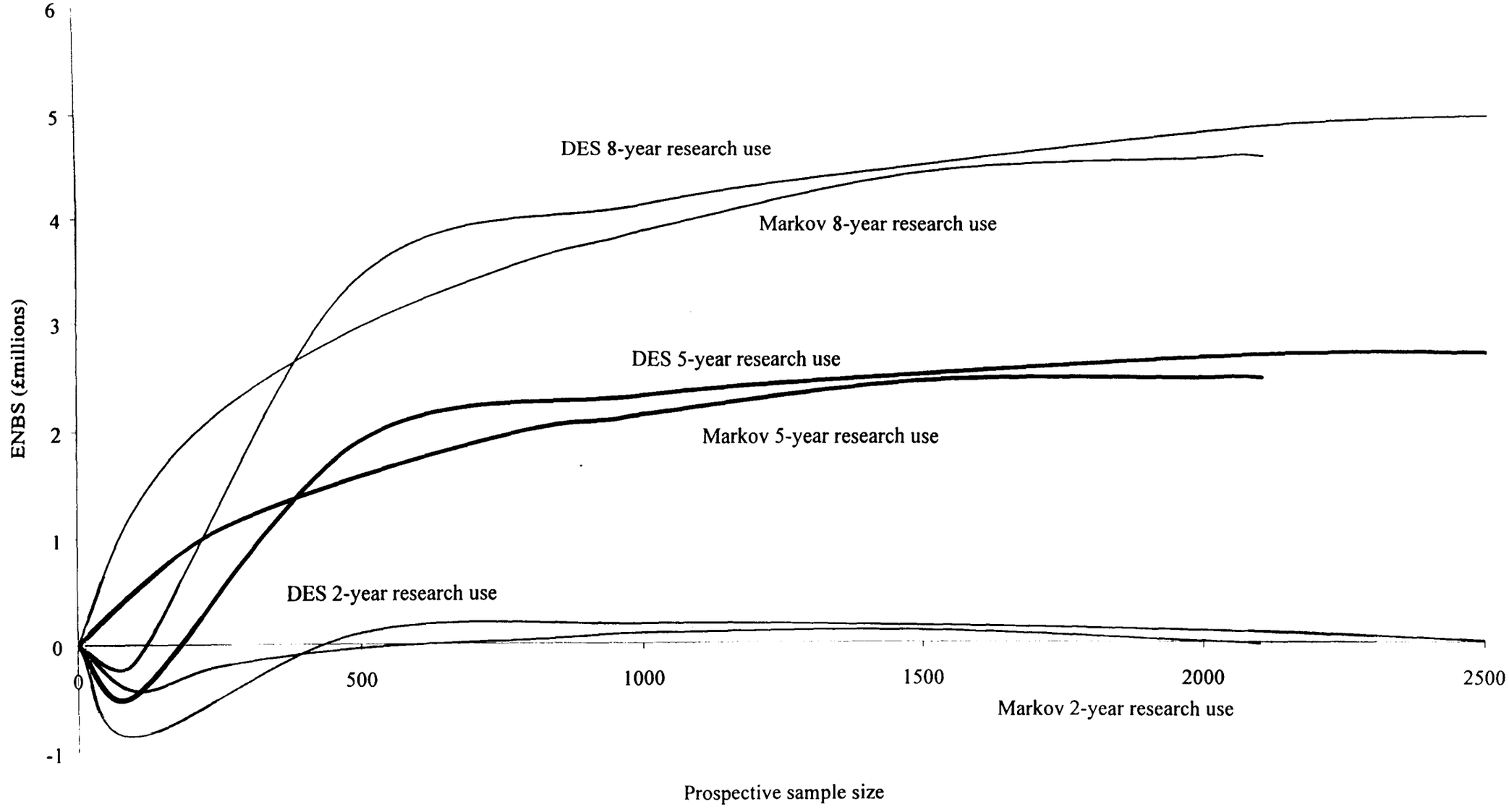


Table 8.8 Calculating the $ENBS_{population}$ for both ABC models

	Markov process			DES model			
Size of trial	$EVSI_{population}$	*Sampling Cost*	$ENBS_{population}$		$EVSI_{population}$	*Sampling Cost*	$ENBS_{population}$
100	£1,664,626	£1,170,625	£494,001	100	£686,394	£1,170,625	-£484,231
500	£2,946,781	£1,353,125	£1,593,656	500	£3,268,146	£1,353,125	£1,915,021
1000	£3,737,909	£1,581,250	£2,156,659	1000	£3,913,835	£1,581,250	£2,332,585
1500	£4,282,003	£1,809,375	£2,472,628	1500	£4,343,201	£1,809,375	£2,533,826
2000	£4,533,611	£2,037,500	£2,496,111	2000	£4,722,612	£2,037,500	£2,685,112
2050	£4,568,562	£2,060,313	£2,508,249	2250	£4,874,716	£2,151,563	£2,723,153
2100	£4,575,442	£2,083,125	£2,492,317	2500	£4,982,320	£2,265,625	£2,716,695

$ENBS_{population} = EVSI_{population} - Sampling\ Cost$

8.5 Conclusions

The results derived from experimentation with the two decision models have been presented and compared in this Chapter. The analyses compared the administration of tamoxifen and chemotherapy versus tamoxifen alone in a subset of breast cancer patients – postmenopausal women with node positive early breast cancer. Two broad objectives were addressed. Firstly, the question of whether to allocate resources to the provision of chemotherapy, given only the identified data, was investigated. Both case study models were evaluated three times to inform an immediate resource allocation decision, using alternative methods for pooling and formatting the identified data into probability distributions. The second issue covered the analysis of stochastic decision models with the objective of valuing the collection of further data, using a Bayesian value of information analysis. This section describes the policy implications that could be drawn from the various results presented in this Chapter, a discussion of the potential causes of differences in the results derived from competing methodologies is presented in Chapter 9.

Comparing the alternative input data analysis techniques the results derived from both models showed that the theoretical specification of probability distributions provided lower mean estimates of the cost per additional QALY from adding chemotherapy to tamoxifen. However, the maximum difference between the mean ICERs was only around £1,400 so the likelihood of alternative decisions being made on the basis of the mean results is small. However, if the level of uncertainty around the mean estimates was incorporated into the decision making process, the choice of input data analysis method could have a significant impact on the resource allocation decision. Comparing

the CEAc curves estimated by either model (Figures 8.2 and 8.4), it is clear that the results derived using the fitted distributions and the empirical data direct indicate far less certainty about the cost-effectiveness of tamoxifen and chemotherapy. These analyses showed that over 30 per cent of observations indicated that tamoxifen was the dominant strategy, whilst the corresponding percentage using the theoretical distributions was less than 10 per cent. Such uncertainty about the cost-effectiveness of the more expensive intervention could prevent the allocation of resources to chemotherapy, a decision that is more likely using the fitted distributions and the empirical data direct.

Comparison of the results derived from the two models showed that the DES model consistently estimated higher values for both costs and effects, but because the differences were proportionally similar there was little difference between the estimated incremental cost-effectiveness ratios (ICERs). Indeed, the largest difference between the respective ICERs was only £400. It is unlikely, therefore, that decision-makers would come to contrary decisions due to the use of alternative modelling techniques using the results that are presented in this thesis.

The second main section in this Chapter covered the results of the analyses of the value of information (VoI) within a Bayesian framework. The estimation of the VoI was split into three sections. The first stage estimated the expected value of perfect information (EVPI), for which all three methods for specifying probability distributions were used. The results revealed a huge difference in the potential value of additional research depending on the method used to pool and format the identified data. As shown in Figure 8.5, at a value of an additional QALY of £10,000 the EVPI estimated using the theoretically specified distributions was around £3.5 million, whereas the corresponding estimates produced using the fitted distributions and the empirical data direct were around £43.5 million. The variation in the estimates of the EVPI due to the choice of methods over the analysis of the available data has serious implications for the role of VoI analyses in informing funding decisions over future research.

The methodology driving the second stage of Bayesian VoI analyses is under development. This stage requires that the baseline probability distributions describing uncertainty in the input parameters be updated to reflect the hypothetical collection of

further data. At present, simple methods for updating only the theoretically specified distributions using the properties of conjugate families of prior distributions are available so the full VoI analysis was only undertaken using the theoretical distributions. Given the divergent estimates of the EVPI, it is likely that the alternative methods would value the collection of further information differently. If the difference in the valuation of information was of a similar magnitude to the difference in the EVPI the choice of input data analysis method could lead to very different research agendas. Indeed, such differences would necessitate a firm judgement on the appropriate set of probability distributions, a decision that could possibly be avoided if a model were only informing the immediate allocation of resources.

Using the theoretically specified distributions both models estimated a similar optimal size for a prospective trial of around 2,050 and between 2,250 and 2,500 for the Markov process and the DES model, respectively. However, the predicted VoI differed to a larger extent for prospective samples below 1,000. For example, using the baseline assumption of the 5-year length of applicability for the proposed research, the Markov process predicted continuously positive net benefits of sampling, whilst the DES model predicted negative net benefits up to a sample size of around 400. Further discussion of the practical and theoretical issues relating to the analysis of the VoI are presented in Chapter 10.

Chapter 9 Case study: methodological issues arising

9.1 Introduction

The main objective of this Chapter is the exploration of the causes of differences in the results presented in the previous Chapter and the associated methodological implications for the conduct of economic HTA decision models. Three methodological issues are discussed in this Chapter, though the primary discussion centres on the main objective of this thesis - the comparison of alternative modelling techniques. The comparison of the DES model and the Markov process is described in the context of an overall economic HTA decision modelling project, relating to the five statements of model characteristics defined at the end of Chapter 2.

One of the secondary objectives concerned the identification of differences between the results obtained from alternative methods for pooling and formatting data to populate decision models. The following sections discuss the results derived from using the three identified methods - theoretically-based distributions, fitted distributions, and direct empirical data.

Within the case study evaluation the value of collecting further data on input parameters was estimated within a Bayesian value of information (VoI) analysis. Such analyses are relatively new to the field of economic evaluation and the methods used were adapted from those advocated by the health economists who introduced VoI analysis to economic HTA decision modelling. The methodological implications concerning the use of the alternative modelling techniques and methods for pooling and formatting the

input data in a VoI analysis are presented in this Chapter. However, the main objective underlying the application of the VoI analyses was to critique the current methodology and to specify particular issues requiring further research, which is presented in Chapter 10.

9.2 Comparing the alternative modelling techniques

At the outset of this thesis the primary objective was to gain a better understanding of the relative advantages and disadvantages of discrete event simulation models and Markov processes as tools for the economic evaluation of health care technologies. Chapter 8 presented the results derived from the alternative techniques and discussed the policy implications of the observed differences. To recap, the DES model estimated slightly higher costs per additional QALY from adding chemotherapy to tamoxifen, but the closeness of the results suggested that it was unlikely that the use of one model's results over the other would lead to an alternative decision. However, the lack of any significant difference in the results produced by the alternative models does not mean there are no significant differences between the two modelling techniques. During the course of this thesis several potentially important differences have been revealed.

Chapter 2 set out the, *a priori,* state of knowledge about the available modelling techniques that concluded with five statements regarding potential factors that may influence the choice of decision modelling technique. The five statements are discussed below in the light of the work completed in this thesis.

1. *If anything other than short-term outcomes are to be modelled decision trees are an inappropriate choice of modelling technique.*

This first statement set the scene for the comparison of DES models and Markov processes. The case study employed in this thesis covered a long time horizon - 50 years - and it is inconceivable that a decision trees would be used to model patient pathways for any intervention over such an extended time horizon. At present decision trees, Markov processes and DES models are the three main modelling techniques that have been applied to economic HTA decision models, though it is possible other decision modelling techniques may be applied in the future.

2. *If model parameters are a function of the time spent in particular states DES will more accurately reflect the true relationships between health states.*

This statement relates to the constraining factor in Markov models known as the Markovian assumption, whereby the pathway of a patient from their current state is dependent only on the current state (and possibly the total length of time spent in the model). The Markovian assumption prevents the application of differential probabilities to patients within the same health state. In effect, a patient is a patient is a patient in every state within the model. DES allows patient pathways to be influenced by any factor within the model, including the representation of alternative patient characteristics on entry to the model.

Within the case study evaluation described in this thesis the Markovian assumption limited the description of patient pathways after the experience of a locoregional relapse. In the ABC models patients experiencing a locoregional relapse progressed to a more severe site of relapse (metastases) or straight to death. The identified data describing progression from locoregional relapse were mainly presented in the form of disease-free survival curves representing the probability of progressing in successive time periods from the point of diagnosis with a locoregional relapse. To apply data directly from a survival curve to patients within a decision model it is necessary to know when each patient entered the state of interest in order to apply the differential probabilities of experiencing an event over time. In the DES model, patients remaining disease free at the end of the available follow-up adopted the survival profile of the general population. This assumption was based on qualitative data reported in the literature. The DES model noted the time at which patients entered the locoregional relapse health state, as well as recording the age of each patient. This enabled the application of age-specific mortality rates to each patient that remained in remission to the last point of identified follow-up.

The Markovian assumption requires that the data available in the survival curve must be transformed to a constant probability that is applied to every patient within the health state. The ideal output from the conversion of the survival data is a mean length of disease-free survival (DFS), which would produce, in the absence of discounting, the

same mean results as those derived from the use of the survival curve directly. However, if a proportion of patients do not experience the event reported by the survival curve, the estimation of a constant probability can only be approximate because the mean 'survival' cannot be fully estimated. In the case study Markov process it was not possible to determine the age of the patient at the end of the period of follow-up because the state was within the model and a constant probability of experiencing an event was applied to all patients in all time periods. A specific time period spent in the health state was required for the whole patient group, which necessarily resulted in some inappropriate estimates of the time spent in the state.

Two options were considered to represent the period of disease-free survival following a locoregional relapse in the Markov process. Firstly, in each of the studies presenting data on DFS following a locoregional relapse the median patient had experienced a relapse, so the median DFS was used as an estimate of the mean DFS. However, when the results of the two decision models were compared the DES model produced considerably larger estimates of overall survival and QALYs associated with both interventions. After a thorough inspection of the two models it was discovered that the use of the median DFS in the Markov process was the main cause of the observed differences in the model outputs. The distribution of DFS was significantly skewed to the right and the approximated mean used in the DES model was substantially higher than the median estimate.

An alternative method for estimating the mean DFS for the Markov process was then adopted, which assumed a maximum length of survival of 20 years from the point of locoregional relapse. For each study a mean survival time within the specified 20 years was estimated. The specification of a maximum survival period was necessary because the survival curve presented by the majority of the available studies had plateauxed by the end of the reported follow-up period so it was difficult to extrapolate the curves.

Table 9.1 presents the model outputs from the Markov processes analysed using the median length of disease-free survival and the approximate mean. The data show that the effect on the aggregate costs is negligible, but there is a more substantial effect on the total QALYs and life years associated with both therapies. The differences give an indication of the importance of the assumptions made with respect to this one input

parameter. The two methods employed to estimate the length of disease-free survival required stronger assumptions about the pathways of patients following a locoregional relapse than were employed in the DES model. Indeed, these were just two of a multitude of alternative approaches that could have been adopted to describe disease-free survival.

Table 9.1 Comparison of model outputs from the Markov process using the median length of disease-free survival and the approximate mean disease-free survival

	Costs	QALYs	Life years
Mean (tamoxifen+chemotherapy)	£8,862	11.62	15.72
Median (tamoxifen+chemotherapy)	£8,934	11.23	15.13
Difference	-£72	0.39	0.58
Mean (tamoxifen alone)	£6,721	11.07	14.90
Median (tamoxifen alone)	£6,741	10.70	14.33
Difference	-£20	0.37	0.57
ICER (mean)	£3,896		
ICER (median)	£4,159		
Difference	£263		

In the case study evaluation, the majority of data identified in the literature described survival times for the separate metastatic sites as median survival times. The survival times for patients diagnosed with metastases were relatively short - soft tissue metastases had the longest median survival of around 31 months. It is likely that a few outlying patients with long survival times would increase the mean values, but without data to support this assumption the most conservative approach was to use the non-adjusted medians. If survival curves become available to describe time to events in other states, such as survival time from metastases, the assumptions imposed on the data employed in Markov processes could considerably undermine confidence in the model's outputs.

The representation of disease-free survival as a constant probability of experiencing an event also had another consequence. In the DES model it was possible to make the assumption that if patients had not experienced a relapse after 11 years in remission they were 'cured' and the next event of interest to the decision model was death. In the Markov process it was not possible to split the destination of patients according to time spent in the state, so all patients leaving the state were subject to the same probabilities of experiencing a more severe relapse, or progressing straight to death. The impact of this shortcoming was limited in the case study because only a small proportion of

patients remained disease free at the end of the follow-up period. However, modelling scenarios in which this issue has a far larger impact can be envisaged and it is more difficult to correct for this deficiency than to revise estimates of the median survival period.

There were other areas in the case study evaluation in which the development of patient pathways that were influenced by the time spent in a earlier states could have demonstrated further advantages of the DES model over the Markov process if more detailed data were available for the evaluation. For example, disease-free interval (DFI) is known to be a prognostic factor that reflects the intrinsic growth rate of the tumour [Borner et al, 1994]. It may be plausible to assume that patients with short lengths of DFI would be more likely to experience a relapse, rather than dying with no evidence of disease, than patients with a longer period of DFI. Moreover, it has been reported that the length of DFI affects a patient's prognosis from the time of diagnosis with a recurrence [Ingle et al, 1994; Wong and Henderson, 1994]. Unfortunately the secondary data sources used in this thesis did not produce data that described possible relationships in enough detail to warrant inclusion in the models. If more detailed primary data becomes available it is possible that certain patient characteristics could be employed as attributes within the DES model, which may provide an improved representation of the treatment area that would enhance the advantages of DES models.

3. *If the specification of similar health states that differ only with respect to the experience of previous states compromises the clarity of the model, the use of DES should be considered.*

The third statement refers to the possible proliferation of health states in a Markov process in order to represent states that have similar characteristics, but where subsequent patient pathways are influenced by a patient's treatment history. Only one such example occurred in the case study evaluation, which involved the description of the experience of toxicity. Three forms of toxicity were described - major, grade 3 or 4, and grade 1 or 2 - that could be experienced simultaneously. The available data described the proportion of patients experiencing each form of toxicity.

In the DES model, no separate health states were defined to represent toxicity, rather the experience of toxicity was assigned to each patient within the health state DFI. Each patient sampled from three binary distributions representing the probabilities of experiencing the three forms of toxicity, so each patient could experience any combination of the toxicity categories. The attributes containing the information on the experience of toxicity were incorporated into the estimation of the costs and QALYs associated with each patient's time in DFI as they left the state. Modelling events such as the experience of toxicity as attributes could be viewed as increasing the black box nature of the modelling process, because these events are not explicitly presented in any structural representation of the model. However, the incorporation of such events can be clearly explained in the text and, in any case, the presentation of states in a diagram does not guarantee their proper representation in the model itself.

In the Markov process it would have been possible to create seven states that described each possible combination of the different types of toxicity, though this would have required probability estimates for each combination of toxicity classifications, which were not available in the literature. To avoid further strong assumptions only three toxicity states were included in the Markov process representing the experience of each type of toxicity independently. Hence, patients could experience only one form of toxicity. If, in any run of the Markov process, the sum of the probabilities of experiencing the three forms of toxicity exceeded 1 the probability of experiencing the least severe form of toxicity (grade 1 or 2) was reduced.

In the analysis of second-order uncertainty, which concerns differences across populations, the aggregate effects of toxicity for a patient group would only differ slightly, due to the adjustments to the probability of experiencing grade 1 or 2 toxicity when the sum of the toxicity probabilities exceeded 1. Differences may have occurred in analyses of first-order uncertainty (if the Markov process had been analysed using first-order Monte Carlo simulations) because there would be more variation between the individual patients reported by the DES model. However, as described in Chapter 3, for the analysis of cost-effectiveness to inform resource allocation first-order uncertainty is irrelevant (see section 3.4.3).

4. *If the data describing the timing of events are not in the form of transition probabilities then DES will provide a truer representation of reality.*

Issues around the format of the available data are only applicable to modelling studies that employ secondary clinical data. In such circumstances the analyst is normally bound by the presentational norms for particular types of data. The case study evaluation highlighted clinical data presented as set survival times, rather than as the probability of experiencing an event in a particular interval, as one area in which the format of the identified data led to significant differences in the results of the two models. The DES model incorporated such data in exactly the format that the data were available, whilst the Markov process required the conversion of set survival times to constant probabilities of experiencing death at any point following the diagnosis of metastases. The impact of these different approaches on the outputs of the models was first noted during the verification of the action of discounting within the models (see section 7.2.3), where differences between the results of the DES model and the Markov process were observed. For example, for a set survival period of 8 months the DES model discounted the full survival period for each patient at the same rate (assuming the 8 months fell within the same year for each patient). However, in the Markov process the distribution of outputs associated with the survival time for each patient were spread over a longer period because all the patients within the state were subject to a constant probability of dying.

To illustrate the impact on costs, quality adjusted life months (QALMs) and life years, which were all subject to alternative discount rates, a macro was set up in an Excel spreadsheet to estimate the outputs associated with varying lengths of survival derived from the two models. Set survival times of between 1 and 40 months were employed. To ascertain the impact on costs, a discount rate of 6% was applied to monthly costs of £469 (the mean monthly cost of bone metastases). For QALMs, a utility value of 0.5 was discounted at a rate of 1.5%, whilst life years were not discounted. The ABC models covered a time horizon of 50 years, which was the maximum period over which the constant probabilities were applied to patients remaining in a metastatic state. The macro was set up to link the length of follow-up in the assumed metastatic state to the start year in the state. For example, if the start year in the state was sampled as year 10 then the outputs from the state were summed for the following 40 years in the state.

This set-up provided the most accurate portrayal of the effect of converting set survival times to constant probabilities within the ABC models.

Figures 9.1a, 9.1b and 9.1c represent the difference in outputs as a function of the survival time and the start year in the assumed health state. The differential effects of discounting, survival, and the time horizon of the model relative to the start year in the assumed state can be drawn from the figures. Figure 9.1a shows that the cost difference between the two modelling techniques, which was subject to the highest discount rate, was mainly affected by the length of survival. The maximum difference was almost £4,000 for patients with a survival time of 40 months entering the metastatic state in year 0 (DES model > Markov process). The cost difference decreased slightly as the start year increased, which reflected the increase in the discount factor. For short survival times of under 4 months the Markov process actually over-estimated costs, though only by a maximum of £38. Figure 9.1b demonstrates that with a low discount rate the start year in a state does not affect the difference in QALMs greatly. For example, the difference in QALMs between the two models for a survival time of 40 months is 0.113 (0.009 QALYs) if patients enter the state in their first year in the model, decreasing to 0.082 (0.007 QALYs) if patients enter the state in their 26th year in the model. Figure 9.1c appears to show the most significant results, but the magnitude of the effects on life years is small, the maximum difference is 0.015 months (roughly half a day). The most significant factor causing the differences in the lifeyear estimates was the time horizon of the model. This stylised example, following the assumptions employed in the case study Markov process, assumed that patients lived for a maximum of 100 years. This meant that patients entering the state, for example, in year 25 with a survival time of 40 months were not all exiting the model (and hence the state) before they reached the age of 100 years.

The data illustrates that, in the absence of discounting, the conversion of set survival times to constant probabilities has little effect, though combinations of longer survival times and late entry into the assumed state started to produce some differences. Around 50 to 60 per cent of patients experienced a metastatic relapse in the ABC models. The differential outputs caused by the conversion of set survival times to constant probabilities in the Markov process was a significant factor in the observed differences between the outputs of the two models.

Figure 9.1a Plot showing the difference in cost estimates between a DES model and a Markov process for survival from a state described as a set survival time, as a function of the survival time and the start year in the state.

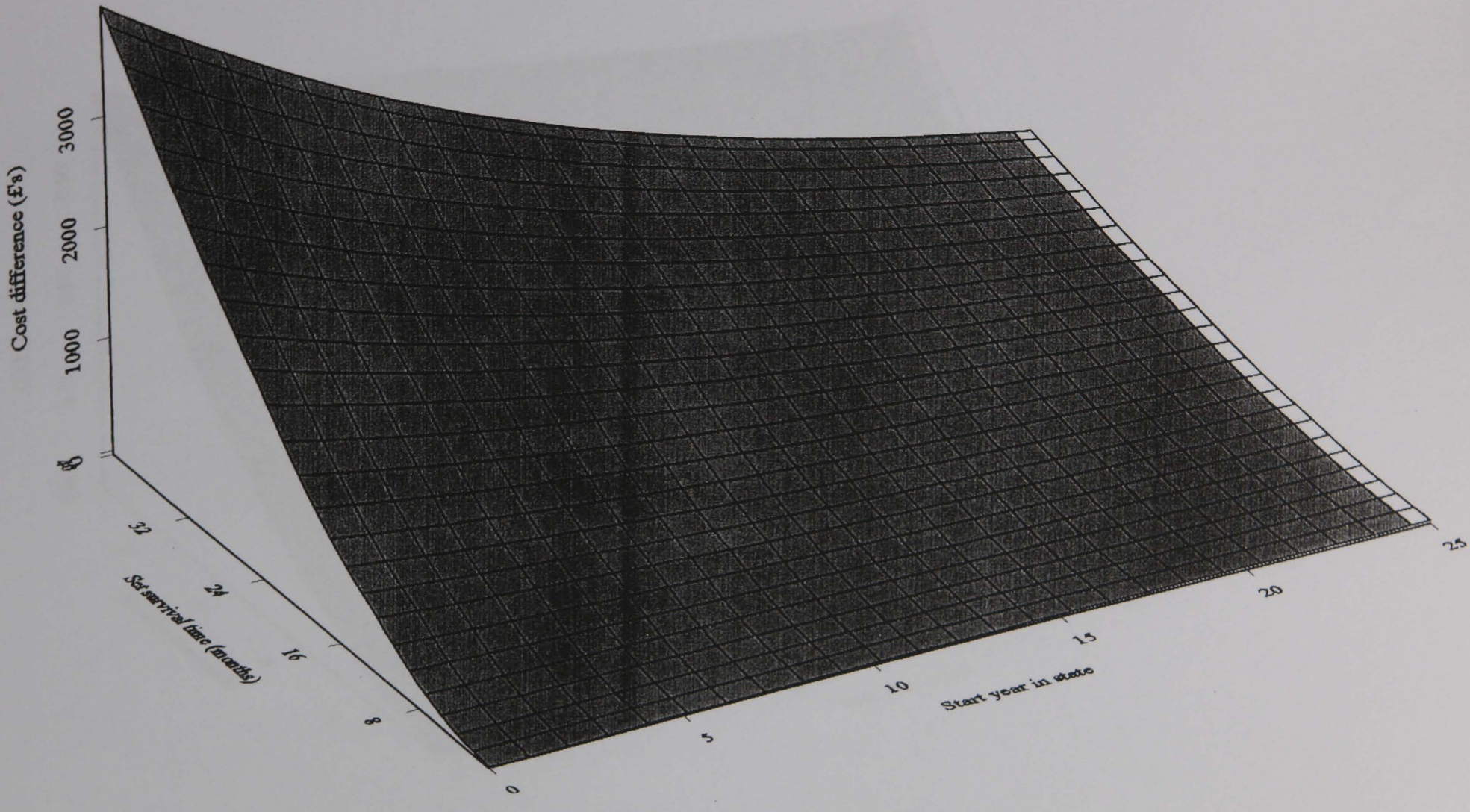


Figure 9.1b Plot showing the difference in QALY estimates between a DES model and a Markov process for survival from a state described as a set survival time, as a function of the survival time and the start year in the state.

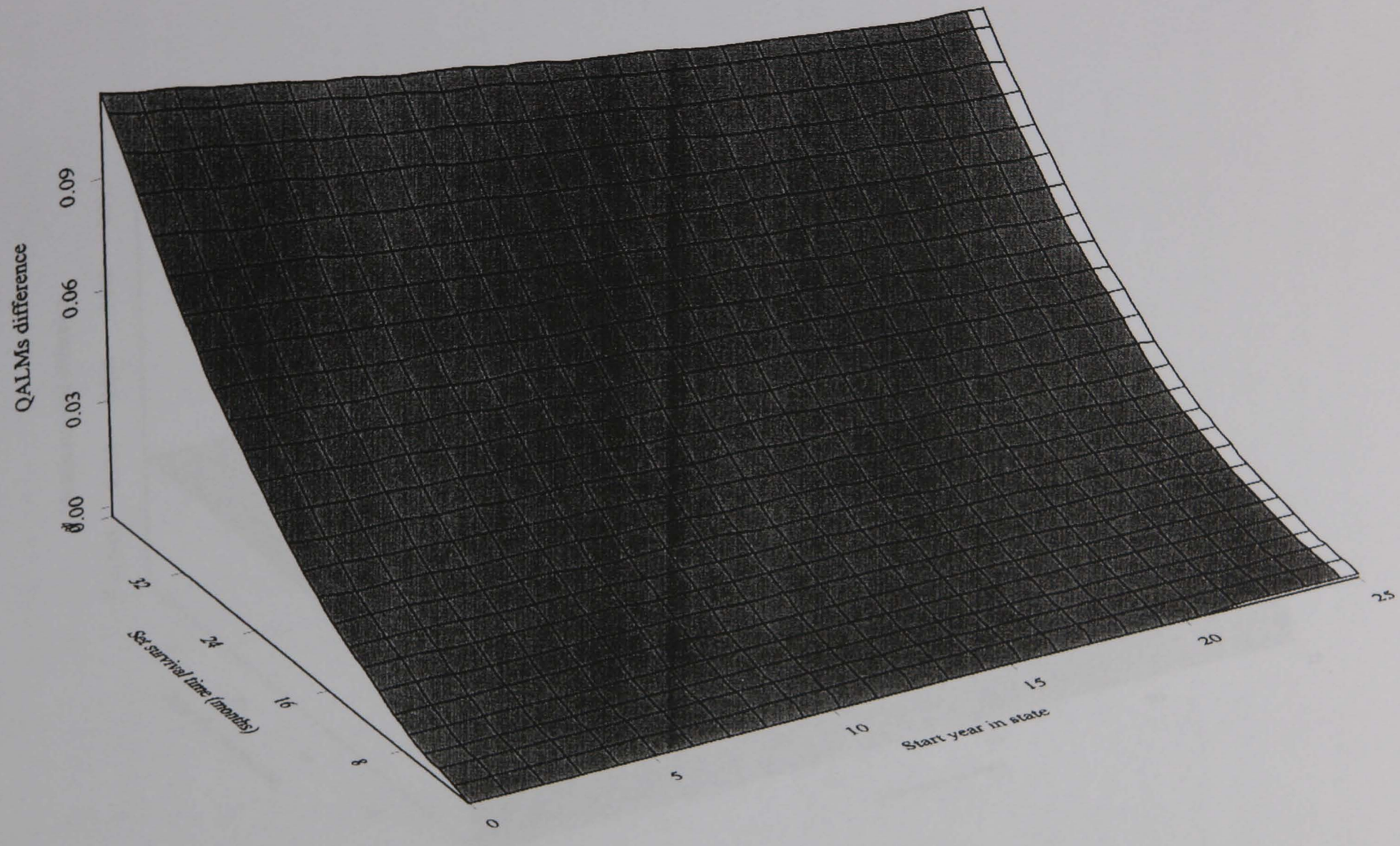
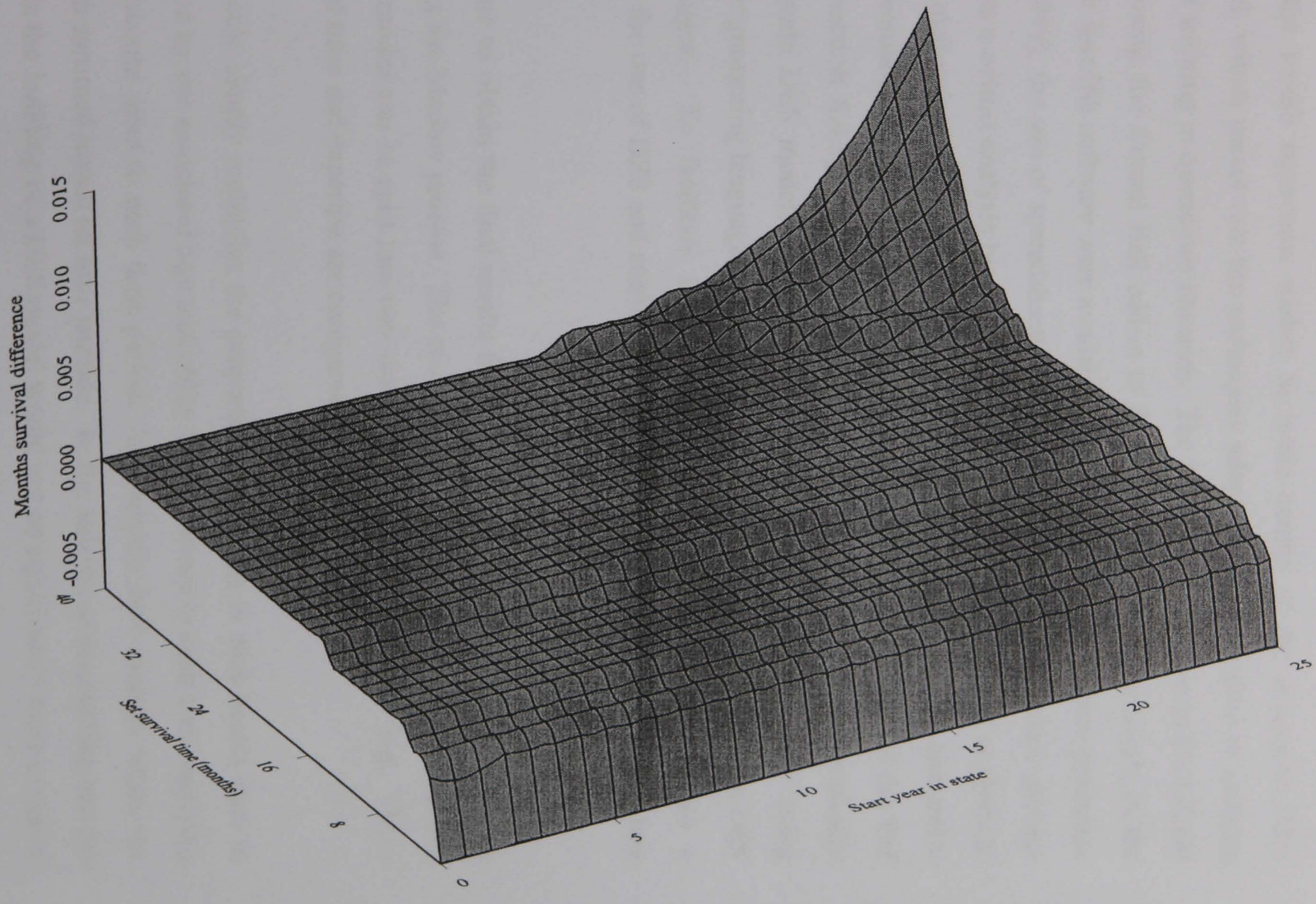


Figure 9.1c Plot showing the difference in life years estimates between a DES model and a Markov process for survival from a state described as a set survival time, as a function of the survival time and the start year in the state.



5. *The advantages of DES need to be weighed against the additional resource requirements. Realistic assessments of the necessary inputs should inform the choice of modelling technique.*

The comparison of modelling techniques undertaken in this thesis was based on the premise that the health economist wishes to retain control of the economic HTA decision model, which meant that the techniques adopted were accessible to analysts with no formal training in operations research. The Markov process was built in Excel spreadsheets, using the Crystal Ball add-in to facilitate the stochastic analysis of the model. Though specific software were available to build and analyse Markov processes [DATA 3.5, 1999], the use of spreadsheets is a common format. The Markov process was analysed as a cohort analysis because this is the most common form of analysis in the economic evaluation literature, but also because the more complicated first-order Monte Carlo simulation approach adds only to one relatively small element of the VoI analysis (see section 8.4.2). The DES model was built using software specifically designed to create DES models [Simul8, 2000]. DES models are often built using independent programming languages that have an even steeper learning curve than DES software packages. To facilitate further use of DES, software that provided a background to the use of DES and enabled the easiest construction of such models was used.

The overall time to obtain the final results from the DES model far exceeded the time employed using the Markov process. The causes of the excess analytic input associated with the DES model can be split into two categories - building and analysing. These two elements of time and expertise are compared below.

A Markov process simply multiplies the proportion of patients in each health state in each time period by the associated cost and utility value attached to each state, applying the relevant discount rates to each time period. To estimate their average value, the outputs are then summed across all time periods. In the absence of discounting and the use of attributes the building of a DES model would not be significantly more complex than a Markov process. The model would monitor when a patient entered a state and when they left the state, the length of stay in each state could then be multiplied by the associated costs and utility values. To discount the associated costs and utilities it is

necessary to distinguish the time spent in a particular state with reference to the time at which an individual patient entered the model. The time spent in a state must then be separated into the individual years since entry into the model for each patient.

DES is further complicated if attributes that affect the cost and utility values associated with health state are included. In the case study, toxicity was modelled as an attribute within the 'DFI' state. This meant that the time spent in the health state 'DFI' had to be separated into the different time periods within the state that individual patients spent with alternative forms of toxicity, or for which they were disease-free with no toxic effects.

A particularly complicated element of building the DES model was uncovered during the validation of the models (see section 7.3). Part of the data employed to validate the models described economic outputs for less than the full lifetime of patients. The use of these data required collecting data for all patients at exactly 4 and 10 years after they entered the model. To collect such data using the Markov process the model was simply stopped at the relevant period and the associated costs and effects collected. This was possible because all patients start at the same time in the model and the outputs are collected at regular time intervals – every month. The collection of model outputs at such specific time points was more complicated in the DES model because it is event-orientated. The collection of data within the model is also event orientated. As a patient leaves a particular state the programming code looks back at the history of the patient within the state she is leaving and assigns the relevant costs and utility values. The DES model could not simply cut-off the data collection period for patients at a specified length. In order to collect data constrained by time it was necessary to run the full model, but collect data at the end of each state up to the specified time period. This was only possible by inserting programming code that effectively asked if the patient had been in the model for the specified period at regular intervals when aggregating the costs and effects at the point of exit of each state. In effect, the process of validation was necessarily subject to a separate process of verification; to ensure that mechanisms put in place for the validation were working properly.

DES, being event-orientated, was necessarily analysed using first-order Monte Carlo simulations. This meant that for each set of input parameter values, data on a large

number of individual patients were collected to inform a mean value for each output. After experimentation, 10,000 patients were found to adequately control for first-order uncertainty, which was a function of the complexity of the model (see section 7.4). Employing cohort analysis, the Markov process estimated the model's outputs for the whole cohort simultaneously. The time required to solve the respective models for a single set of input parameter values using a 700Mhz PC was a couple of seconds for the Markov process, and over a minute for the DES model. In the context of a stochastic analysis, which involved solving the models for a large number of sets of input parameter values such a difference translated to weeks versus one day in terms of total running time for the final experimentation. However, the time to analyse included not only the final 'correct' experimentation, but also the whole process of verification and validation, which required significantly more analytic time than the final process of experimentation.

9.3 Comparing the alternative methods for assembling input distributions for stochastic decision models

The results presented in the Chapter 8 demonstrated that the most significant differences between the input data analysis methods were between the methods based on the creation of weighted datasets, which were employed to describe the empirical data direct and to inform fitted distributions, and the theoretically specified distributions. The following section discusses the causes and methodological implications of the observed differences in the mean values of the model outputs, whilst the subsequent section covers the substantial difference in the variation observed around the mean values estimated using the alternative methods.

More subtle differences were noted between the use of the empirical data direct and the fitted distributions, which are discussed in the third section.

9.3.1 Comparing the mean results using weighted datasets and theoretically specified probability distributions

The primary cause of the differences in the mean values of the model outputs between the use of weighted datasets and theoretically specified probability distributions relates to their respective handling of data presented as proportions. The main causes of the differences were the parameters for which measures of variance were available, i.e. the clinical parameters represented as proportions. To create weighted datasets for such parameters meta-analytic methods for weighting the data were used (see section 3.4.3.1). Using the fixed effects model (assuming no heterogeneity between the studies) the weight for a particular observation was the inverse of the variance. The variance for proportions is $p(1-p)/n$, which decreases as the proportion decreases, but as the weight is the inverse of the variance, the weight attached to individual data observations increases as the proportion of interest gets smaller. The chosen theoretically specified distribution to represent proportions was the beta distribution. The beta distribution parameters – α and β - are simply the number of events and the number of non-events, respectively, thereby applying equal weights to studies of equal sample size, irrespective of the number of events recorded.

Ceteris paribus, the meta-analytic techniques gave greater weight to smaller proportions, which led to smaller mean values for proportion parameters. To illustrate, the data representing the probability of patients receiving tamoxifen alone experiencing grade 3 or 4 toxicity is presented in Table 9.2.

Table 9.2 Data informing the proportion of patients receiving tamoxifen alone who experience grade 3 or 4 toxicity

	Study 1	Study 2	Study 3
(a) study sample	352	145	771
(b) number of events reported	1	7	8
(c) proportion of events reported	0.003	0.05	0.01
(d) meta-analytic weight [a/(b(1-b))]	117686	3053	77879
(e) proportional meta-analytic weight [d/sum(d)]	0.59	0.02	0.39
(f) implicit beta distribution weight [a/sum(a)]	0.28	0.11	0.61

The distribution parameters for a beta distribution representing such data were α = 16 and β = 1252, with a corresponding mean 0.0126 (16/1268). The process for weighting the data according to the fixed effects model employed the following formula:

$w_i = \frac{1}{v_i} = \frac{n_i}{p_i(1-p_i)}$, which led to a mean value of

$$\left[\frac{352}{0.003(1-0.003)} \times 0.003\right] + \left[\frac{145}{0.05(1-0.05)} \times 0.05\right] + \left[\frac{771}{0.01(1-0.01)} \times 0.01\right] = 0.00647$$

The beta distribution estimate is double that of the weighted dataset estimate. In addition to this general finding of divergence between the meta-analytic weighting techniques and the direct incorporation of data into probability distributions using method of moments formulae, other differences relating to the different types of input parameters were also noted. The following three sections outline such effects in parameters describing duration in a state, the type of event to be experienced, and parameters for which only a point estimate and a range are available.

9.3.1.1 Differences in the weighted values of duration parameters

In the case study evaluation, the mean duration in the DFI was lower in the analyses using the theoretical distributions than in the analyses based on meta-analytic techniques, which ran contrary to the general finding presented above. The data describing the probability of experiencing an event were presented as survival curves, from which the data were recorded as the proportion of patients remaining disease free at the end of each time interval. The apparent anomaly in the applied weights was due to the fact that weights were attached to the proportion of patients remaining disease-free at the end of each period. Using the meta-analytic techniques, higher weights were attached to data describing lower proportions of patients remaining disease-free, which led to higher weights for larger probabilities of experiencing an event.

Table 9.3 Example of difference in the estimation of the input data due to alternative methods of defining probability distributions

Study	A. Patients disease-free at start year x	B. Proportion disease-free at start year x	C. Proportion disease-free at end year x	D. Proportion experiencing an event [(B-C)/B]	E. Weighting patients remaining disease-free [A/(C(1-C))]	F. Weighting patients experiencing an event [A/(D(1-D))]
1	148	0.49	0.46	0.03	596	5086
2	148	0.44	0.35	0.09	651	1807

Table 9.3 presents an example of the difference in the mean probability of experiencing an event derived from weighting the data according to the proportion remaining disease-free, and to the proportion experiencing an event. At the start of year x, Study 1 reports the largest proportion of patients remaining disease-free, but during year x more patients' experience an event in Study 2. Applying the alternatively specified weights, the following weighted means were estimated:

Weighting patients remaining disease-free =

$$\left(\frac{596}{(596+651)}\right).0.03+\left(\frac{651}{(596+651)}\right).0.09 = 0.061$$

Weighting patients experiencing an event =

$$\left(\frac{5086}{(5086+1807)}\right).0.03+\left(\frac{1807}{(5086+1807)}\right).0.09 = 0.046$$

The observed difference in the weighted mean of 0.015 is a large difference in the context of the aggregate proportions (24.5% of 0.061 and 32.8% of 0.046). There were two reasons for basing the meta-analytic weighting procedures on the proportion of patients remaining disease-free, rather than the proportion of patients experiencing an event. Firstly, in time periods in which a study reported zero occurrences of an event the meta-analytic weighting procedure attempted to divide the sample by zero, which was obviously infeasible. Secondly, weighting data on the basis of the proportion of events occurring in each time period applied consistently higher weights to lower estimates of the probability of experiencing an event. This approach appeared unwise because higher relative weights were applied to whichever study, in a particular time period, presented the lowest estimate of the probability of an event in that period. Consistently weighting the lowest estimates most highly under-represented the true mean survival curve.

Using the proportion of patients remaining disease-free as the base for the weighting procedure did not consistently weight observations reporting fewer events more heavily, rather it weighted data according to the variance in the proportion of patients remaining disease-free.

9.3.1.2 Differences in the weighted values of type of event parameters

During experimentation with the case study evaluation it was noted that, relative to the theoretical distributions, the mean values of some events following DFI or remission (following locoregional relapse) were increased using the meta-analytic weights. This effect was due to the description of the types of event patients experienced as separate probability distributions for each event, which were sampled independently and the adjusted to equal 1 within the model. Though the meta-analytic weighting procedures assigned higher weights to lower probabilities for all such events, those events with the lowest probability of occurring - soft tissue metastases and death - were increased (relative to the theoretically specified distributions) in order to accommodate larger decreases in the mean values of the other events.

9.3.1.3 Differences in the weighted values for parameters informed by limited data

A third potential area of divergence between the alternative data input analysis techniques are parameters that are informed only by a point estimate and a subjectively defined range. In the case study evaluation, the utility values associated with each of the states included in the model were informed by point estimates obtained from the literature, but the ranges were subjectively defined after consultations with a range of health professionals. The minimum and maximum values that were specified were not necessarily symmetric around the derived point estimate.

The theoretically specified distribution chosen to describe utility values was the beta distribution. The distribution parameters for the beta distribution were calculated using a formula that required estimates of the mean and standard deviation of the distribution of values. The mean was taken as the point estimate derived from the literature, whilst the standard deviation was estimated from the specified range for each utility value (see Appendix 5). Triangular distributions were fitted to these data, taking the established point estimate as the most likely data point (the tip of the triangle). The remaining distribution parameters for triangular distributions were the minimum and maximum values of the distribution, which were informed by the subjectively specified ranges for each utility value. Unless the minimum and maximum values are equidistant from the most likely data point in a triangular distribution, the most likely value will not be the

mean value for the distribution. The non-symmetry of the ranges within some of the triangular distributions specified to describe utility values led to the differences in the mean values of some of the theoretical and fitted probability distributions.

9.3.2 Differences in the variation around the mean results using weighted datasets and theoretically specified probability distributions

The previous section described how differences in the mean values estimated for input parameters caused differences in the model outputs derived from the use of weighted datasets and theoretically specified distributions. This section illustrates how the alternative methods for pooling the identified data caused differences in the variation around the mean values for the input parameters, as reflected by the cost-effectiveness planes presented in Chapter 8. There was substantially more variation around the mean estimate of cost-effectiveness derived from the use of weighted datasets, than from the use of the theoretically specified distributions. Differences in the width of probability distributions were noted for parameters for which a measure of the variance around each estimate was available: the experience of different forms of toxicity, the probability of leaving DFI or remission, and the types of events experienced following DFI or remission. Figures 9.2a and 9.2b display a theoretically specified beta distribution and a normal distribution fitted from the weighted dataset, respectively, to represent the input parameter 'probability patient receiving tamoxifen and chemotherapy experiences major toxicity'. The fitted (normal) distribution is wider than the theoretically specified (beta) distribution because there is increased variation in the weighted dataset due to the larger weights placed on the outlier observations, in this case the smaller proportions.

Figure 9.2a Beta distribution(109, 1855)

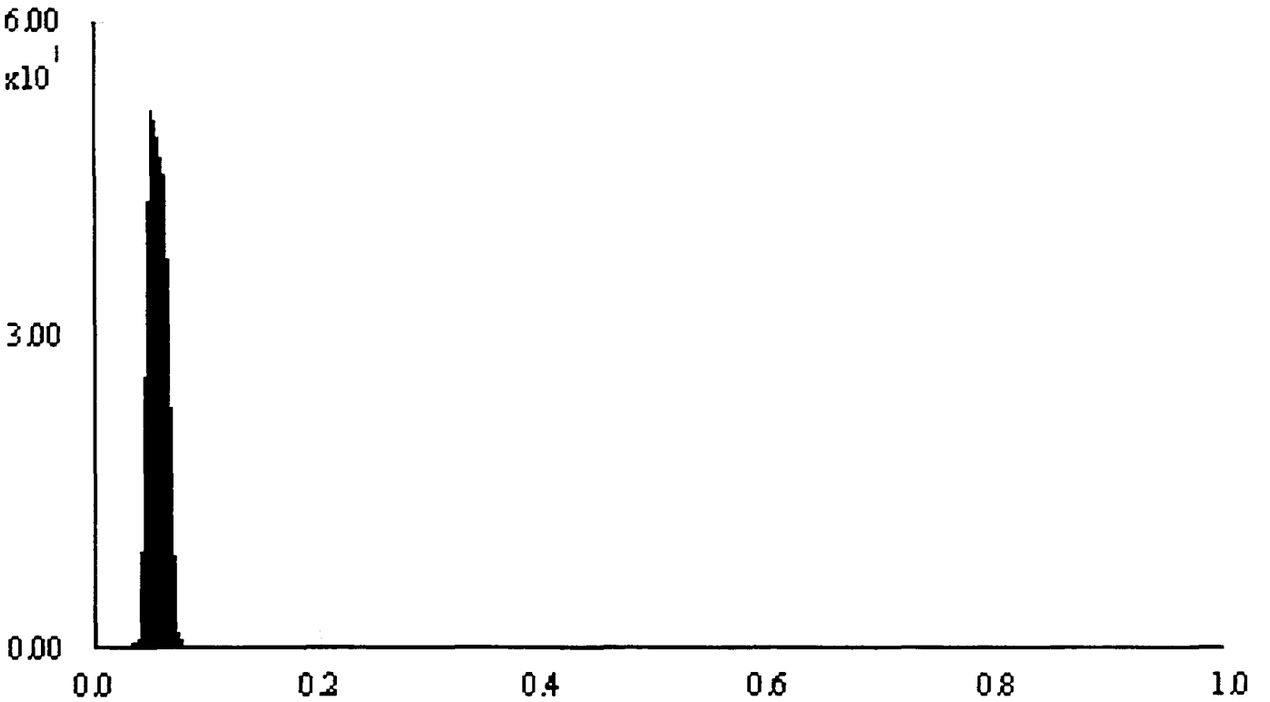
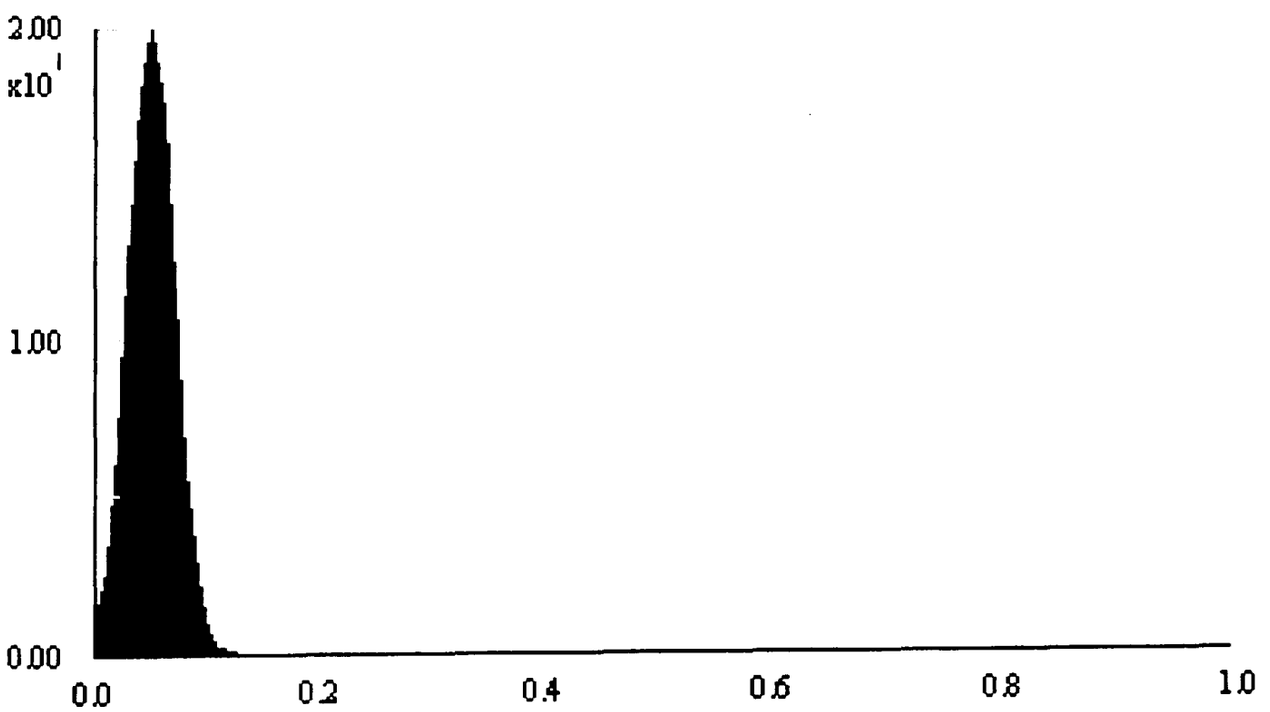


Figure 9.2b Normal (0.0499, 0.02)



9.3.3 Comparison of direct use of empirical data and fitted distributions based on the empirical data

The mean ICERs, the ICERs at the chosen credible intervals, and the cost-effectiveness acceptability curves, were virtually identical using the empirical data directly and the fitted distributions based on the empirical data. This closeness was expected as the fitted distributions had the same mean value as the empirical data and the variance in the empirical data was used to inform the distribution parameters. A difference was only noticeable between the two sets of observations when the differences in costs and effects were plotted on the cost-effectiveness plane (see Figures 8.1b and 8.1c). Though the same range of differences were covered by the two sets of observations the spread was more even in the results derived from the fitted distributions. Again, this was expected because the fitted distributions covered all values within the range specified by the empirical data, whilst only the observed values for the input parameters could be sampled using the empirical data directly.

In the case study evaluation the difference in the spread of observations did not cause any difference in the interpretation of the two sets of results. It is conceivable that more important differences may occur in evaluations where the empirical input data are mainly located at the extremes of the ranges. The mean results between the two approaches to specifying input distributions will remain equal, but the credible intervals may differ more significantly. This is especially likely if the decision-maker incorporates a lower credible range into her decision-making process, for example, a 90% limit rather than a 95% limit.

9.4 Value of information (VoI) analysis

The application of VoI analyses to economic HTA decision models is a recent methodological development - no full VoI analysis was identified in the literature, though a number of studies reporting the expected value of perfect information (EVPI) were identified [Claxton et al, 1998; Felli and Hazen, 1998; Fenwick et al, 2000]. The VoI analysis implemented in this thesis adapted a non-parametric approach to estimate the optimal sample size for a prospective study to inform all the input parameters within the defined models. The non-parametric approach benefits from requiring weaker assumptions than its' parametric equivalent. The examination of the VoI analysis

interprets the observed differences in the results of the VoI analysis between the two models, and between the alternative methods of specifying probability distributions. A further commentary on the methods used to evaluate the VoI and suggestions for further research in this area are presented in Chapter 10.

The first stage in the full estimation of the VoI involves the estimation of the expected value of perfect information (EVPI). In this thesis the EVPI was estimated using both decision models and for all three alternative methods for specifying probability distributions (see section 8.4). The Markov process and the DES model produced similar estimates of the EVPI, but the estimation of EVPI was massively different according to the choice of data analytic technique (see Figure 8.5). The fitted distributions and the empirical data direct resulted in much larger estimates of the EVPI than the theoretically specified distributions.

The EVPI derived from the weighted datasets rose continuously upwards in relation to the value of an additional QALY, whilst the EVPI associated with the theoretical distributions peaked at the value of a QALY at which tamoxifen alone had the highest probability of being the cost effective therapy option. This result reflected the wider dispersion of the net benefit observations using the fitted distributions and the empirical data direct. It is likely, therefore, that the use of meta-analytic techniques to analyse the identified data will always lead to an increased valuation of the VoI, over the use of theoretically specified distributions.

Unfortunately, the methodological tools for updating the fitted distributions and the empirical data, which was required for the next stage of the analysis, were not available at the time of undertaking the VoI analysis reported in this thesis. Therefore, the remainder of the analysis used only the theoretically specified distributions, which could be updated using established formulae (see Appendix 5). Comparing the alternative modelling techniques to the remainder of the VoI analysis, it appeared that the DES model was more precise in the estimation of the expected value of sampling information (EVSI) because it was possible to quantify first-order uncertainties (between patient variation), which were arguments in Neyman's formula. However, the impact of the increased accuracy was minimal as there was a similar level of variation within both therapy groups.

The estimation of the VoI necessitated the repeated stochastic analysis of the decision model, which required extensive running time for the DES model (weeks), whilst the whole process was completed in a day using the Markov process. The comparison of the results derived from the Markov process and the DES model showed that the expected net benefits of sampling (ENBS) were similar, though the DES model estimated a slightly higher optimal prospective sample size. The DES model placed a higher value on the collection of further data to inform the decision-maker. This result was expected as the increased flexibility of the DES model produced slightly more variation in the model's outputs. Indeed, the VoI estimated by DES models should always be at least as large as that estimated by a corresponding Markov process, because a DES model facilitates a more precise depiction of the true patient pathways, which can only lead to more variation in the model's outputs.

9.5 Conclusions

This Chapter has described the methodological implications derived from the results presented in Chapter 8. The primary objective of this thesis was addressed, using the five modelling characteristics statements, which were first defined in Chapter 2, to compare alternative modelling techniques. The first statement set up the choice between a Markov model and a DES model for evaluations covering extended time horizons. The following four statements were each found to have some impact on the results derived from the case study evaluation. The most important factor in terms of affecting the results of the evaluation appeared to be statement 4, which noted that increased data flexibility enabled more accuracy in the incorporation of data into the DES model. The improved accuracy was shown to have a sizeable impact on the differential outputs estimated by the Markov process and the DES model, particularly in the presence of discounting.

Another area of difference between the models was due to the enforced Markovian assumption, which meant that the Markov process was unable to define transition probabilities in terms of how long patients had remained in a particular state. In the case study evaluation, the probability of leaving the state 'remission' (following locoregional relapse) was a function of the time already spent in the state. The DES

model incorporated the exact slope of the disease-free survival curve, whilst the Markov process necessitated the estimation of a constant transition probability. In addition, the DES model allowed for a more realistic extrapolation of the available curve using age-specific mortality rates, whilst the data inputted into the Markov process was based on a single cut-off period for all patients. The DES model also enabled a more realistic assumption that patients would not experience a further relapse after an extended period of remission. Other areas in which the conditional development of patients pathways could lead to more significant differences, given appropriate data, were also identified. The representation of the relationships between the length of DFI and other parameters within the model could have a far larger impact on the models' outputs.

A less significant factor in the case study evaluation concerned the use of attributes in the DES model to represent the experience of treatment side effects, whilst the Markov process defined three separate toxicity states. The attributes did allow a more realistic representation of toxicity, though the impact on the results of the two models was minimal. The final issue related to the analytic input required to build and analyse the alternative modelling techniques. The DES required far greater time inputs to build and analyse the model. However, there is a steep learning curve in building DES models that, once surmounted, will reduce the gap between the analytic input required to build the respective models. The time to analyse the models is less open to improvement as the use of first-order Monte Carlo simulation to analyse DES models cannot be changed. The time to analyse the DES case study model was found to be the most restrictive element of DES because a large amount of time was required to verify and validate the model, and for the process of experimentation.

Following the assessment of the two modelling techniques this Chapter reviewed the methodological implications of using three alternative methods for pooling and formatting data for inputting to a stochastic decision model. Chapter 8 had highlighted the main differences as pertaining to the methods based on the creation of weighted datasets (fitted distributions and using the empirical data direct) and the specification of theoretically defined probability distributions. The major cause of the observed differences was the fact that the meta-analytic techniques used to create weighted datasets for some of the identified data (those for which a measure of variance was available) attached larger weights to smaller proportions than the alternative method. In

the case study evaluation, the impact of this effect was tempered due to anomalies observed in some data categories, in particular, data describing the probability of experiencing an event were weighted on the basis of the proportion of patients remaining event-free. The general conclusion arising from these comparisons is that the application of meta-analytic weights reduces the transparency of the process of pooling and formatting the input data. The use of meta-analytic weighting formulae based upon the variance within samples also caused the substantial difference observed in the spread of the outputs derived using the alternative methods.

The examination of the VoI analysis concentrated on a comparison of the above techniques (models and input data analysis methods) within the application of a VoI analysis. A further commentary on the actual process is presented in Chapter 10. The two models produced broadly similar estimates of the VoI. The ability of the DES model to estimate first-order uncertainty was employed in the allocation to therapy formulae used in the estimation of the EVSI to little effect, though the impact could be larger in other evaluations. However, the main difference between the models analysis of the VoI was the length of time required to obtain the results. The repeated estimation of the ENBS for a series of prospective samples took a few weeks using the DES model, but only a day using the Markov process. Bearing in mind that the first analysis is rarely the final analysis, the time constraints imposed by the use of DES cause more concern within a VoI analysis than for analyses to inform immediate resource allocation decisions.

Chapter 10 Conclusions and recommendations

10.1 Introduction

The aim of this final Chapter is to address the objectives raised at the start of this thesis by summarising the main findings reported in the previous Chapters, and defining recommendations regarding the application of decision models, as well as related areas for further research. The primary objective of this thesis has been the comparison of alternative decision modelling techniques, with respect to the process and outputs of an economic evaluation of health care interventions. The investigation was based on the application of the two main modelling techniques employed to model patient pathways over extended time horizons - Markov processes and DES models – to a case study evaluation comparing alternative adjuvant therapies for early breast cancer. The comparison of alternative modelling techniques was also presented in the context of a stochastic evaluation, which described probability distributions around the outputs of the models - the costs and effects associated with each intervention. The distributions of outputs were informed by randomly sampled sets of input parameter values from probability distributions representing the uncertainty around the values of the model's input parameters.

Three secondary objectives for this thesis were also defined relating to the application of stochastic decision models. The most general objective was the development of a complete description of the methodology informing the application of decision modelling to the economic evaluation of alternative health care interventions. Secondly, alternative methods for pooling and formatting the identified data into probability

distributions to populate decision models were compared. Finally, techniques available to estimate the value of information (VoI) were adapted and empirically applied.

The four objectives are addressed in separate sections. Each section presents the methodological insight gained from the development of a framework for the modelling process, the implications of the empirical evidence derived from the application of the defined methodology to the case study evaluation, and areas for future research.

10.2 Comparison of alternative modelling techniques

10.2.1 Methodological insight

Chapter 2 presented a general introduction to the three main modelling techniques – decision trees, Markov processes and DES models – supported by examples of the use of the different modelling techniques, which informed a preliminary assessment of the strengths and weaknesses of the alternatives. The characteristics of decision trees and Markov processes are very different and the choice between the two techniques in alternative treatment settings is relatively straightforward. With the introduction of DES to the field of economic evaluation in health care, the issue of choosing the appropriate technique could become an important decision in the initial stages of modelling projects. The issue of choosing the correct modelling technique has been referred to in the health economics literature [Chaussalet et al, 1999; Sonnenberg et al, 1994], but the consequences of the choice have not been fully explored.

On the basis of the review reported in Chapter 2 some criteria for the comparison of alternative modelling techniques were developed. Five statements were defined that related to the characteristics of the treatment area being modelled, or the type of data available to populate the model, which were used as the basis for the empirical comparison of the alternative modelling techniques (see next section).

10.2.2 Empirical evidence

Table 10.1 presents a summary of the differences observed between the DES model and the Markov process in the context of the case study evaluation. Despite the relative

closeness of the aggregate cost-effectiveness results derived from the two models, the structural differences between the models were potentially important. It would appear to be a matter of good fortune that the divergences between the models acted in opposite directions that almost cancelled each other out. However, the outputs estimated by the DES model were uniformly higher than those produced by the Markov process, which indicated that the most significant advantage of the DES model was the increased flexibility with respect to the handling of time to event data. Another important advantage of the DES model concerned the linking of parameter values to the time spent in states, which provided a more precise approach to the representation of other forms of data.

The choice of decision model should be judged on the characteristics of alternative treatment areas but, on the basis of these results, it is recommended that the use of DES should be strongly considered if either of the two issues highlighted in the previous paragraph (statements 2 and 4) appear relevant. Statement 3, relating to the use of attributes to reduce the number of health states included in a decision model, did not appear to have a great impact on the results of the case study evaluation. The additional analytic input to the use of DES can be addressed in two stages. Firstly, the increased analytic expertise is not as great as may be assumed, because specialist software packages are available that provide a framework for the development of DES models. Though the initial learning curve is steep relative to the development of a Markov process, most health economists who have used Markov processes should be able to build a DES model. Secondly, the additional time to analyse a DES model should only influence the choice of decision model if the results from an evaluation are required quickly and the additional time is simply not available.

One final recommendation is that the simultaneous application of both modelling techniques to a single evaluation provides an excellent basis for checking that both models are working correctly. Within the case study evaluation, the process of the verification was greatly aided by the ability to compare the respective outputs of the two models.

Table 10.1 Summary of differences between the case study Markov process and DES model with respect to the five modelling characteristics statements

	Statement	Model difference	Impact
1	*If anything other than short-term outcomes are to be modelled decision trees are an inappropriate choice of modelling technique.*	No difference, both models enabled the representation of extended time horizons.	-
2	*If model parameters are a function of the time spent in particular states DES will more accurately reflect the true relationships between health states.*	DES: inputted data from survival curves in exactly same format to describe disease free survival after experience of a locoregional relapse. Markov process: survival curve data was converted to a constant probability of experiencing event.	The mean length of DFS was longer in the DES model, but in the Markov process all patients leaving remission had a probability of experiencing a further relapse (metastases) because it was impossible to distinguish between patients within the health state 'remission'. In the DES model patients who remained disease free after 11 years were subject to the mortality rate in the general population and went straight to death from remission. It is likely that this led to lower costs in the DES model, though the effect on life years and QALYs is less clear.
3	*If the specification of similar health states that differ only with respect to the experience of previous states compromises the clarity of the model, the use of DES should be considered.*	DES: used attributes to describe simultaneous experience of separate toxicity categories. Markov process: patients could only experience toxicities independently.	Small increase in costs estimated by the DES model for patients receiving tamoxifen and chemotherapy because combined probability of toxicity can be greater than 1.
4	*If the data describing the timing of events are not in the form of transition probabilities then DES will provide a truer representation of reality.*	DES: survival from metastases was specified as set survival times, as described in the literature. Markov process: set survival times were converted to constant probability of experiencing death.	The Markov process underestimated costs and QALYs when survival times from the metastatic sites over 5 months were converted from set survival times to constant probabilities. The magnitude of the effect was influenced by discount rates. Assuming discounts rates of 6%, 1.5% and 0% for costs, QALYs and life years, respectively, the largest impact was on the costs estimated by the model. The likely aggregate impact on the ABC models is that the DES model produced higher estimates of all three outputs, but the effect will be greatest for costs, then QALYs and finally life years.
5	*The advantages of DES need to be weighed against the additional resource requirements. Realistic assessments of the necessary inputs should inform the choice of modelling technique.*	DES: increased analytic input required more time and expertise. Markov process: low level of expertise required, very fast running time.	There was a learning curve associated with the use of DES, though the use of specialist software reduced its' steepness. The analysis of DES models was substantially longer than for the Markov process. The experimentation time may be partly due to the use of specialist software.

10.2.3 Future research

The empirical comparison of the alternative modelling techniques presented in this thesis has highlighted three general areas in which DES provided an improved representation of the patient pathways associated with alternative adjuvant therapies for early breast cancer, given the available information. However, the model structure applied within the case study evaluation represented just one option. The findings reported in this thesis would be re-enforced if a separate evaluation of the same therapies using an alternative model structure arrived at the same conclusions. Moreover, the relative importance of each of the three issues is likely to differ between treatment areas. Further empirical comparisons of the use of Markov processes and DES models in evaluations of different treatment areas will provide additional evidence on the extent of the benefits offered by DES, as well as on the relative importance of the three identified areas of advantage.

Another potentially important area is the comparison of the alternative modelling techniques using patient-level data to describe some, if not all, of the input parameters within a model. This issue was referred to in section 9.2 (statement 2) where it was indicated that more detailed data might promote the advantages of DES further through the representation of relationships between patients treatment history and future prognosis. A comparison of a DES model and a Markov process using patient-level data to inform input parameters would provide further evidence on which to base the choice of appropriate modelling technique.

10.3 Developing a complete methodology for the modelling process

10.3.1 Methodological insight

There is a need for explicit guidelines to inform the use of modelling, especially in the context of submissions to official bodies, such as the National Institute for Clinical Excellence (NICE) in the UK. At present NICE simply specify that it is important that modelling is used appropriately and carried out to the highest standards. The development of a complete description of the necessary methods for applying an economic HTA model was borne of the realisation that guidance on the different

components were available, but that the advice was located in disparate sources. Furthermore, different sources of advice relating to the same stage of the modelling process recommended alternative procedures. Chapter 3 described the whole modelling process, in chronological order moving through five main stages, which were applied to the case study in the following Chapters:

1. Specifying the theoretical model (Chapter 4);
2. Undertaking a literature review to obtain input data from the model (Chapter 4);
3. Analysis of the available data to populate the model (Chapter 5);
4. Implementation of the model (Chapters 6 and 7);
5. Experimentation with the model (Chapter 8).

Previous work undertaken in the field of health economics was sought, though much of the process drew on issues common to the general area of clinical research. Insights into individual stages of the modelling process were also obtained from other disciplines including the social sciences and operations research.

10.3.2 Empirical evidence

This section summarises the experience gained from the application of the modelling process to the case study evaluation, incorporating various examples of good practice that were established. The review of the modelling process did not attempt to provide a prescriptive methodology for all economic HTA decision modelling projects, but rather to bring together the available methods relating to each element of the modelling process. The characteristics of the treatment area being evaluated, the available data, and the focus of the evaluation will influence the methods adopted for different parts of the process. To inform the use of the stated methods the accompanying theoretical and pragmatic arguments have been stated explicitly so the reader has a clear basis on which to base their own judgement.

In Chapter 4, the specification of a preliminary model structure was based on information obtained from oncology clinicians, a preparatory review of the literature, and the Internet, which limited the potential for bias because the structure of the model was informed by consensus between the alternative sources. The literature review was based on the preliminary model, which informed a series of separate

literature reviews relating to different aspects of the patient pathways. The conduct of the separate literature reviews was pragmatic. Explicit study inclusion criteria relating to year of publication, publication type and language were specified in order to limit the scale of the necessary review, though extensions to the criteria were enacted if insufficient data were identified to populate particular areas of the model. During the extraction of data from the literature any prognostic information reported about the identified patient groups were recorded in order to evaluate sub-groups of the population.

Chapter 5 described the procedures for pooling and formatting the identified data to populate a decision model. The first task involved the explicit harmonisation of data that described similar events, but differed slightly in their definition of the event of interest. Harmonisation facilitated the combination of data by making explicit alterations to the values reported by one or more of the relevant studies to improve the comparability of the underlying parameter definitions. The process of harmonising the data cannot be subject to hard and fast rules as the adjustments made to the data will depend on the event described and the format of the available data. However, the harmonisation of the ABC data illustrated a range of issues with potential relevance to other disease areas. For example, during the harmonisation of the DFI data, the studies that treated death as a censored event reported the number of deaths observed, which were then used to adjust the original disease-free survival data. If the relevant data to harmonise a parameter were not presented by a study, data from studies with the most similar patient and treatment characteristics were used to adjust the initial estimate.

Chapter 5 also described the implementation of alternative methods for the specification of probability distributions to represent the uncertainty around the values of the input parameters. The evaluation of this issue developed into a secondary objective of the thesis, which is discussed separately in the next section.

Chapter 6 described the development of a Markov process and a DES model as computer-based decision models. The relevant issues relating to building the alternative models were presented in the section above comparing the two modelling techniques. Chapter 6 also addressed two issues relating to the actual use of the

models. Firstly, the DES model was necessarily analysed using first-order Monte Carlo simulations, which required an adequate number of patients in each (first-order) run of the model. Testing alternative run sizes, 10,000 patients were found to be sufficient to minimise the impact of first-order uncertainty on the mean values estimated for each set of input parameters. However, it was noted that the necessary run size is linked to the complexity of the model and specific testing for the minimum run size should be undertaken for different models. Secondly, the stochastic sampling of monthly versus annual probabilities of experiencing an event was discussed. The Chapter concluded that the appropriate method of sampling should follow the intervals in the original data. For example, if the identified data are presented as annual probabilities they should also be sampled annually and then converted to monthly probabilities within the model.

Chapter 7 reported the application of methods to verify and validate decision models. Three categories of verification were employed. The main form comprised the verification of logic, which checked that the internal mechanisms of the models were working correctly. The explicit presentation of three classes of logic testing - clinical parameters, costs and utility values, and discounting - provided ample evidence of the analyst's attention to detail. Secondly, sensitivity testing compared expected and observed effects of alternative input parameters on the models' outputs. The sensitivity tests backed up the logic checks, but also confirmed that the two decision models were producing consistent outputs. Finally, stress testing verified that the models recognised nonsensical input data, alerting the analyst to data entry errors.

The validation of the decision models compared the models' outputs to a range of relevant outputs presented in previously published economic studies, as well as comparing clinical endpoints such as survival at different cut-off points. Reasons for any differences between the compared outputs were sought in the context of methodological and data driven differences between the case study evaluation and the identified studies. The process of validating economic HTA models is notoriously difficult and the methods described in Chapter 7 are not heralded as a definitive approach. However, the applied methods did result in the revision of the original values for a set of cost parameters within the case study evaluation, which demonstrates some success in this approach.

The outputs from a stochastic evaluation comprise a series of matched observations of the costs and effects associated with the relevant interventions. Chapter 8 presented the results of the case study evaluation using the alternative presentational forms presented in Chapter 3. Mean incremental cost-effectiveness ratios (ICERs) were supplemented by plotting the cost and effects differences on cost-effectiveness planes, the presentation of the credible intervals for the ICERs, and the use of cost-effectiveness acceptability (CEAc) curves. Each method of presentation added to the general interpretation of the model outputs, though the relative effectiveness of the alternative methods in aiding the comprehension of decision-makers is an empirical question beyond the scope of this thesis. The application of the final methodological area covered in the modelling process, the analysis of the value of information (VoI), developed into a secondary objective of this thesis and is discussed in the final section of this Chapter.

10.3.3 Future research

The collection of methodologies for the modelling process highlighted certain areas that would benefit from further research, either due to a lack of defined methods, or due to uncertainty regarding the appropriate methods to employ. The definition of economically relevant sub-groups is an important issue because it may be cost-effective to provide an intervention to a sub-set of a patient population, but not to the whole population. Recent recommendations issued by NICE have highlighted the importance of sub-group analyses, but research is required on methods for identifying which sub-groups should be analysed separately. Indeed,

Various methods were suggested for the validation of decision models that did provide some assistance in identifying parameters values that could be represented more realistically, though such modifications did not provide complete assurance of the validity of the models. Other authors have suggested that validity can only be proved to the point that the modelling process was undertaken to a sufficiently high standard [Sculpher et al, 2000; McCabe and Dixon, 2000], which appears to settle for a second-best solution. To increase the acceptance of economic HTA models, it is

necessary to explore this area further in order to convince potential users of such evaluations that the results really do answer the question asked.

The final area in need of further research follows on from the need to properly validate models, to the need to present the outputs of models in the most appropriate format for the user to be able to incorporate the information to their requirements. A wide range of possible formats for presenting cost-effectiveness information were described, but there is no real understanding of the best format, which will persuade decision-makers to actually use the available information. Future research is, therefore, required on how best to present cost-effectiveness information.

10.4 Comparison of alternative data analysis methods

10.4.1 Methodological insight

Four alternative methods for pooling and formatting the identified input data into probability distributions to populate stochastic decision models were described in Chapter 3. The first method, labelled 'theoretically specified distributions', assigned the same type of distribution to similar groups of input parameters. The distributions were chosen by matching the characteristics of the input parameters to the characteristics of different probability distributions. The parameters for each distribution were estimated by applying the identified data to established formulae that estimated the relevant distribution parameters.

The second and third methods employed meta-analytic techniques to weight the identified data, which were then described in datasets comprising weighted representations of the identified data for each input parameter. The weighted datasets were either inputted directly into the decision models ('empirical data direct'), or they were inputted into statistical fitting software, which fitted the data to the best matching probability distribution ('fitted distributions'). The final approach involved bootstrapping the identified set of observations to create a bootstrap distribution.

10.4.2 Empirical evidence

The bootstrapping approach was more applicable to the analysis of primary data and was not applied to the case study evaluation, but the remaining three methods were applied to the case study evaluation. Substantial differences were observed between the probability distributions specified using meta-analytic techniques to weight the available data (the fitted and empirical distributions) and the theoretically specified distributions for parameters for which a measure of variance within each identified study was available, i.e. parameters described as proportions that were informed by clinical trials or observational studies. The general effect of the meta-analytic weighting procedure was to give greater weight to lower proportions. Table 10.2 shows that the effect on the models' outputs of the different parameter categories was mixed, which meant that the respective outputs produced by the three methods did not vary greatly. However, the alternative methods did have a substantial effect on the variation described in the outputs of the models, which was far greater using the meta-analytic techniques.

Table 10.2 Summary of the differences in the mean values of input parameters due to the alternative methods of specifying probability distributions

Parameter category	Impact
Experience of toxicity	Probability of experiencing toxicity lower using meta-analytic methods, which led to higher costs and lower QALYs using the theoretical distributions. The magnitude of the differences in the mean values was small.
Time to event	Probability of experiencing an event was generally lower using theoretical distributions, which led to higher life years and QALYs, though impact on costs was more difficult to determine.
Type of event	Lower probability of progressing straight to death with no relapse in the theoretically defined distributions, particularly for patients receiving tamoxifen and chemotherapy. Resulted in higher costs, life years and QALYs using the theoretical distributions.
Informed only by a point estimate and a non-symmetric range	Higher utility values using theoretical probability distributions, because the range around the mean was assumed to be symmetric unlike the alternatively specified triangular distributions. This led to higher estimates of QALYs using the theoretical distributions.

It is not clear which data input analysis method provides the most reliable data, and hence, the results that engender the most confidence in their accuracy. The fact that the implications of the cost-effectiveness results arising from the alternative methods for specifying input distributions could differ raises the question of which method might be preferred. The work presented in this thesis suggests that there are two disadvantages to the use of meta-analytic techniques to weight the identified data over

the specification of theoretical distributions. Firstly, and most importantly, alternative approaches to the application of the meta-analytic weighting techniques were possible that increased the scope for differential results between analysts employing seemingly similar techniques. The only element that could vary between evaluations using the theoretically specified distributions is the choice of the theoretical distributions, which should converge to a common set of distributions as their application increases. There was no subjectivity apparent in the analysis of the distribution parameters, because well-defined formulae were employed to estimate the parameter values. Secondly, the analysis of the data to create weighted datasets is more complicated, and more prone to errors than the simpler process of estimating distribution parameters for the theoretically specified probability distributions.

If only one form of analysing the data is to be chosen, on practical grounds the general application of theoretically specified distributions is more likely to provide a common baseline for the comparison of alternative economic HTA decision models. It is also the only method that enables a full analysis of the VoI at present. Ideally, however, all three input data analysis methods should be employed in an economic HTA decision model, because this provides an excellent basis for the testing the sensitivity of the cost-effectiveness results to the alternative methods of populating a stochastic decision model.

10.4.3 Future research

The empirical evidence suggested that there is potential for far greater discrepancies in the estimated mean values between the alternative data analysis methods than was found in the case study evaluation. In addition, the observed difference in the variation around the cost-effectiveness results is a cause for concern. Though the theoretically specified distributions are most likely to provide a common baseline for comparisons between evaluations, this choice was based primarily on the ease of use and the ability to undertake full VoI analyses. The central question that remains to be addressed is which method is the most theoretically correct method?

10.5 Analysis of the value of information (VoI)

10.5.1 Methodological insight

The final element of the research reported in this thesis was the application of techniques to estimate the value of conducting further research in a particular treatment area. The main objective was to critique the underlying methodology in order to highlight areas that could benefit most from further research, though a couple of adaptations to the current methodology were developed within the context of the case study. The methods for the analysis of the VoI were presented in Chapter 3, incorporating a three-stage process: estimating the expected value of perfect information (EVPI), estimating the expected value of sampling information (EVSI), and estimating the expected net benefits of sampling (ENBS).

Both methodological alterations related to the estimation of the EVSI. The EVSI is based on the premise that hypothetical samples of data can be assumed that are analysed within a model to re-estimate the EVPI, which then informs the EVSI. The methods adopted in this thesis employed only two types of probability distributions, the beta and the gamma, which could incorporate hypothetical data using established formulae. The mean values of the new data for each parameter were assumed to be the same as the values derived from the identified data, because the mean represented the best estimate of the true value of the individual parameters. The hypothetical representation of additional data is a key issue in ongoing research. Secondly, Claxton stated that every feasible allocation of patients between the relevant therapy options, for each prospective sample, should be evaluated in order to estimate the optimal prospective trial [Claxton, 1999]. Given the time required to run the DES model, in particular, such a detailed level of analysis was not feasible. A simpler, and quicker, method of allocating a proposed sample between the treatment options being evaluated was adopted in this thesis, Neyman's allocation to strata formula, which accounted for the variance within the outputs of each intervention and the costs of sampling between the interventions [Cochrane, 1977]. A constant ratio of the sample allocation between two treatment options was estimated for all prospective samples.

10.5.2 Empirical evidence

The results of the VoI analysis derived from the alternative modelling techniques were relatively close, but the application of the alternative input data analysis methods within the first section of the VoI analysis – the estimation of the EVPI – revealed huge differences of magnitude. These differences would have persisted if all three methods had been able to estimate the full VoI, which reinforces the need to define a consistent method for the specification of probability distributions around the input parameters across all evaluations.

The revised methodology for the analysis of the VoI was applied to both decision models. The assumption of basing the hypothetical data on the mean values of the identified data estimated the minimum possible optimal sample size, as the assumed sample led to the maximum reduction in the variation described in the original probability distribution. Other options for the specification of hypothetical data to inform the estimation of the EVSI are currently being developed, which will hopefully provide a more solid base for the analysis, but also enable the use of a wider range of probability distributions than those employed in the case study evaluation. It was difficult to assess the impact of assuming a constant allocation ratio between the interventions, though it is likely that the optimal sample was reduced because the estimated net benefits for each sample were underestimated. However, the use of a constant allocation ratio reduced the complexity and duration of the VoI analysis significantly, which may encourage the application of further VoI analyses.

A final issue derived from the empirical evidence related to one of the most contentious elements of the modelling process - should the final structure of a decision model be influenced by the amount of data available to populate the model? The omission of qualitatively informed events or relationships in the structure of decision models is based on the definition of the practical model as 'the most detailed model that can be constructed given the limitations of available data' [Sonnenberg et al, 1994]pgJS54. The objective of most models is to inform a resource allocation decision based only on the data identified at the present time. However, VoI analyses aim to inform the collection of additional data so to restrict the structure of the model on the basis of the currently identified data appears to be nonsensical. It may,

therefore, be necessary to adopt different model structures to inform an immediate resource decision and to inform a VoI analysis.

10.5.3 Future research

The most important output from the application of the VoI analyses was the insight that it offered into how the technique could be improved. A number of issues arose during the VoI analyses that were beyond the scope of this thesis to explore, but which should be addressed during future research on this topic. The most prominent issue noted from the results of the VoI analysis concerned the impact of the assumptions made with respect to the size of the relevant patient population to whom the benefits of further information would apply. Previous estimates of the useful lifetime of information included 5 years for interventions to control the symptoms of urinary tract infection [Fenwick et al, 2000] and 3 years for treatments of Alzeimer's disease [Claxton et al, 1998], though no basis for these choices were provided. A baseline length of research application of 5 years was employed in this thesis, which was chosen on the basis that five years appeared to be a reasonable estimate of the time between the availability of new therapy options. Sensitivity analyses undertaken to test the impact of the assumed length of usefulness of the prospective research demonstrated that the magnitude of the VoI was extremely sensitive to the length of applicability. However, in order to persuade decision-makers of the relevance of VoI analyses in setting research agendas it will be necessary to develop standardised methods for the estimation of the length of research use in alternative treatment areas.

The length of time required to obtain additional information on specific parameters within the model will also affect the length of applicability of the research. The implicit assumption in the application of the relevant patient population to the VoI was that the prospective sample, n, consisted of the next n patients to require treatment for the condition evaluated. The subsequent patients were all assumed to benefit from the additional information provided by the n patients until the assumed length of research usefulness expired. Ideally, data on all the parameters within the decision model could be obtained within the same time horizon within individual patients; for example, from the point of entry to a trial the required data from a single patient would be available within one year. The estimation of the length of research

usefulness could then build in factors relating to the estimated time to recruit the required sample of patients, as well as the time to inform all parameters within the model.

Data describing events occurring later in the case study evaluation would only be available after time periods of up to 20 to 30 years, such as survival from metastases or events experienced following a locoregional relapse. To incorporate such effects into the VoI analyses would require the additional estimation of the VoI for different lengths of follow-up for prospective trials. The updated probability distributions could be estimated to reflect the assumed amount of data that would be available to inform alternative input parameters at differing lengths of follow-up. The relevant patient population would be re-estimated for every length of follow-up to reflect the revised estimates of the length of usefulness for the research.
The following function provides a foundation for the estimation of the length of usefulness of research (U):

$$U = t - r - f$$

Where t is the anticipated time to the availability of a new intervention, r is the time required to recruit prospective sample and f is the length of trial follow-up.

Other issues raised with respect to the analysis of the VoI encompassed means of making the analysis more precise, but also considerably more complicated. For example, the above discussion over the estimation of the usefulness of research assumed that the choice of intervention for patients who refused to enter a prospective trial would not be informed by the additional data obtained to the date at which their treatment decision is made. The VoI for non-participating patients could be included on the basis that the allocation decision would be updated at regular intervals within the follow-up period.

Additional data on some input parameters may be collected in trials involving other patient groups. For example, in the ABC models data relating to the experience of treatment side effects were assumed to be specific to the intervention rather than the

patient group. Data relating to such parameters as survival from metastases, or the cost of treating locoregional relapses could also be incorporated from studies set up to investigate treatment in other patient groups. The availability of such data should also be included within VoI analyses, which would require the identification of prospective trials and an assessment of the likely quantity and timing of the data to be obtained.

At this stage in the development of the analysis of the VoI, this area of research appears to be an exciting prospect that will offer useful information to decision-makers allocating resources to research, but many issues need addressing before such analyses are likely to be commonly employed for such tasks.

Appendix 1 Literature review details

A1.1 Introduction

Two broad literature reviews were undertaken, which are described in the following sections. The first review identified work that informed the methodology for the modelling process, whilst the second identified data that could be used to populate the decision models used in the case study evaluation.

A1.2 Review of the modelling process

A large-scale review of the health economics and clinical literature was undertaken to identify research describing any methodological issues involving the application of decision models to clinical evaluation. Relevant research was also sought from disciplines such as operations research and social sciences. The purpose of the literature review was to inform the discussion of the characteristics of the alternative modelling techniques (see Chapter 2), and to identify research relating to individual aspects of the modelling process (see Chapter 3).

The literature review included orthodox searches of the main medical databases (Medline and EMBASE). No relevant index terms were identified so the search of the databases comprised a range of free-text terms: model*, decision analysis, decision trees, Markov, simulation. In addition, manual searches of the most prominent journals were undertken.

including Medical Decision Making, Health Economics, Pharmacoeconomics, Journal of Health Economics, the International Journal of Health Technology Assessment in Health Care and the Journal of Health Services Research and Policy. The references lists of the identified studies were then searched for further studies, and citation searches for leading authors were conducted using the Science Citation Index. Internal publications from academic departments and papers presented at relevant conferences and workshops also proved to be an important source of additional work.

OR textbooks were included in the review for general insights into the comparison of the modelling techniques. Numerous sources of advice were identified from Brunel University library's catalogue and a selection of textbooks was chosen to provide an alternative perspective on the modelling process. Social science texts were sought for further advice on the conduct of literature reviews.

A1.3 Review to populate the decision models

To identify data to populate the case study models Medline was searched initially in May 1998 using the following search terms:

- Index terms, Breast neoplasms/, Disease-Free Survival/, Survival/ or survival analysis/ or survival rate/, Mortality/, Recurrence/, Neoplasm recurrence, local/, Lymphatic metastasis/, Neoplasm metastasis/, Neoplasm metastasis/, Tamoxifen/, Chemotherapy, adjuvant/, Drug therapy/, Ovariectomy/, Menopause/, Menopause, premature/, Castration/, Postmenopause/, Drug toxicity/, Quality of life/, Costs and cost analysis/
- Free text terms, early, site, first recurrence, adjuvant, ovarian, suppression, ablation, ovarian suppression, ovarian ablation, estrogen replacement therapy, utility

Some examples of the searches made using these terms, including the term combinations and associated number of hits, are presented below. The identified reference lists and the Science Citation Index were also searched to obtain further references.

A1.4 Literature search results

Tamoxifen and survival

1	Breast neoplasms/	25074
2	limit 1 to (human and english language)	18451
3	Survival/ or survival analysis/ or survival rate/	43428
4	limit 3 to (human and english language)	34944
5	Tamoxifen/	3350
6	limit 5 to (human and english language)	2381
7	2 and 4 and 6	165
8	from 7 keep	114

****

Tamoxifen and disease-free survival

1	Breast neoplasms/	25074
2	limit 1 to (human and english language)	22019
3	Disease-Free Survival/	3147
4	limit 3 to (human and english language)	2901
5	early.ti,ab,sh.	160807
6	limit 5 to (human and english language)	93442
7	Tamoxifen/	3383
8	limit 7 to (human and english language)	2399
9	2 and 4 and 8	47
10	from 9 keep	32

****

Chemotherapy and survival

1	Breast neoplasms/	25074
2	limit 1 to (human and english language)	22013
3	Survival/ or survival analysis/ or survival rate/	43428
4	limit 3 to (human and english language)	34944
5	Chemotherapy, adjuvant/	5044
6	limit 6 to (human and english language)	3662
7	early.ti,ab,sh.	160807

8	limit 9 to (human and english language)	93442
9	2 and 4 and 7	255
10	2 and 4 and 7 and 10	51
11	from 12 keep	32

****

Chemotherapy and disease-free survival

1	Breast neoplasms/	25074
2	limit 1 to (human and english language)	22019
3	DISEASE-FREE SURVIVAL/	3147
4	limit 3 to (human and english language)	2901
5	2 and 4	380
6	early.ti,ab,sh.	160807
7	limit 6 to (human and english language)	93442
8	Chemotherapy, adjuvant/	5044
9	limit 8 to (human and english language)	3662
10	2 and 4 and 9	86
11	from 10 keep	65

****

Site of relapse

1	Breast neoplasms/	25074
2	limit 1 to (human and english language)	21076
3	Recurrence/	30651
4	limit 3 to (human and english language)	23144
5	site.ti,ab,sh.	122394
6	limit 5 to (human and english language)	53315
7	first recurrence.ti,ab,sh.	213
8	limit 7 to (human and english language)	188
9	2 and 4 and 6	8
10	2 and 8	25
11	9 or 10	32
12	from 11 keep	20

****

Survival from relapse

1	Breast neoplasms/	25074
2	limit 1 to (human and english language)	21076
3	Recurrence/	30651
4	limit 3 to (human and english language)	23144
5	site.ti,ab,sh.	122394
6	limit 5 to (human and english language)	53315
7	first recurrence.ti,ab,sh.	213
8	limit 7 to (human and english language)	188
9	2 and 4 and 6	8
10	2 and 8	25
11	9 or 10	32
12	from 11 keep 1-4,6,9,14,16-21,23-25,28-31	20
13	Survival/ or survival analysis/ or survival rate/	43428
14	limit 13 to (human and english language)	34944
15	Neoplasm recurrence, local/	7212
16	limit 15 to (human and english language)	5395
17	Neoplasm recurrence, local/mo [Mortality]	942
18	limit 17 to (human and english language)	649
19	2 and 14 and 16	198
20	2 and 17	99
21	from 11 keep 31	1
22	from 20 keep	68
24	Lymphatic metastasis/	10474
25	limit 24 to (human and english language)	7225
26	Neoplasm metastasis/	7728
27	limit 26 to (human and english language)	5273
28	2 and 14 and 25	499
29	Mortality/	3897
30	2 and 24 and 29	1
31	from 20 keep 91,93-96	5
32	from 30 keep 1	1
33	or/31-32	6
34	2 and 26 and 29	0
35	2 and 14 and 27	212
36	from 35 keep	137

Quality of life

1	Breast neoplasms/	25074
2	Quality of life	
3	1 and 2	671
4	limit 3 to (human and english language)	210
5	Adjuvant therapy	
6	1 and 2 and 5	35
7	Chemotherapy	
8	1 and 2 and 7	67

****

Appendix 2 Data extraction methods

Using previously published examples as a framework [Dowie, 1996; Murphy et al, 1994], an evidential database was set up to collect and manage the relevant data obtained from the research identified through the literature review. Three main information management criteria were identified as relevant to the case study evaluation:

- an ability to download references from other electronic databases,
- a facility to search for, and retrieve references stored in the database, and
- a facility for appending text to references.

At the time of the current project the software developed by ProCite had been improved so that it met each of the three criterion listed above to a high standard. A twin software package – Bibliolink 2 – had been developed that enabled the automatic transfer of records found by most electronic databases, whilst searching within a database created in ProCite was straightforward [Procite 5, 1999]. Most importantly workforms could be easily adapted for data entry and formatted output.

Using the flexible workforms, a data extraction form was developed in ProCite that categorised the identified data according to the model parameters that they informed. Figure 4.1 (section 4.4.4) presents the extraction form. The first section of the forma captured the reference details of the papers, as well as the objectives of the paper and the characteristics of the patient group included in the study. The aim of a paper provided an immediate indication of the type(s) of data (relating to the model parameters) that might be found in the paper, as well as allowing a judgement to be

made on the authors success in meeting the original objective of the study. Careful documentation of the patient characteristics within a study was essential in order to ensure that the analyses of the model were as homogeneous as possible. Information on the location of the study and the time period were also recorded as potential sub-groups.

Variable definitions referred to any definition given in a paper that might vary across studies, covering such topics as estrogen receptor status [Zambetti et al, 1992], types of toxicity [Alonso et al, 1995], and types of relapse [Kamby and Sengelov, 1997]. The category 'omissions in context' covered all additional information about the study that was not described in the previous sections, such as the protocol restrictions on adjuvant therapies in trials of treatment for metastases [Falkson et al, 1995], or missing information such as the number of cycles of chemotherapy undertaken [Senn et al, 1997] or types of costs excluded [Lober et al, 1988]. The 'Comments' category incorporated a wide range of information including the conclusions drawn by each study. Also captured in this category were comments relating the current results to previous studies [Swedish breast cancer cooperative group, 1996] or any general statements about breast cancer that were not relevant to any of the other categories, such as hypotheses about the spread of cancer cells [Boccardo et al, 1997]. The assessors' category simply notes the assessor (all JK) and the date on which the study was reviewed.

The second section of the extraction form covered all the parameters in the model. The parameters were split into three types that were identified by their prefix – 'id' covered clinical parameters, and 'qu' related to quality of life data. The 'co' category collected any available data on the cost of treatments related to breast cancer, but it also contained any relevant information on the type and frequency of resource use associated with the different stages of breast cancer. The category titles are mostly self-explanatory relating to the parameters included in the preliminary model. 'Disease free interval' and 'rates of relapse' both cover the probability and timing of experiencing a relapse or death with no evidence of disease, reflecting the two main formats of these data. The category 'overall survival' captured data describing survival from the point of diagnosis with early breast cancer.

Within the established categories data entry was left unstructured. This flexible form of data entry was used because of the need to capture both quantitative and qualitative data. As there was only one reviewer, who was involved with the mechanics of the model, the style of extraction remained similar and was geared towards the type of data required for the model. The data extraction form created in ProCite was adequately tractable with respect to the format of data that was entered. There was no limit to the size of each category. Tables of data had to be tab-created in Wordpad and pasted into ProCite, which sometimes led to messy tables due to the width restriction in ProCite.

The final category listed – Parameter Keywords – collected the categories that contained data for each paper included in the database. For example, a paper presenting disease-free interval curves and the proportion of patients experiencing different types of toxicity would be marked 'idDFI/ idTox'. Identifying each paper according to the data categories they contained made searching the database for particular parameter information extremely simple.

Though the review was ongoing in the sense that new studies were reviewed and added to the database as they were identified, the analysis of the data extracted began in May 1999. Firstly, qualitative data contained in the three categories 'variable definition', 'omissions in context' and 'comments' were assessed to provide background information on the disease area prior to commencing the quantitative analysis of the individual parameters. Next the database was searched separately for each parameter by entering the necessary keyword in the 'Parameter Keyword' category. These searches picked out the papers in the database that held data on the parameters of interest. Quantitative data was exported to an excel spreadsheet for analysis.

Appendix 3 Meta-analytic formulae for weighting data

The methods available for weighting data depend on the format of the identified data. If good primary data that incorporates measures of the within sample variance are available then methods for weighting data based on meta-analytic methods may be employed. The fixed effects model should be used if there is limited heterogeneity between the data, whilst the random effects model attempts to control for assumed heterogeneity. If it is unclear whether heterogeneity is present between the identified studies, then a formal test can be applied that tests the hypothesis that the underlying effects are equal:

$$Q = \sum_{i=1}^{k} w_i \left(\theta_i - \bar{\theta}\right)^2$$

Where k is the number of studies, θ_i is the treatment effect of the individual study and $\bar{\theta}$ is the mean value of all the studies. The null hypothesis is rejected if the observed variance is higher than would be expected due to chance alone – if value of Q exceeds the critical value of the chi squared distribution associated with $k-1$ degrees of freedom. It is recognised that this test lacks power and the chance of falsely defining studies as homogeneous is quite large [Sutton et al, 1998]pg77.

In heterogeneity is not identified, the fixed effects model involves a simple method of weighting the constituent data using the inverse of the variance [Sutton et al, 1998]pg55. The formulae for calculating weights are:

$w_i = \frac{1}{v_i} = \frac{n_i}{p_i(1-p_i)}$, for input parameters defined as proportions, and

$w_i = \frac{1}{v_i} = \frac{n_i}{\sigma_i^2}$, for input parameters measured on a continuous scale.

The random effects model also weights the different data according to their variance, but as well as the variance within the studies, the random effects model incorporates variance between studies to control for the assumed heterogeneity. There are various procedures for combining data assuming random effects [Hasselblad and McCrory, 1995], though the most commonly used method appears to be the weighted non-iterative approach, probably due to its relative simplicity [Sutton et al, 1998]pg72. This approach consists of the following formulas:

$$w_i^* = \frac{1}{\left[\left(1/w_i\right) + \hat{\tau}^2\right]}$$

Where $\hat{\tau}^2 = \left(Q - (k-1)\right)/U$

and $U = (k-1)\left(\bar{w} - \frac{s_w^2}{kw}\right)$

and $\bar{w} = \sum_{i=1}^{k} w_i / k$

and $s_w^2 = \frac{1}{k-1}\left(\sum_{i=1}^{k} w_i^2 - k\bar{w}^2\right)$

Appendix 4 Analysing survival curves

A4.1 Introduction

Data on 'time to event' parameters in the literature are often presented as survival curves, including disease-free or event-free survival curves. The process of collecting, analysing and inputting such data into decision models comprises three separate stages – collating the data from the literature, combining the data, and transforming the data to an appropriate format for inputting into a decision model. The following sections describe the process of incorporating data described in the form of survival curves into decision models.

A4.2 Collating the data

Firstly, the survival estimates and the respective sample sizes must be collated. In cases where only graphical data are available the data represented by the curves can either be read directly off the presented curves or digitised images of the survival curves can be imported into a graphics package enabling a more exact measurement of the proportions [Earle and Wells, 2000]. The proportions experiencing an event during each time interval must be revised to reflect the proportion of patients remaining event free. To explain: the proportion of patients leaving health state A in any given period is read from survival curves as the proportion of the original patient cohort (in year 0) who experience an event. In a decision model, however, only the patients remaining in state A at the start of each time period sample from the assigned probability distribution describing the proportion of patients experiencing an event in the ensuing period. Applying the probabilities that relate to the original cohort to the cohort of patients

remaining in state A will systematically underestimate the number of patients leaving the state.

Detailed information on censored data is rarely published in journals. To obtain a proxy for the numbers at risk for each time interval the following formula transforms the estimates from describing the proportion of the original cohort that leave a state to reflect the probability of the remaining cohort leaving a state. This approach has been shown to be the most reasonable assumption in the absence of information on censored data [Earle and Wells, 2000]:

$$P[event/remainingcohort]_i = \frac{P[event/originalcohort]_i}{P[remaining/originalcohort]_i}$$

where $P[event/remainingcohort]_i$ is the probability that a patient remaining in a state in year i will experience an event in year i;

$P[event/originalcohort]_i$ is the probability that a patient in the original patient cohort will experience an event in year i;

$P[remaining/originalcohort]_i$ is the probability that a patient in the original patient cohort remains event free at the beginning of year i;

A4.3 Combining the data

Earle and Wells [2000] presented and compared five possible methods for combining published survival curves. It was noted that the five methods produced similar results up to the point at which the trial with the shortest length of follow-up finished. The combination of trials with variable lengths of duration is common and a recommendation to restrict the combined curve to the duration of the shortest trial is unrealistic, especially for their incorporation into decision models. The preferred method when it is important to estimate survival beyond the shortest duration was termed the 'meta-analysis of failure time data' method (MFD). MFD involves pooling the number of patients at risk and the number of events in each time interval, from which the hazard and survival functions in each interval can be calculated. The process of pooling the data incorporates a weighting procedure based on the sample sizes of the respective studies. MFD was least affected by the need to extend the period of analysis

because it recalculates the number of subjects at risk at every time interval [Earle and Wells, 2000].

A4.4 Formatting the data

Undertaking a MFD of alternative survival curves weights the data appropriately but it produces only a single aggregate survival curve. No measure of variation is produced that can be used to represent the uncertainty in the survival estimates. However, the MFD method can be adapted to either create datasets that comprise weighted probabilities of experiencing an event within specified time intervals, or to estimate the relevant parameters to fit beta distributions that inform the probability of experiencing an event in each time period. To establish weighted datasets, the number of patients at risk and the number of events in each time interval reported by each identified study are documented separately to produce a range of estimates of the proportion of patients experiencing an event in each time interval. The individual estimates are weighted using the meta-analytic formulae presented in Appendix 3. More simply, the alpha and beta parameters for beta distributions are estimated as the number of patients experiencing an event and the number not experiencing an event within a period, respectively.

Appendix 5 Bayesian distribution theory

A5.1 Introduction

This appendix describes the process of specifying probability distributions for different types of parameters on the basis of theoretical considerations and presents the formulae employed to estimate the distribution's parameters using the available data, employing method of moments formulae. The process of updating the probability distributions derived from the available data to reflect the assumed impact of additional samples (as employed in the value of information (VoI) analysis) is also described.

Both processes are grounded in Bayesian methodology, whereby the estimation of the input distributions combines a prior distribution with the additional information to produce a posterior distribution for the input parameters. To estimate the baseline probability distributions a non-informative prior distribution is usually assumed, which is updated with the data obtained from the literature review. For the VoI analysis the original probability distributions are treated as the prior distributions and are updated under the assumption that new data will not alter the mean value of the distribution, only the level of variation described by the distribution.

The probability distributions that best represent different groups of input parameters are presented in the following sections. The first describes the specification of the *beta* distribution as a suitable representation of all input parameters that take the form of a proportion bounded by 0 and 1. The following sections link other clinical parameters

and cost parameters to the gamma distribution, whilst utility values are suitably described by a beta distribution.

A5.2 Proportion parameters

The following description is adapted from a text on Bayesian statistical inference [Iverson, 1984]. For the analysis of a population proportion (π) it is particularly convenient to look at functions that are polynomials in π. Polynomials in π can be written as a product of three parts: a numerical constant (C'), π to some exponent, and 1-π to some exponent. The general formula is:

$$f(\pi) = C'\pi^{a-1}(1-\pi)^{b-1}$$

Which is the general expression for the *beta* distribution. The mean and standard deviation of the assumed distribution of π can be used to estimate the distribution parameters, a and b, using the formulae:

$$a = \mu\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right] \text{ and } b = [1-\mu]\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right],$$

which are calculated from the fact that:

$$\mu = \frac{a}{a+b} \text{ and } b = \frac{\mu(1-\mu)}{a+b+1}$$

These calculations may not estimate integer values for a and b but it is usual to round off the values to the nearest integer. If the relevant data represents a series of patients who have either experienced an event, or not, then a simpler method of estimating the parameters for the beta distribution invokes the Bayesian use of a non-informative prior distribution. If no prior information is available a uniform prior can be assumed with beta distribution parameters $a = 1$ and $b = 1$. The beta distribution parameters are '$x+a$' and 'n-x+b', where x is the number of observations with the event of interest and n-x the number without the event.

A5.2.1 Updating the beta distribution

To update the beta distribution to reflect additional data in a VoI analysis the mean value from the relevant sample is added to the original a and b proportionately. For example, the parameters for a prior beta distribution informed by a meta-analysis of a series of trials that reported 100 events in 300 patients would be $a = 100$, $b = 200$. If the mean value from a prospective sample of 100 is 0.25, the updated distribution parameters can be updated proportionately, i.e. $a = 125$, $b = 275$.

A5.3 Non-proportions clinical parameters

All clinical parameters within a Markov process are described as proportions, but in a DES model some clinical parameters may be described as event times. For example, in the case study DES model survival from metastases was described as the median length of survival, rather than the probability of dying in successive time periods. Discussing the specification of a posterior distribution for the mean value of a variable between any two extremes Iverson assumed that the variable has a normal distribution [Iverson, 1984]. Assuming a non-informative prior distribution, the posterior distribution is informed solely by the mean and standard deviation of the observed data. However, the normal distribution is not bounded by zero and specifying a normal distribution for survival parameters may result in negative values being sampled. Alternatively, the gamma distribution is bounded by zero and provides a more flexible range of shapes for describing non-negative random variables [Rice, 1995]. The gamma distribution has appealing properties in that the distribution parameters α and β are linked to mean and variance [Berry and Stangl, 1996]:

$$\mu = \alpha / \beta \text{ and } \sigma^2 = \alpha / \beta^2$$

The distribution parameters can be back solved after specifying a mean and a variance for the specified distribution.

A5.3.1 Updating the gamma distribution

The easiest model for the likelihood function for event times is the Poisson distribution because it resembles a gamma density. The gamma distribution is a conjugate prior

when the observations are Poisson distributed. This means that the posterior is also a gamma distribution with updated parameters $\alpha + \sum t_i$ and $[n+1/\beta]^{-1}$, where n is the increased sample and $\sum t_i$ is the sum of the survival times for the increased sample [Berger, 1980]. The original mean value can simply be used to fit the updated α parameter as $\sum t_i = n\bar{t}$.

Within the case study evaluation, the updated parameters of the gamma distribution $\left[\alpha + \sum t_i, [n+1/\beta]^{-1}\right]$ were estimated assuming that the hypothetical data was Poisson distributed, which assumed that the mean of the data was equal to the variance. The hypothetical data may be over-dispersed compared to the Poisson distribution that explicitly acknowledged the similarity of the two means, which led to a much tighter posterior distribution. An alternative approach was to employ the negative binomial to update the Gamma distribution, which incorporated a dispersion factor specified as the ratio of the mean to the variance. Employing a dispersion factor would reduce the narrowness of the updated distribution, but the level of dispersion could only be subjectively specified. Despite the assumed narrowness of the resulting updated distributions, the Poisson distribution was employed to update Gamma distributions because the use of the negative binomial approach would have added further subjectivity to the VoI process.

A5.4 Cost parameters

Health care costs generally follow a skewed distribution for individual patients [Barber and Thompson, 1998; Delong and Simons, 1999], though Briggs and Gray [1998] found that a normal distribution was an adequate assumption for the sampling distribution of mean per patient costs. A number of authors have assumed that the logarithm of a cost is normally distributed, and the log-normal distribution has often been used to describe uncertainty in the mean value of cost parameters [Pasta et al, 1999; Fenwick et al, 2000]. However, as cost parameters display similar properties to the event times described in the previous section the gamma distribution can also be chosen to describe cost parameters.

To estimate the distribution parameters for the baseline distribution in the case study evaluation, the mean and the standard deviation from the corresponding weighted dataset for each cost parameter were inserted into the formulae presented in the previous section. Updating the gamma distribution for cost parameters on the basis of additional (hypothetical) data also follows the procedure developed above for the non-proportions clinical parameters. In the case of cost parameters $\sum t_i$ becomes $\sum c_i$, where the c_i are the individual estimates of the cost parameter. Again, $\sum c_i = n\bar{c}$.

A5.5 Utility value parameters

In the majority of cases, utility values are subject to the same constraints as proportions parameters – they are bounded between 0 and 1. The sample mean and standard deviation can be calculated and used to estimate the parameters of the beta distribution (a and b). If the assumption of zero as the minimum possible value of a utility value is not justified, it is unlikely that the upper bound of the utility value will approach 1. An adjustment solution may be possible, whereby the lowest possible utility value is specified, which is then added to each possible utility value in order to specify a beta distribution. The value added to fit the distribution is subtracted when applied to the relevant health state. For example, a range of –0.2 to +0.5 becomes a range of 0 to 0.7 in order to fit the distribution. Alternatively, a scale parameter can be applied to the value sampled from the beta distribution to expand the range outside 0 to 1. To update the probability distributions describing utility values the same procedure is employed as for the proportions clinical data.

A5.6 Estimating the standard deviation

The discussion in the previous sections has assumed that a selection of data is available to estimate the standard deviation, which can be employed to estimate the chosen distributions' parameters. If, for example, one utility value estimate with no measure of variance is available, the required standard deviation can be estimated directly. Alternatively, it may be more intuitive to specify a range around the observed value. A standard deviation can be estimated as one quarter of the full range, which, together with assumed mean value, can be employed to estimate the parameters of the appropriate probability distribution.

Appendix 6 Study characteristics tables

Table A6.1 Details of studies reporting length of DFI following tamoxifen alone

Study	Ther-apy	Regimen details	DFI details*	IBTR% reported	IBTR% assumed	Lgth of DFI	f-up	Age	Percentages						
									<50	50-9	60+	<66	65+	Pre m	Post m
[Fisher et al, 1996]	Tam	20mg/day, 5yrs	DND and 2ndP	4		10	10.42		100						
[Fisher et al, 1996]	Tam	20mg/day, 5yrs	DND and 2ndP	4		10	10.42			50	50				
[Fisher et al, 1997]	Tam	20mg/day, 5yrs	DND and 2ndP	11		5	4.00		45	28	27				
[Pritchard et al, 1997]	Tam	30mg/day, 2yrs	Neither DND or 2ndP	8		9								0	100
[Stewart, 1992]	Tam	20mg/day, 5 yrs	DND and likely not 2ndP	0		9	6.75	58						29	71
[Ribeiro and Swindell, 1992]	Tam	20mg/day, 1 yr	DND only	0		15	10.00							0	100
[Ribeiro and Swindell, 1992]	Tam	20mg/day, 1 yr	DND only	0		15	10.00							100	0
[Cummings et al, 1993]	Tam	20mg/day, 2 yrs	DND and likely 2ndP	0		12	10.00	70						0	100
[Baum et al, 1992]	Tam+/-peri-CT	20mg/day, 2yrs,	DND and contra' and likely not 2ndP	?	5	8	7.80	54.8						37	63
[Gerard et al, 1993]	Tam	30mg/day, 3 yrs	DND and likely not 2ndP	?	5	5	3.08							0	100
[Rivkin et al, 1994]	Tam	20mg/day, 1 yr	DND only	0		12	6.50	61						0	100
[Castiglione-Gertsch et al, 1994]	Tam + P	20mg/day, 1 yr,	DND and 2ndP	0		14	10.00						100	0	100
[Castiglione-Gertsch et al, 1994]	Tam + P	20mg/day, 1 yr	DND and 2ndP	0		14	10.00					100		0	100
[Mustacchi et al, 1994]	Tam	20mg/day die	DND and 2ndP	?	5	6	3.08	76						0	100
[Gundersen et al, 1995]	Tam	20mg/day, 2 yrs	Neither DND or 2ndP	0		10	6.33		32	34	34				
[Martelli et al, 1995]	Tam	20mg/day, indefinitely	DND and 2ndP	9		6	5.58	77						0	100

IBTR% reported, % of events named as ipsilateral breast relapse ending DFI; IBTR% assumed, % of ipsilateral breast relapse assumed to end DFI; f-up, follow-up; prem, premenopausal; postm, postmenopausal.

* DFI details describes whether studies included death with no evidence of disease (dnd) and 2nd primary tumours (2ndP) as ending DFI.

Table A6.1 Details of studies reporting length of DFI following tamoxifen alone (continued)

Study	Percentages															
	Node +	1 to 3 positive	4+ positive	Node -	T size	SD	pT1	pT2	pT3+	pT2/3	ER+	PgR+	Mast	Lump	Lump +RT	RT
[Fisher et al, 1996]	0	0	0	100	2	1				na	100	78	62	38	38	38
[Fisher et al, 1996]	0	0	0	100	2	1				na	100	78	62	38	38	38
[Fisher et al, 1997]	0	0	0	100	70		70	27	3	30	100	82	45	55	55	55
[Pritchard et al, 1997]	100	59	41	0			27	70	3	73	100	100	64	36	27	27
[Stewart, 1992]	0	0	0	100			36	59	5	64	63		100	0	0	0
[Ribeiro and Swindell, 1992]	0	0	0	100						na			100	0	0	0
[Ribeiro and Swindell, 1992]	0	0	0	100						na			100	0	0	0
[Cummings et al, 1993]	100	55	45	0			67			33	98	70	100	0	0	0
[Baum et al, 1992]	42			58			28			72						
[Gerard et al, 1993]	100	0	0	0						na						
[Rivkin et al, 1994]	100	49	51	0					6	na	100		96	4	4	4
[Castiglione-Gertsch et al, 1994]	100	54	46	0			39			61	6		100	0	0	0
[Castiglione-Gertsch et al, 1994]	100	54	46	0			39			61	6		100	0	0	0
[Mustacchi et al, 1994]	42	30 - 1a	12 - 1b	58			41	54	5	59						
[Gundersen et al, 1995]	100	67	33	0			27	73	0	73	100		97	3	0	0
[Martelli et al, 1995]	0	0	0	100			68	24	8	32	90	71	7	93	0	0

ER+, estrogen receptor positive; PgR+, progesterone receptor positive; Mast., mastectomy; Lump., lumpectomy; RT, radiotherapy.
* T refers to different tumour grades.

Table A6.2 Details of studies reporting length of DFI following tamoxifen and chemotherapy

Study	Therapy	Regimen details	DFS details†	IBTR% reported	IBTR% assumed	Lgth of DFI	f-up	age	Percentages					
									<50	50-9	60+	65+	pre m	post m
[Fisher et al, 1997]	Tam + MF	100-600, 6 cycles	DND and 2ndP	9		5	4.00		45	29	26			
[Fisher et al, 1997]	Tam + CMF	100-40-600, 6 cycles	DND and 2ndP	4		5	4.00		46	27	27			
[Pritchard et al, 1997]	Tam + CMF	600-40-600, 8 cycles	Neither DND or 2ndP	3		9							0	100
[Tormey et al, 1992]	Tam + CMFP	10mg/day, 1-5 yrs, 100-40-600-40	DND only	0		8	5.10	42					100	0
[Tormey et al, 1992]	Tam + CMFPH/VAT	4.5-45-12	DND only	0		8	5.10	44					100	0
[Baum et al, 1992]	Tam/Tam+ peri-CT	20mg/day, 2yrs, 5mg/kg 6 days	DND and contra' and likely not 2ndP	?	5	8	7.80	54.8					37	63
[Gerard et al, 1993]	Tam + FEC		DND and likely not 2ndP	?	5	5	3.08						0	100
[Gelber et al, 1993]	Tam + CMF	6 cycles (inc. Tam)	DND and unclear 2ndP	?	5	7	7.00						0	100
[Hupperets et al, 1993]	CAF + MPA	500-40-500, 6 cycles, MPA 6 months	DND and likely not 2ndP	?	5	8	3.50						62	38
[Rivkin et al, 1994]	Tam + CMFVP		DND only	0		12	6.50	60					0	100
[Schumacher et al, 1994]	CMF (w/wo Tam)	500-40-600, 3 cycles (30mg/day, 1 yr)	DND and 2ndP	0		6	4.67		42	30	28		42	58
[Castiglione-Gertsch et al, 1994]	Tam + CMFP	100-40-600, 12 cycles	DND and 2ndP	0		14	10.00						0	100
[Fukutomi et al, 1995]	Tam + ACMF	30mg/day, 2 yrs, 26-130-26-600, 6 cycles	DND and unclear 2ndP	0		7	5.40	45					100	0
[Fukutomi et al, 1995]	Tam + ACMF	30mg/day, 2 yrs, 13-65-13-300, 12 cycles	DND and unclear 2ndP	0		7	5.40	45					100	0

IBTR% reported, % of events named as ipsilateral breast relapse ending DFI; IBTR% assumed, % of ipsilateral breast relapse assumed to end DFI; f-up, follow-up; prem, premenopausal; postm, postmenopausal.

* DFI details describes whether studies included death with no evidence of disease (dnd) and 2nd primary tumours (2ndP) as ending DFI.

Table A6.2 Details of studies reporting length of DFI following tamoxifen and chemotherapy (continued)

Study	Percentages															
	Node +	1 to 3 positive	4+ positive	Node -	T size	SD	pT1	pT2	pT3+	pT2/3	ER+	PgR+	Mast	Lump	Lump +RT	RT
[Fisher et al, 1997]	0	0	0	100	68		68	29	3	32	100	83	45	55	55	55
[Fisher et al, 1997]	0	0	0	100	70		70	26	4	30	100	82	44	56	56	56
[Pritchard et al, 1997]	100	61	39	0			29	65	7	71	100	100	66	34	25	25
[Tormey et al, 1992]	100	54	46	0			52			48	63	61	100	0	0	
[Tormey et al, 1992]	100	53	47	0			52			48	63	65	100	0	0	
[Baum et al, 1992]	42			58			28			72						
[Gerard et al, 1993]	100	0	0	0						na						
[Gelber et al, 1993]	100	0	0	0						na						
[Hupperets et al, 1993]	100	68	32	0			32	58	10	68	66					
[Rivkin et al, 1994]	100	51	49	0					9	na	100		95	5	5	5
[Schumacher et al, 1994]	100	58	42	0			12	66	22	88	40	41	100	0	0	0
[Castiglione-Gertsch et al, 1994]	100	58	42	0			36			64	75		100	0	0	0
[Fukutomi et al, 1995]	100	54	46	0			25	58	17	75	62	60	100	0	0	0
[Fukutomi et al, 1995]	100	56	44	0			32	56	12	68	63	61	100	0	0	0

ER+, estrogen receptor positive; PgR+, progesterone receptor positive; Mast., mastectomy; Lump., lumpectomy; RT, radiotherapy.
* T refers to different tumour grades.

Table A6.3 Details of studies reporting type of event experienced following tamoxifen alone

Study	Therapy	Regimen details	Relapse details*	f-up	Age	SD	Percentages								
							<40	40-65	<50	50-9	60+	<66	65+	Prem	Postm
[Fisher et al, 1996]	Tam	20mg/day, 5yrs	L, r, con, dsep, pr, dnd	10.42	55	10								29	71
[Fisher et al, 1997]	Tam	20mg/day, 5yrs	Lr, dagg, con, pr, dnd	4.00					45	28	27				
[Pritchard et al, 1997]	Tam	30mg/day, 2yrs	L, lr, dagg, con, pr											0	100
[Stewart, 1992]	Tam	20mg/day, 5 yrs	Lr, con, dagg, dnd	6.75										100	0
[Stewart, 1992]	Tam	20mg/day, 5 yrs	Lr, con, dagg, dnd	6.75										0	100
[Cummings et al, 1993]	Tam	20mg/day, 2 yrs	Lr, dsep, con, pr, dnd	10.00	70									0	100
[Rivkin et al, 1994]	Tam	20mg/day, 1 yr	L, r, con, dsep, pr, dnd	6.50	61									0	100
[Castiglione-Gertsch et al, 1994]	Tam+ P	20mg/day, 1 yr,	Lr, con, dsep, pr, dnd	10.00									100	0	100
[Castiglione-Gertsch et al, 1994]	Tam+ P	20mg/day, 1 yr,	Lr, con, dsep, pr, dnd	10.00								100		0	100
[Mustacchi et al, 1994]	Tam	20mg/day die	L, dagg, dnd	3.08	76									0	100
[Gundersen et al, 1995]	Tam+ peri-op CT	20mg/day, 2 yrs	L, r, con, dagg, dnd	6.33					32						
[Martelli et al, 1995]	Tam	20mg/day, indefinitely	L, r, dagg, pr, dnd	5.58	77									0	100
[Kamby et al, 1988]	Tam	30mg/day 1year	L, r, con, dsep						26	25	49			37	63

P, prednisolone; f-up, follow-up (years); SD, standard deviation; prem, premenopausal; postm, postmenopausal.
Relapse details describe the types of relapse reported by each study: L, local; r, regional; Lr, locoregional; con, contralateral; dsep, separate sites of metastases; dagg, aggregate metastases; pr, primary tumours; dnd, death with no evidence of disease.

Table A6.3 Details of studies reporting type of event experienced following tamoxifen alone (continued)

Study	Percentages														
	Node +	1 to 3 nodes	4+ nodes	Node -	Tumour size	pT1*	pT2*	pT3+*	pT2/3*	ER+	PgR+	Mast.	Lump.	lump+RT	RT
[Fisher et al, 1996]	0	0	0	100	2				na	100	78	62	38	38	38
[Fisher et al, 1997]	0	0	0	100		70	27	3	30	100	82	45	55	55	55
[Pritchard et al, 1997]	100	59	41	0		27	70	3	73	100	100	64	36	27	27
[Stewart, 1992]	0	0	0	100		36	59	5	64	63		100	0	0	0
[Stewart, 1992]	0	0	0	100		36	59	5	64	63		100	0	0	0
[Cummings et al, 1993]	100	55	45	0		67			33	98	70	100	0	0	0
[Rivkin et al, 1994]	100	49	51	0				6	na	100		96	4	4	4
[Castiglione-Gertsch et al, 1994]	100	54	46	0		39			61	6					
[Castiglione-Gertsch et al, 1994]	100	54	46	0		39			61	6					
[Mustacchi et al, 1994]	42	30 - 1a	12 - 1b	58		41	54	5	59						
[Gundersen et al, 1995]	100	67	33	0		27	73	0	73	100		97	3	0	0
[Martelli et al, 1995]	0	0	0	100		68	24	8	32	90	71	7	93	0	0
[Kamby et al, 1988]	83	56	27	17		5	71	24	95			100	0	0	100

ER+, estrogen receptor positive; PgR+, progesterone receptor positive; Mast., mastectomy; Lump., lumpectomy; RT, radiotherapy.
* T refers to different tumour grades.

Table A6.4 Details of studies reporting type of event experienced following tamoxifen and chemotherapy

Study	Therapy	Regimen details	Relapse details*	f-up	Age	Proportions					
						<50	50-9	60+	65+	Prem	Postm
[Fisher et al, 1997]	Tam + MF	100-600, 6 cycles	Lr, dagg, con, pr, dnd	4.00		45	29	26			
[Fisher et al, 1997]	Tam + CMF	100-40-600, 6 cycles	Lr, dagg, con, pr, dnd	4.00		46	27	27			
[Pritchard et al, 1997]	Tam + CMF	600-40-600, 8 cycles	L, lr, dagg, con, pr							0	100
[Tormey et al, 1992]	Tam + CMFP	10mg/day, 1-5 yrs, 100-40-600-40	L, r, dsep, pr, dnd	5.10	42					100	0
[Tormey et al, 1992]	Tam + CMFPH/VAT	4.5-45-12	L, r, dsep, pr, dnd	5.10	44					100	0
[Lindeman et al, 1992]	CMF (+tam in ER+)	100-40-600, 60-40-600, 6 or 12 cycles	Lr, dsep	5.17	44	76					
[Rivkin et al, 1994]	Tam + CMFVP	20mg/day, 1 yr, 60/day 15-400/week, 1 yr	L, r, con, dsep, pr, dnd	6.50	60					0	100
[Schumacher et al, 1994]	CMF (w/wo Tam)	500-40-600, 3 or 6 cycles (30mg/day, 1 yr)	L, r, dagg, pr, dnd	4.67		42	30	28		42	58
[Castiglione-Gertsch et al, 1994]	Tam + CMFP	100-40-600, 12 cycles	Lr, con, dsep, pr, dnd	10.00						0	100
[Gundersen et al, 1995]	Tam (with peri-op CT)	20mg/day, 2 yrs	L, r, con, dagg, dnd	6.33		32					
[Fukutomi et al, 1995]	Tam + ACMF	30mg/day, 2 yrs, 26-130-26-600, 6 cycles	L, r, dsep	5.40	45					100	0
[Fukutomi et al, 1995]	Tam + ACMF	30mg/day, 2 yrs, 13-65-13-300, 12 cycles	L, r, dsep	5.40	45					100	0
[Powles et al, 1995]	3M/2M + tamoxifen	Mitom 7 every 6w, mitox 7 e 3w, meth 35 e 3w, 8 cycles, 20mg/day 5 years	L,r, dagg	2.33	56					40	60

P, prednisolone; f-up, follow-up (years); SD, standard deviation; prem, premenopausal; postm, postmenopausal.
Relapse details describe the types of relapse reported by each study: L, local; r, regional; Lr, locoregional; con, contralateral; dsep, separate sites of metastases; dagg, aggregate metastases; pr, primary tumours; dnd, death with no evidence of disease.

Table A6.4 Details of studies reporting type of event experienced following tamoxifen and chemotherapy (continued)

Study	Percentages														
	Node +	1 to 3 positive	4+ positive	Node -	T size	pT1	pT2	pT3+	pT2/3	ER+	PgR+	Mast	Lump	Lump +RT	RT
[Fisher et al, 1997]	0	0	0	100	68	68	29	3	32	100	83	45	55	55	55
[Fisher et al, 1997]	0	0	0	100	70	70	26	4	30	100	82	44	56	56	56
[Pritchard et al, 1997]	100	61	39	0		29	65	7	71	100	100	66	34	25	25
[Tormey et al, 1992]	100	54	46	0		52			48	63	61	100			
[Tormey et al, 1992]	100	53	47	0		52			48	63	65	100			
[Lindeman et al, 1992]	100	73	27	0	3		53		na	56		100	0	0	0
[Rivkin et al, 1994]	100	51	49	0				9	na	100		95	5	5	5
[Schumacher et al, 1994]	100	58	42	0		12	66	22	88	40	41	100	0	0	0
[Castiglione-Gertsch et al, 1994]	100	58	42	0		36			64	75					
[Gundersen et al, 1995]	100	67	33	0		27	73	0	73	100		97	3	0	0
[Fukutomi et al, 1995]	100	54	46	0		25	58	17	75	62	60	100	0	0	0
[Fukutomi et al, 1995]	100	56	44	0		32	56	12	68	63	61	100	0	0	0
[Powles et al, 1995]	14	8 - 1a	6 - 1b	86		10	82	8	90			28	72	72	72

ER+, estrogen receptor positive; PgR+, progesterone receptor positive; Mast., mastectomy; Lump., lumpectomy; RT, radiotherapy.
* T refers to different tumour grades.

Appendix 7 The input data

A7.1 Introduction

From the literature review, data were collected for a wide range of patient groups experiencing early breast cancer, reflecting many patient characteristics (as described in Appendix 6). The process of examining the data to defined relevant sub-groups to evaluate was described in Chapter 4. From that process, the patient group chosen for the case study evaluation was postmenopausal women with node positive early breast cancer. The relevant therapy comparison for this patient group was defined as tamoxifen and chemotherapy versus tamoxifen alone. This Appendix describes the probability distributions used to populate the case study models, which were derived from the data analysis described in Chapter 5.

The data are presented in the same order as the methods of data analysis were discussed, namely, the clinical parameter values are followed by the cost parameters, and then the utility values.

A7.2 Clinical parameters

The clinical parameters consist of the different forms of toxicity, disease-free interval (DFI) and the clinical events ending DFI, and progress from relapse. Each are descibed in the following sections.

A7.2.1 Toxicity parameters

The data identified describing the probability of experiencing different categories of toxicity (major toxicity, grade 3 or 4 toxicity, and grade 1 or 2 toxicity), and the types of conditions included in each of the categories are described in the following sections. A final section describes the representation of treatment compliance, which was linked to the experience of toxicity.

A7.2.1.1 Major toxic events

Four studies were identified that presented data on major toxic events associated with tamoxifen and chemotherapy, including a total of 1964 patients [Tormey et al, 1992; Rivkin et al, 1994; Pritchard et al, 1996; Fisher et al, 1997]. The same number of studies was included in the analysis of major toxicity and tamoxifen, but the total number of patients was almost double (3733 patients) [Fisher et al, 1996a; Rivkin et al, 1994; Fisher et al, 1997; Pritchard et al, 1996]. Table A7.1 presents the point estimates for the incidence of major toxicity for both therapy options alongside the number of patients included in each study, and the resulting weight attached to each estimate. A relatively wide range of estimates were identified for tamoxifen and chemotherapy that might reflect differences in the age and the burden of disease of the patients included in the respective trials, as these factors are known to have an impact on the incidence of thromboembolism [Pritchard et al, 1996]. The tamoxifen figures are very small and there is not much scope for variation between the estimates.

Table A7.1 Point estimates, and corresponding sample size and weights, for data describing the experience of major toxicity

Tamoxifen+chemotherapy			Tamoxifen alone		
Data	Sample	Weight	Data	Sample	Weight
0.096	353	10	0.012	1422	29
0.070	263	10	0.006	1188	49
0.036	300	21	0.012	771	16
0.045	778	43	0.014	352	6
0.040	270	17			

Table A7.2 shows the types of major toxic events that were experienced. Deep vein thrombosis (DVT) is clearly the prominent major toxic event. In patients receiving tamoxifen and chemotherapy all major events were thromboembolic, whilst a significant proportion of patients receiving tamoxifen alone experienced cardiac events.

Table A7.2 Proportion of specific major toxic events experienced by patients receiving alkylating drugs-based regimens

Condition	Tamoxifen+chemotherapy	Tamoxifen alone
Pulmonary Embolism†	0.196	0.144
Deep Vein Thrombosis	0.741	0.462
Arterial Thrombosis	0.047	0.076
Mesenteric Vein Thrombosis	0.016	-
Unspecified cardiac event	-	0.318
Total	1.000	1.000

† pulmonary infarction and embolism are included in this category

A7.2.1.2 Grade 3 or 4 toxicity

The same four studies reporting major toxicity also presented grade 3 or 4 toxicity in patients receiving tamoxifen and chemotherapy, involving 1964 patients [Tormey et al, 1992; Rivkin et al, 1994; Pritchard et al, 1996; Fisher et al, 1997]. The largest study reporting on tamoxifen alone did not report grade 3 or 4 toxicity so only 1268 patients informed this analysis [Rivkin et al, 1994; Fisher et al, 1997; Pritchard et al, 1997]. Table A7.3 presents the separate point estimates for the incidence of grade 3 or 4 toxicity. The estimates for tamoxifen and chemotherapy appear stable apart from one low estimate, which came from the largest trial [Fisher et al, 1997]. The highest estimate for tamoxifen alone is 0.05, but the trial reporting this estimate was small relative to the remaining trials and the associated weight reduces it's impact [Rivkin et al, 1994].

Table A7.3 Point estimates, and corresponding sample size and weights, for data describing the experience of grade 3/4 toxicity

Tamoxifen+chemotherapy			Tamoxifen alone		
Data	Sample	Weight	Data	Sample	Weight
0.482	353	16	0.003	352	59
0.510	263	12	0.050	145	2
0.500	270	12	0.010	771	39
0.610	300	14			
0.250	768	46			

The grade 3 or 4 toxicity conditions experienced are presented in table A7.4. The total indicates the average number of conditions experienced by a single patient. The composition analysis was complicated by the studies that did not present an aggregate proportion experiencing grade 3 or 4 toxicity (see section 5.3.1). Using the maximum proportion as a proxy for the aggregate proportion overestimated the proportion of patients experiencing such conditions, though few resources were associated with two of the main events, leukopenia and thrombocytopenia, so the impact of the overestimation was reduced. Apart from these two conditions, the main toxicity for patients receiving tamoxifen and chemotherapy was nausea and vomiting. The dominance of infection in tamoxifen alone was initially surprising, but understandable given the very small incidence of grade 3 or 4 toxicity.

Table A7.4 Proportions of patients experiencing toxicity conditions in the grade 3 or 4 toxicity category.

Condition	Tamoxifen+chemotherapy	Tamoxifen alone
Nausea/vomiting	0.280	0.266
Stomatitis/mucositis	0.037	-
Diarrhea	0.108	0.061
Leukopenia*	0.541	-
Thrombocytopenia†	0.259	0.278
Neurotoxicity‡	0.131	-
Infection	0.09	0.521
Neutropenia‡	-	0.061
Total	1.447	1.186

* leucocyte is included in this toxicity condition.
† includes platelet count < 50 x 10^9/L
‡ includes paresthesias/neuropathy

A7.2.1.3 Grade 1 or 2 toxicity

Only three studies including tamoxifen and chemotherapy reported grade 1 or 2 toxicity, totalling 1421 patients [Rivkin et al, 1994; Pritchard et al, 1996; Fisher et al, 1997]. Five trials with tamoxifen alone presented data on grade 1 or 2 toxicity with 2775 patients [Fisher et al, 1996a; Rivkin et al, 1994; Fisher et al, 1997; Pritchard et al, 1997; Bergman et al, 1995]. Table A7.5 presents the separate point estimates for the incidence of grade 1 or 2 toxicity. There was a large amount of variation in the tamoxifen and chemotherapy data but the low estimate of 0.34 was actually a maximum estimate used as a proxy for the

aggregate proportion (the proportion of patients experiencing mucositis) [Rivkin et al, 1994] and was almost certainly an underestimate.

Table A7.5 Point estimates, and corresponding sample size and weights, for data describing the experience of grade 1/2 toxicity

Tamoxifen+chemotherapy			Tamoxifen alone		
Data	Sample	Weight	Data	Sample	Weight
0.620	353	23	0.305	85	3
0.340	300	20	0.640	1422	41
0.720	768	57	0.085	352	30
			0.300	145	5
			0.430	771	21

The interpretation of the proportions presented for patients receiving tamoxifen alone is difficult. In disaggregating the data the lower proportions were reported in studies comparing tamoxifen with a chemotherapy-based treatment arm, which may be due to the choice of side effects reported. Comparing tamoxifen with a placebo the occurrence of hot flashes, vaginal discharge, fluid retention, and skin changes - all effects that may be described as hormonal – were described [Fisher et al, 1996a]. Two other trials compared tamoxifen with a chemotherapy treatment arm, and of the hormonal conditions only 'hot flashes' was recorded [Pritchard et al, 1997; Rivkin et al, 1994]. Fisher *et al* [1996] labelled side effects as 'undesirable sequelae' which gave no indication of their intensity and it is likely that only a small proportion would be classified as equivalent to a grade 1 or 2 toxicity.

Table A7.6 presents the conditions experienced within grade 1 or 2 toxicity. The data reported by one study was excluded from this analysis because the use of the maximum proportion led to impossibly high estimates of four conditions [Rivkin et al, 1994]. For patients receiving tamoxifen and chemotherapy leukopenia was common, though nausea/vomiting was by far the most common form of toxicity with real impact. Hot flushes and nausea/vomiting dominated the grade 1 or 2 conditions experienced for tamoxifen alone.

Table A7.6 Proportions of patients experiencing toxicity conditions in the grade 1 or 2 toxicity category.

Condition	Tamoxifen+chemotherapy	Tamoxifen alone
Nausea/vomiting	0.619	0.356
Stomatitis/mucositis	0.077	0.003
Diarrhea	0.024	0.093
Leukopenia*	0.563	-
Anemia	0.210	-
Constipation	0.023	-
Thrombocytopenia†	0.233	0.004
Neurotoxicity‡	-	-
Granulocytopenia	-	-
Fatigue	-	-
Hot flashes	-	0.565
Vaginal discharge	-	0.246
Superficial phlebitis	0.046	0.026
Total	1.794	1.294

* leucocyte was included in this toxicity condition.
† includes platelet count < 50 x 10^9/L
‡ includes paresthesias/neuropathy

A7.2.1.4 Treatment compliance

Aggregate proportions of patients completing specific cycles of chemotherapy were reported [Schumacher et al, 1994; Coombes et al, 1996], but the length of treatment was not linked to the type of toxicity experienced. In order to model the relationship between non-compliers and the severity of their toxicity experiences, it was assumed that those patients experiencing the worst toxicity would receive the fewest cycles of chemotherapy. Table A7.7 describes the proportion of each class of toxicity receiving specific cycles of chemotherapy.

Table A7.7 Assignment of compliance rates across types of toxicity experienced

Chemotherapy cycles	Aggregate proportions receiving specific cycles	Grade1/2 toxicities (0.57)	Grade 3/4 toxicities (0.38)	Major toxicities (0.05)
1	0.0055	0.00	0.00	0.11
2	0.0055	0.00	0.00	0.11
3	0.045	0.00	0.07	0.39
4	0.0375	0.00	0.05	0.39
5	0.0375	0.00	0.10	0.00
6	0.869	1.00	0.79	0.00

Only four studies referred to treatment compliance for tamoxifen alone. Three of these studies were review papers [Bryson and Plosker, 1993; EBCTCG, 1998; Jaiyesimi et al. 1995], whilst one reported findings from a trial containing 352 patients receiving tamoxifen only [Pritchard et al, 1997]. The latter reports a discontinuation proportion of 0.04, whilst the figures in the reviews range from 0.03 to 0.05, with only seldom exceptions. One review paper cited a large trial in which about 20% of women either failed to start tamoxifen or discontinued it prematurely [EBCTCG, 1998], whilst another study found that 10% of patients did not complete 1 year due to side effects [Jaiyesimi et al, 1995].

A7.2.2 Disease free interval and clinical events ending DFI

Input distributions were specified for each year following primary surgery for the analysis of DFI, and for each type of event for the clinical events ending DFI. Table A7.8 presents the separate estimates of the proportions leaving DFI in each year following diagnosis, whilst Figure A7.1 shows the aggregate survival curves for each therapy option. Table A7.9 presents the separate estimates for the proportions of patients experiencing the alternative events that follow DFI.

Table A7.8 Point estimates and corresponding sample sizes, for data describing the annual probability of the remaining cohort leaving the health state 'disease free interval'

Tamoxifen+ chemotherapy														
Year	1*		2*		3*		4*		5*		6*		7*	
[Pritchard et al, 1997]	0.930	352	0.076	327	0.113	303	0.057	268	0.135	253	0.191	219	0.107	177
[Gerard et al, 1993]	0.970	182	0.074	176	0.057	163	0.036	154	0.007	148	-	-	-	-
[Gelber et al, 1993]	0.924	341	0.103	315	0.149	283	0.081	241	0.095	221	0.057	200	0.086	189
[Jansen et al, 1998]	0.960	303	0.073	291	0.135	270	0.117	233	0.118	206	0.067	182	0.071	170
[Castiglione-Gertsch et al, 1994]	0.887	97	0.137	86	0.123	74	0.042	65	0.087	62	0.064	57	0.034	53
Year	8		9		10		11		12		13		14	
[Pritchard et al, 1997]	0.168	158	0.029	132	-	-	-	-	-	-	-	-	-	-
[Gerard et al, 1993]	-	-	-	-	-	-	-	-	-	-	-	-	-	-
[Gelber et al, 1993]	-	-	-	-	-	-	-	-	-	-	-	-	-	-
[Jansen et al, 1998]	0.058	158	0.061	148	0.130	139	0.025	121	0.000	118	-	-	-	-
[Castiglione-Gertsch et al, 1994]	0.124	52	0.081	45	0.033	41	0.011	40	0.023	40	0.023	39	0.000	38
Tamoxifen alone														
Year	1*		2*		3*		4*		5*		6*		7*	
[Pritchard et al, 1997]	0.896	353	0.093	316	0.090	287	0.070	261	0.136	243	0.052	210	0.111	199
[Gerard et al, 1993]	0.947	192	0.088	182	0.066	166	0.057	155	0.010	146	-	-	-	-
[Jansen et al, 1998]	0.940	295	0.106	277	0.119	248	0.149	218	0.111	186	0.089	165	0.078	150
[Castiglione-Gertsch et al, 1994]	0.763	116	0.174	89	0.135	73	0.087	63	0.114	58	0.075	51	0.046	47
Year	8		9		10		11		12		13		14	
[Pritchard et al, 1997]	0.062	177	0.044	166	-	-	-	-	-	-	-	-	-	-
[Gerard et al, 1993]	-	-	-	-	-	-	-	-	-	-	-	-	-	-
[Jansen et al, 1998]	0.085	139	0.186	127	0.029	103	0.118	100	0.100	89	-	-	-	-
[Castiglione-Gertsch et al, 1994]	0.085	45	0.053	41	0.056	39	0.030	37	0.015	36	0.016	35	0.032	35

* The two columns represents the reported value and the sample. The corresponding weights are not displayed, but they can be estimated using the following formula

$$\frac{p_i(1-p_i)}{n_i}.$$

Figure A7.1 Disease-free survival curves for tamoxifen and chemotherapy, and for tamoxifen alone, as adjuvant therapies for node-positive, postmenopausal women with early breast cancer

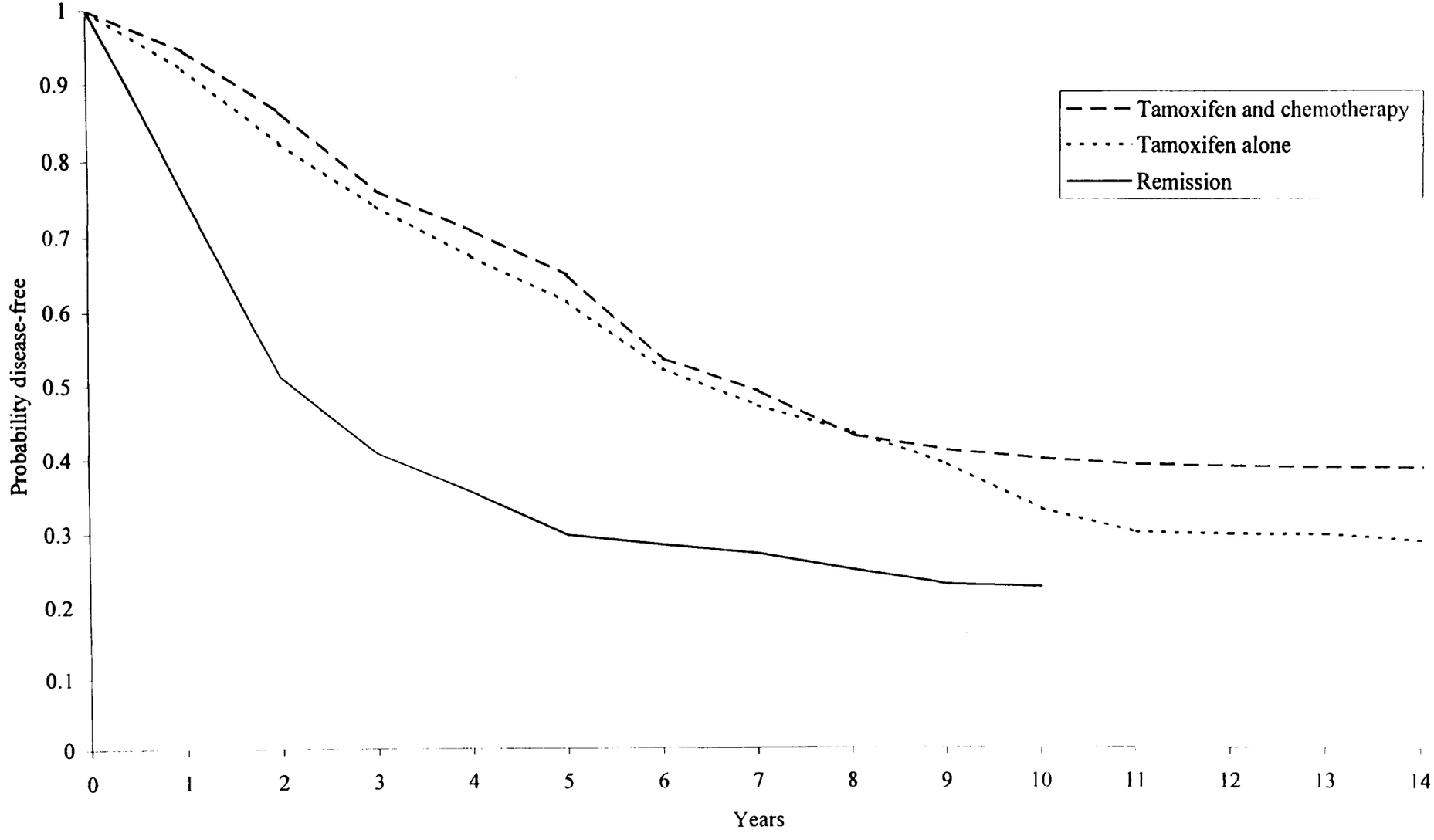


Table A7.9 Point estimates and corresponding sample sizes and weights, for data describing the destination state of patients leaving the health state 'disease free interval'

Tamoxifen+chemotherapy	Locoregional*			Soft tissue*			Bone*			Visceral*			DNED*		
[Pritchard et al, 1997]	0.332	160	721	-	-	-	-	-	-	-	-	-	0.080	160	2171
[Rivkin et al, 1994]	0.223	141	814	-	-	-	0.344	141	625	0.297	141	675	0.006	141	23140
[Castiglione-Gertsch et al, 1994]	0.254	91	479	0.034	91	2774	0.271	91	460	0.305	91	429	0.136	91	775
Tamoxifen alone	Locoregional*			Soft tissue*			Bone*			Visceral*			DNED*		
[Pritchard et al, 1997]	0.355	160	700	-	-	-	-	-	-	-	-	-	0.040	160	4178
[Rivkin et al, 1994]	0.320	150	689	-	-	-	0.330	150	678	0.310	150	701	0.113	150	1490
[Castiglione-Gertsch et al, 1994]	0.306	110	519	0.014	110	8043	0.306	110	519	0.306	110	519	0.056	110	2100

* The three columns represents the reported value, the sample, and the corresponding weight applied to the value

A7.2.3 Progress from relapse

The data describing progression from relapse is presented separately for progression from locoregional relapse and from metastases.

A7.2.3.1 Locoregional relapse

Nine studies presented data on the time to metastatic dissemination of locoregional relapse [Kamby and Sengelov, 1997; Borner et al, 1996; Crowe et al, 1991; Beck et al, 1983; Janjan et al, 1986; Bedwinek et al, 1981; Schwaibold et al, 1991; Toonkel et al, 1983; Toi et al, 1997]. The representation of the time spent in remission (disease-free survival (DFS)) following a locoregional relapse was the major difference between the two modelling techniques - the Markov process and the DES model. In the DES model, DFS was represented in an identical manner to the representation of DFI, but in the Markov process it was only possible to describe DFS as a constant probability of experiencing a further relapse (see section 5.4.3). The raw data employed to represent DFS is presented in Table A7.10, whilst the mean survival curve is presented in Figure A7.1.

Table A7.10 Point estimates and corresponding sample sizes, for data describing the annual probability of the remaining cohort leaving the health state 'remission'

Year	1*			2*			3*			4*			5*		
[Borner et al, 1996]	0.740	61	317	0.470	45	181	0.390	29	121	0.320	24	109	0.290	20	95
[Borner et al, 1996]	0.850	71	557	0.760	60	331	0.700	54	257	0.650	50	218	0.600	46	192
[Crowe et al, 1991]	0.570	81	330	0.325	46	210	0.260	26	137	0.200	21	132	0.150	16	127
[Beck et al, 1983]	0.670	121	547	0.400	81	338	0.315	48	224	0.250	38	203	0.200	30	189
[Janjan et al, 1986]	0.650	57	251	0.480	37	148	0.325	27	125	0.300	19	88	0.225	17	98
[Janjan et al, 1986]	0.850	50	392	0.570	43	173	0.460	29	115	0.370	23	99	0.370	19	79
[Janjan et al, 1986]	0.850	57	447	0.550	48	196	0.400	31	131	0.350	23	100	0.325	20	91
[Bedwinek et al, 1981]	0.760	32	175	0.630	24	104	0.510	20	81	0.400	16	68	0.360	13	56
[Schwaibold et al, 1991]	0.652	128	564	0.460	83	336	0.343	59	261	0.297	44	210	0.235	38	211
[Toonkel et al, 1983]	0.835	124	900	0.505	104	414	0.355	63	273	0.335	44	198	0.250	42	222
Year	6*			7*			8*			9*			10*		
[Borner et al, 1996]	0.290	18	86	0.290	18	86	0.290	18	86	0.290	18	86	0.290	18	86
[Borner et al, 1996]	0.540	43	171	0.540	38	154	0.480	38	154	0.400	34	142	0.400	28	118
[Crowe et al, 1991]	0.125	12	111	0.100	10	113	0.100	8	90	0.100	8	90	0.090	8	99
[Beck et al, 1983]	0.170	24	172	0.150	21	161	0.105	18	193	0.105	13	135	0.105	13	135
[Janjan et al, 1986]	0.190	13	83	0.190	11	70	0.190	11	70	0.190	11	70	-	-	-
[Janjan et al, 1986]	0.325	19	84	0.325	16	74	0.325	16	74	0.325	16	74	0.325	16	74
[Janjan et al, 1986]	0.300	19	88	0.300	17	81	0.300	17	81	-	-	-	-	-	-
[Bedwinek et al, 1981]	0.230	12	65	0.200	7	46	0.200	6	40	0.150	6	50	0.150	5	38
[Schwaibold et al, 1991]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
[Toonkel et al, 1983]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

* The three columns represents the reported value, the sample, and the corresponding weight applied to the value.

A7.2.3.2 Metastases

Substantial amounts of data describing the median length of survival from the point of diagnosis with metastases were identified. From clinical trials comparing alternative therapies for metastatic cancer, 113 separate treatment arms containing patients diagnosed with metastases were identified. 68 reported the component proportions of the different metastatic sites, whilst the remaining treatment arms reported data on individual sites – 11 soft tissue, 15 bone and 19 visceral. Given the reasonable number of studies reporting separate survival times for the three metastatic sites, survival data on specific metastatic sites were used to estimate median survival times. No information other than median survival times were presented in the vast majority of studies reporting such data, no attempt was made to estimate mean survival.

A number of studies presented data on survival from metastases differentiating with respect to nodal status [Koenders et al, 1992; Venturini et al, 1996]ER status receptor [Koenders et al, 1992; Vogel et al, 1992; Alonso et al, 1995; Venturini et al, 1996], PgR receptor status [Koenders et al, 1992], menopausal status [Venturini et al, 1996], and the administration of prior adjuvant therapies [de Takats et al, 1993; Venturini et al, 1996]. These data were used to estimate survival multipliers for different patients groups. For example, node negative patients with visceral metastases were estimated to live 33 per cent longer than the average survival length for visceral metastases, conversely node positive patients lived 33 per cent less than the average. Table A7.11 presents the patient categories and their associated multipliers for node positive patients.

Table A7.11 Survival from metastases multipliers for node positive patients

Patient group	Soft tissue	Bone	Visceral
Node positive	0.82	0.82	0.67

The respective multipliers were applied to the individual estimates of survival derived from the identified studies. The identified data for the three metastatic sites are presented in Table A7.12.

Table A7.12 Point estimates and corresponding sample sizes, for data describing median survival from the metastatic health states

Soft tissue	Data	Sample
[Koenders et al, 1992]	33.62	52
[Vogel et al, 1992]	38.54	70
[Alonso et al, 1995]	16.40	32
[Alonso et al, 1995]	16.40	30
[Goldhirsch et al, 1988]	11.56	26
[Leonard et al, 1994]	12.30	42
[Kamby et al, 1988]	34.06	277
[Clark et al, 1987]	32.64	284
[Paridaens et al, 1993]	27.06	13
[Paridaens et al, 1993]	27.06	15
[Castiglione-Gertsch et al, 1997]	26.08	52

Bone	Data	Sample
[Koenders et al, 1992]	27.88	70
[Vogel et al, 1992]	19.68	48
[Rasmusson et al, 1995]	9.84	100
[Rasmusson et al, 1995]	9.02	100
[Alonso et al, 1995]	16.40	23
[Alonso et al, 1995]	16.40	39
[Goldhirsch et al, 1988]	17.88	90
[Goldhirsch et al, 1988]	12.87	22
[Coleman and Rubens, 1987]	19.68	150
[Kamby et al, 1988]	17.03	225
[Clark et al, 1987]	17.55	274
[Paridaens et al, 1993]	25.17	27
[Paridaens et al, 1993]	25.17	19
[Castiglione-Gertsch et al, 1997]	20.66	10

Visceral	Data	Sample
[Koenders et al, 1992]	10.72	137
[Vogel et al, 1992]	10.72	75
[Goldhirsch et al, 1988]	8.78	24
[Goldhirsch et al, 1988]	8.31	24
[Kamby et al, 1988]	7.73	361
[Clark et al, 1987]	9.38	386
[Paridaens et al, 1993]	14.22	42
[Paridaens et al, 1993]	14.22	41
[Castiglione-Gertsch et al, 1997]	8.04	34
[Alonso et al, 1995]	8.04	71
[Alonso et al, 1995]	8.04	55
[Leonard et al, 1994]	6.70	55
[Arai et al, 1994]	8.38	56
[Goldhirsch et al, 1988]	3.62	18
[Goldhirsch et al, 1988]	5.09	16
[Coleman and Rubens, 1987]	2.01	75
[Leonard et al, 1994]	5.36	55
[Kocher et al, 1995]	2.14	190
[Kocher et al, 1995]	8.04	45

Estimates of the time to progression (TTP) were required indirectly for the model, in order to calculate the monthly costs of the metastases states (see section 5.4.4.4). Three categories of patients were specified for whom data describing TTP were required: Patients receiving hormonal therapy for bone and soft tissue metastases; chemotherapy for visceral metastases; and chemotherapy for bone and soft tissue metastases. Fewer trials presented data on the TTP or TTF. For example, TTP estimates were available for only 2 of the 19 treatment arms presenting specific data on survival from visceral metastases. TTP was estimated, therefore, using all studies in which a single metastatic site was dominant in over 50% of patients. Table A7.13 presents the separate point estimates of the median TTP for the three identified patient groups.

Table A7.13 Point estimates and corresponding sample sizes, for data describing median time to progression from the metastatic health states

Bone/Soft tissue (hormonal)			Visceral (chemotherapy)		
[Pyrhonen et al, 1997]	7.30	214	[Swain et al, 1997]	8.55	168
[Pyrhonen et al, 1997]	10.20	201	[Swain et al, 1997]	8.55	181
[Pyrhonen et al, 1997]	6.00	91	[Swain et al, 1997]	7.73	81
[Muss et al, 1994]	6.50	91	[Swain et al, 1997]	7.73	104
[Jonat et al, 1995]	5.31	159	[Gabra et al, 1996]	10.00	56
[Jonat et al, 1995]	6.46	159	[Stewart et al, 1997]	5.30	128
[Thurlimann et al, 1996]	19.80	107	[Stewart et al, 1997]	3.20	121
[Thurlimann et al, 1996]	15.00	105	[Richards et al, 1992]	4.00	28
[Hayes et al, 1995]	5.75	215	[Richards et al, 1992]	5.00	31
[Hayes et al, 1995]	5.52	221	[Blomqvist et al, 1993]	9.20	86
[Hayes et al, 1995]	5.49	212	[Blomqvist et al, 1993]	5.40	84
Bone/Soft tissue (chemotherapy)			[Ingle et al, 1994]	4.10	80
[Jones et al, 1995]	2.77	115	[Ingle et al, 1994]	4.30	83
[Jones et al, 1995]	1.85	64	[Venturini et al, 1996]	9.80	326
[Pavesi et al, 1995]	8.50	71	[Paridaens et al, 1993]	13.85	82
[Pavesi et al, 1995]	7.50	70	[Ibrahim et al, 1996]	12.00	767
[Cobau et al, 1996]	7.00	135	[Ibrahim et al, 1996]	10.50	244
[Cobau et al, 1996]	4.00	131			
[Bastholt et al, 1996]	4.40	75			
[Bastholt et al, 1996]	4.70	66			
[Bastholt et al, 1996]	8.40	64			
[Bastholt et al, 1996]	8.40	58			
[Aisner et al, 1995]	8.24	165			
[Aisner et al, 1995]	8.48	164			
[Aisner et al, 1995]	8.90	162			
[Chu et al, 1996]	2.77	21			
[Paridaens et al, 1993]	8.28	73			

A7.3 Cost parameters

Four categories of cost parameters are presented in the following sections: adjuvant therapies, surveillance, toxicity, and relapse.

A7.3.1 Adjuvant therapies

To represent the cost of chemotherapy primary cost estimates for the individual components of chemotherapy were identified, as well as aggregate estimates derived directly from the literature. For the former, three resource elements were identified – drugs, anti-emetics and administration. Drug costs were based on protocols presented in the literature, unit costs from various sources were then applied to the protocol to estimate

total drug costs. Tamoxifen is self-administered so no costs other than the cost of the treatment itself were considered. Table A7.14 presents the baseline estimates for the alternative adjuvant therapies.

Table A7.14 Adjuvant therapy costs

Therapy and source	Constituents*	Unit cost (£'s)	Cost (£'s)
Tamoxifen	20/30mg	0.12/0.22	3.60/6.54 per month
CMF[1] – standard	Cyclophosphamide 100mg/m^2 days 1-14	0.11 – 50mg	4.93 per cycle
Milan regimen	Methotrexate 40 mg/m^2 days 1&8 iv.	2.62 – 50mg	10.48 per cycle
	5-Fluorouracil 600 mg/m^2 days 1&8 iv.	3.20 – 250mg	19.20 per cycle
CAF[2]	Cyclophosphamide 400 mg/m^2 day 1 iv.	1.65 – 200mg	3.30 per cycle
	Adriamycin 40 mg/m^2 day 1 iv.	20.60 – 10mg	82.40 per cycle
	5-Fluorouracil 400 mg/m^2 days 1&8 iv.	3.20 – 250mg	12.80 per cycle

* All costs taken from British National Formulary, apart from cost of fraction of radiotherapy [Warde and Murphy, 1996], outpatient visit (see text) and surgical oophorectomy (national schedule of reference costs), which are from an NHS Trust. The average female body surface was assumed to be 1.6m^2.

† Second year costs have been discounted at 6%

[1] [Fisher et al, 1996b]

[2] [Alonso et al, 1995]

The use of standard anti-emetics alongside the chemotherapy regimens was not reported in any study, but a handbook published by the Royal Marsden NHS Trust provided details of appropriate schedules [Price et al, 1995]. Estimates of the necessary inputs from health professionals to the administration of chemotherapy were sought from clinicians. Two studies were identified that presented the cost of disposable items employed in the administration of chemotherapy, and associated laboratory tests, such as blood tests for monitoring tumours [Lober et al, 1988; Lokich et al, 1996]. Disposable items included syringes, drug reservoirs, intravenous tubings, and solutions.

The cost of an outpatient visit was a key unit cost in calculating the cost of the adjuvant therapies. Using data collected by the Institute of Public Finance estimates of net expenditure on outpatient attendances within Medical Oncology specialities were divided by the number of outpatient visits to estimate a mean cost per outpatient attendance [CIPFA, 1999]. Data were available from 100 Trusts, leading to an average cost of £86.35, though the standard deviation of £89.84 highlights a wide range of variation.

Table A7.14 presents the resource use estimated for standard anti-emetics administered alongside the chemotherapy, the assumed health service contact per cycle of chemotherapy, and the disposable items employed per cycle.

Table A7.14 Estimates for anti-emetic schedules for CMF and CAF regimens per cycle, disposable items, and clinicians for health professional's time

CMF	Unit cost (£'s)	Cost (£'s)	CAF	Unit cost (£'s)	Cost (£'s)
Dexamethasone 8mg iv days 1+8	1.27 – 8mg	2.54	Granisetron 1mg iv	12.00 – 1mg	12.00
Dexamethasone 2mg po tds for 3 days following days 1+8	0.085 – 2mg	1.53	Dexamethasone 8mg iv	1.27 – 8mg	1.27
Domperidone 20mg po qds for 3 days following days 1+8	0.17 – 20mg	4.08	Dexamethasone 4mg po tds for 3 days	0.17 – 4mg	1.53
Folonic acid 15mg q.d.s. 3 days	3.71 – 15mg	44.52	Domperidone 20mg po qds for 3 days	0.17 – 20mg	2.04
Disposables[1]		14.47	Disposables[2]		11.31
Tests[1]		10.49	Tests[1]		10.49
2 breast clinic visit		172.70	1 breast clinic visit		86.35
Total		250.33			124.99

* Drug costs taken from the British National Formulary

[1] [Lober et al, 1988] Costs are converted using 1987 Danish crowns:£ exchange rate, and uprated to 1999 levels. Laboratory test costs are taken as the difference between the total lab test costs for CMF and the total lab test costs for the control group.

[2] [Lokich et al, 1996], Charges are converted using 1995 $:£ exchange rate, and uprated to 1999 levels. Clinic visits incorporate the administration of the drugs.

The baseline cost for a cycle of chemotherapy was calculated separately for an alkylating drugs-based regimen (CMF) and a cytotoxic antibiotics-based regimen (CAF), which are presented in Table A7.15.

Table A7.15 Total costs per cycle for CMF and CAF chemotherapy regimens

Components	CMF (£'s)	CAF (£'s)
Chemotherapeutic drugs	34.61	98.50
Anti-emetic drugs	52.67	16.84
Disposables	14.47	11.31
Tests	10.49	10.49
Clinic visits	172.70	86.35
Total	284.94	223.49

The cost of chemotherapy is an integral parameter in the model and a wide search for alternative estimates of the cost of a cycle was undertaken. Aggregate estimates combining the individual components, as well as the identified aggregated estimates are presented in Table A7.16.

Table A7.16 Cost estimates of cycle of chemotherapy

Method	Source	Rank	Weight	Cost
Separate	CMF, baseline	1	10	284.94
	CAF, baseline	2	9	226.65
	CMF, baseline (1 clinic visit)	3	8	198.59
	CMF, health professionals time costs[1]	6	5	141.86
	CMF, health professionals time costs[2]	7	4	700.46
	CAF, health professionals time costs[2]	8	3	467.60
Aggregate	NHS reference cost (HRG v.3)	4	7	269
	E Anglia Trust cost	5	6	67
	cost[3] CMF	9	2	596
	cost[3] CAF	10	1	648

[1] [Lober et al, 1988], [2] [Lokich et al, 1996], [3] [Silva and Zurrida, 1999]

Tamoxifen, 30mg/day, costs less than £7.00 per month, so no distribution was specified for such a small total cost, which is incurred by the majority of patients.

A7.3.2 Surveillance

Table A7.17 describes a selection of the follow-up schedules employed to survey patients in the absence of symptoms following primary surgery, and their associated costs.

Table A7.17 Components of follow-up reported in the literature for patients with no symptoms of disease

Treatment arms	Physical exam	Biochemical tests	Chest & bone x-rays	Mammograms	Electrocar-diograms	Total
Unit costs	89.55*	7.08[1]	122.91[1]	11.29[2]	16.94‡	
12 cycles CMF/ control[3]	3 weekly for 9 months, then 6 monthly	6 monthly for 3 years, then annually	6 monthly	Annually		
First year†	97.28	1.17	20.49	0.94	0.00	119.88
Subsequent years†	14.93	0.59	20.49	0.94	0.00	36.95
Radiotherapy/ Chemotherapy/ Tamoxifen[4]	3 monthly for 2 years, 6 monthly to 5 years, then annually	Only in case of symptoms	Only in case of symptoms	Annually		
First year†	29.85	0.00	0.00	0.94	0.00	30.79
Subsequent years†	14.93	0.00	0.00	0.94	0.00	15.87
12 cycles A/CMF[5]	3 weekly for 1 year, 6 monthly to 5 years, then annually	6 monthly for five years, then annually	6 monthly for five years, then annually	Annually	Annually from end of 1 year	
First year†	97.28	1.17	20.49	0.94	0.00	119.88
Subsequent years†	14.93	1.17	20.49	0.94	1.41	38.94

* Physical exam is assumed to incorporate a visit to a breast clinic or a medical oncology outpatient visit.
† cost per month.
‡ Hospital Trust finance department.
[1] [Robertson et al, 1995], [2] [Johnston et al, 1998], [3] [Zambetti et al, 1992], [4] [Arriagada et al, 1992], [5] [Bonadonna et al, 1995]

An UK-based study compared two options for the surveillance of patients with metastatic breast cancer [Robertson et al, 1995]. The reported costs are presented in table A7.18. The surveillance costs following a locoregional or a metastatic relapse were assumed to be the same, and the four estimates reported were entered into a discrete uniform distribution.

Table A7.18 Annual surveillance costs reported by Robertson *et al* 1995[1]

Year	UICC assessment	Serum marker assessment
One	441.57	252.50
Two and beyond	304.62	123.50

UICC - the International Union against Cancer.
[1] [Robertson et al, 1995]

A7.3.3 Toxicity parameters

The costs of toxicity were estimated separately for the different categories of toxicity severity, for both adjuvant therapies. Within each category, a weighted cost was estimated that reflected the proportion of patients experiencing the various conditions defined within the group (see section A7.2). Various conditions that required inpatient episodes were specified within the major toxicity category. Cost estimates for each event were identified through a search of the NHS economic evaluation database [NHS Economic Evaluation Database], and the Office of Health Economics' economic evaluation database [OHE Health Economic Evaluations Database]. One Italian study provided adequate estimates for both life threatening (massive) Pulmonary Embolism (PE) and Deep Vein Thrombosis (DVT) [Lloyd et al, 1997]. The presented costs were converted to 1994 pounds, and uprated to £4036.80 and £1653.97, respectively. Holloway *et al* presented inpatient costs of four cerebrovascular events, of which 'ischemic cerebral infarction' and 'transient ischemic attack' were taken as the closest to the types of events experienced as a consequence of chemotherapy [Holloway et al, 1996]. The mean costs of the two events were £5981.06 and £2816.22, converted from US dollars and uprated from 1992 values. Finally, Quirk *et al* compared the cost of treating acute upper gastrointestinal bleeding by physician speciality [Quirk et al, 1997]. The cost of being treated by a gastroenterologist was £1900.42, converted from US dollars and uprated from 1995 values, which was the lowest estimate. Table A7.19 presents the resulting composition for major toxic events.

Table A7.19 Health events included in the major toxicity category, and related cost estimates

Major toxic events	Tamoxifen alone	Chemotherapy alone	Chemotherapy+ Tamoxifen	Cost Estimate (£'s)
Pulmonary Embolism	0.144	0.235	0.196	4037
Thrombosis	0.538	0.336	0.794	1654
Myocardial infarction		0.030	-	2816
Cerebral vascular accident		0.180	-	5981
Thromboembolic event*		0.099	-	2633
Cardiac event*	0.318	-	-	4398
Gastrointestinal bleeding		0.120	-	1900
Weighted cost (£'s)	2870	3153	2105	

* unspecified, cost taken as the weighted mean of the other thromboembolic/cardiac events

Table A7.20 presents detailed information on the anti-emetics used to control nausea and vomiting, and the anti-biotics associated with infection. Table A7.21 describes the assumptions made with respect to the treatment of all conditions included in the graded toxicity categories, and their respective costs.

Table A7.20 Components and costs of the treatment for nausea/vomiting and infection

Cost component	Unit cost (£'s)	Grade 1 or 2 toxicity	Cost (£'s)	Grade 3 or 4 toxicity	Cost (£'s)
Nausea/vomiting					
Ondansetron 8mg i.v.	13.50	8mg per cycle	13.50	16mg per cycle	27.00
Dexamethasone 8mg i.v.	1.76	8mg per cycle	1.76	16mg per cycle	3.52
Metoclopramide 10mg po	0.11			180mg per cycle	2.01
Lorazepam 2mg	0.02			2mg per day	0.54
Total per cycle			**15.26**		**33.07**
Infection					
Gentamicin, 2mL vial (80mg)	1.54			240mg/day for 7 days	32.34
Ceftazidime, 2g vial	19.80			4g/day for 7 days	277.20
Total					**309.54**

Table 21 Health events included in the graded toxicity categories, and related cost and utility estimates

Graded events	Grade 3 or 4 treatment	Tamoxifen		Tam+Che		Grade 1 or 2 treatment	Tamoxifen		Tam+Che	
		Propo-rtion	Cost per cycle	Propo-rtion	Cost per cycle		Propo-rtion	Cost per cycle	Propo-rtion	Cost per cycle
Nausea/vomiting	Anti-emetics, 4 days following each cycle	0.266	8.80	0.567	18.75	Anti-emetics, 4 days following each cycle	0.356	5.43	0.850	15.26
Stomatitis/mucositis	Mouth washes, 7 days following each cycle (Povidone-iodine)	-	-	0.166	0.19	No treatment	0.003	-	0.287	-
Diarrhea	Self-treatment	0.061	-	0.041	-	Self-treatment	0.093	-	0.266	-
Leukopenia	Postpone chemotherapy until recovery	-	-	0.385	-	Postpone chemotherapy for a week	-	-	0.920	-
Thrombocytopenia	No treatment	0.278	-	0.056	-	No treatment	0.004	-	0.150	-
Neurotoxicity	No treatment	-	-	0.061	-	No treatment	-	-	0.322	-
Anemia	Transfusion	-	-	0.140	28.00	No treatment	-	-	0.112	-
Infection*	Anti-biotics , 7 days (Gentamicin)	0.521	161.27	0.144	44.57	-	-	-	-	-
Fatigue	No treatment	-	-	0.018		No treatment	-	-	0.144	-
Neutropenia*	Inpatient stay[1]	0.061	99.35	0.048	78.18	-	-	-	-	-
Alopecia	Scalp cooling system	-	-	0.179	-	Scalp cooling system	-	-	0.464	-
Outpatient visits	An extra visit every cycle, plus an appointment with a dietician every cycle	1.000	129.55	1.000	129.55	An extra visit every two cycles	1.00	43.18	1.000	43.18
Total			**398.97**		**299.24**			**48.61**		**58.44**

* It was assumed that patients experienced infection or neutropenia only once. Thus, the cost of treatment was spread over the six cycles.

[1] [Leese, 1993]

A7.3.4 Costs of relapse

Differential costs were estimated for each of the four sites of relapse included in the structure of the model – locoregional, soft tissue, bone, and visceral.

A7.3.4.1 Locoregional

Protocols for the treatment of locoregional relapses were derived from the literature. Excisional biopsy is the recommended primary treatment for chest wall relapse, which may involve the resection of the chest wall, though in some cases incisional biopsy may be the only option [Jardines et al, 1993]. Radiation therapy should follow, with a minimum dose of 45-50Gy to elective sites and 55Gy to unexcised relapses [Willner et al, 1997]. Systemic treatments are commonly implemented in patients following a chest wall relapse, though their effect is unproven [Jardines et al, 1993; Willner et al, 1997]. Similar treatment patterns were identified for regional relapses, ipsilateral axillary relapses requiring repeat axillary dissection, whilst supraclavicular lymph node recurrences are excised when possible and radiated [Jardines et al, 1993; Silva and Zurrida, 1999].

Surgery for locoregional relapses most closely matched the 'Intermediate Breast Surgery' category in the National Schedule of Reference Costs [The new NHS - 1998 Reference Costs, 1998]. The number and cost of fractions of radiotherapy were obtained from the literature, which reported a mean number of 20 fractions [Aberzik et al, 1986; Toonkel et al, 1983], at a cost of a single fraction of radiotherapy of £41.92 [Read, 1994]. The cost of chemotherapy following locoregional relapse employed the same estimates derived for the cost of adjuvant chemotherapy. The baseline cost for the treatment of locoregional relapse, and the constituent parts, are presented in table A7.22 below.

Table A7.22 Baseline cost of treating a locoregional relapse

Treatment	Cost (£'s)
Surgery	780.00
Radiotherapy (20 fractions @ £41.92 each)	838.40
Chemotherapy	285.00
Total	1903.40

A7.3.4.2 Metastases

Protocols were also developed from the literature to estimate separate monthly costs for the treatment of the three sites of metastases. The costing of surveillance for patients with metastases has been reported above, and the same systemic therapies were assumed as described above. Table A7.23 presents details of the assumed local treatment of metastases. Data from the literature reported that bone metastases should be treated with radiation therapy to palliate pain [Jardines et al, 1993], though another study differentiated between localised and widespread bone pain and offered a range of alternative interventions [Leonard et al, 1994]. Pathological fractures are an infrequent problem and should be treated with internal [Jardines et al, 1993] or external [Leonard et al, 1994] fixation and radiation. Bisphosphonates (pamidronate and clodronate) are added to systemic therapy and are useful in reducing pain, analgesic use, fractures and hypercalcemia [Silva and Zurrida, 1999; Hortobagyi, 1998]. The assumed local treatment schedule for bone metastases included radiation therapy and bisphosphonates alongside systemic therapy. Following the validation process and the ensuing examination of a series of patient notes, the original estimate of patients with bone metastases receiving five episodes of radiotherapy per month was thought to be excessive. Instead, patients were assigned a 0.5 probability of undergoing one fraction of radiotherapy per month.

Little data was available on the specific interventions for patients with soft tissue metastases, only pain through soft tissue infiltration was mentioned, which was treated with non-steroidal anti-inflammatory drugs (prednisolone) [Leonard et al, 1994]. Patients with visceral metastases have the least favourable prognosis, local treatment is applied only to palliate symptoms, though only around 50 per cent of such patients will develop signs, such as coughs and dysponoea (difficulty in breathing) [Jardines et al, 1993]. Treatment consists of morphine and diazepam (BNF).

Table A7.23 Local treatments used in patients with metastases

	Unit price (£'s)	Dose	Cost per day
Bone metastases			
Sodium clodronate 400mg	1.52	1.6g per day	6.08
Radiotherapy (fraction)	41.92	1 per month	1.40
Soft tissue			
Prednisolone 5mg	0.016	10mg per day	0.03
Visceral			
Diazepam 5mg	0.0035	10mg per day	0.01
Morphine 10mg	0.116	60mg per day	0.69

Various sources were identified that provided data on the costs associated with inpatient episodes for patients experiencing metastases. The three UK-based studies are described first. Richards *et al* presented retrospective data on the aggregate cost of treating for 50 patients with advanced breast cancer [Richards et al, 1993]. The costs were presented in eight categories, and a total cost of £7620 (1991 costs) from the point of diagnosis of advanced breast cancer to death is calculated. Wolstenholme and Whynes [1998] presented the mean four-yearly costs of breast cancer treatment by stage. Taking stage 4 as the relevant stage, a mean of zero was presented for inpatient stay investigations, and £72 (1991 costs) for inpatient stay complications, though a sample size of 6 restricts the use of such estimates. The other UK study compared two options for the surveillance of patients with metastatic breast cancer (see section A7.3.2).

Hurley *et al* [1992] identified 128 patients and collected resource use information from the point of diagnosis until death or the last point of contact (records were treated as censored if patients were still alive) in Australia. Patients were classified into five groups - visceral, CNS, bone, local, and other. The final category included regional and soft tissue relapses. The classes were ordered hierarchically, and separate relapse episodes were defined by the occurrence of a more serious site of relapse. Inpatients days, daypatient attendance's, outpatient visits and common investigations for each category of relapse were presented. These resource estimates were collected and equivalent costs from the UK NHS applied. Data from English Trust returns including 55 Trusts shows that the mean cost per patient day in medical oncology specialities is £302.40 (uprated from 1996/7 prices), and a mean cost per outpatient visit as £86.35. Patient day includes both inpatients and day case, so the calculated costs were used for both types of hospital visit. Table A7.24 presents the details of the information used.

Table A7.24 Median monthly resource use data extracted from Hurley *et al* [1992], and the costs applied from the UK NHS

Cost category	Visceral				Bone			
	Units*		Cost (£'s)*		Units*		Cost (£'s)*	
Inpatient days	0.63	2.67	190.51	807.41	0.05	1.09	15.12	329.62
Daypatients attendance's	0.33	0.67	99.79	202.61	0	0.33	0.00	99.79
Outpatient visits	1.17	1.57	101.03	135.57	1	1.33	86.35	114.85
Total			391.33	1145.59			101.47	544.25
Cost category	Other				Local			
	Units*		Cost (£'s)*		Units*		Cost (£'s)*	
Inpatient days	0.15	1.06	45.36	320.54	0	0.19	0.00	57.46
Daypatients attendance's	0.07	0.15	21.17	45.36	0	0.1	0.00	30.24
Outpatient visits	1.17	1.53	101.03	132.12	1	1.33	86.35	114.85
Total			167.56	498.02			86.35	202.54

* The left column presents the median number of units/cost, the right column presents the 75th percentile number of units/cost (1999£'s), inpatient day/day case = £302.40, outpatient visit = £86.35

A7.4 Utility values

Quality of life in breast cancer patients has been investigated, but the majority of studies have used condition- or symptom- specific measures [Macquart-Moulin et al, 1999; Fairclough, 1997]. The non-generic QoL data identified included the Rotterdam Symptom Checklist [Jonat et al, 1996; Richards et al, 1992], the Functional Living Index [Chu et al, 1996; The Givio investigators, 1994], the Perceived Adjustment to Chronic Illness Scale (PACIS) [Bernhard et al, 1997; Hurny et al, 1996], the Symptom Checklist-90 Revised [Tross et al, 1996]. In addition, a range of individual dimensions such as body image, emotional wellbeing and satisfaction with care, were valued using a Visual Analogue Scale (VAS) [The Givio investigators, 1994; Hayes et al, 1995]. The conversion of such measures to utilities was not attempted, though the data provided could be used to inform descriptions of health states for the primary collection of health state utilities.

A number of previous modelling studies had assigned utility values using either focus groups of oncology professionals [Hillner et al, 1992; Hillner and Smith, 1991; Smith and Hillner, 1993; Desch et al, 1993], or direct measurement techniques from oncology nurses [Hutton et al, 1996]. Table A7.25 summarises the data available from these studies.

Table A7.25 Utility values derived from previous modelling studies

From start of adjuvant therapy[1]		First-line therapy for metastases[2]		Second-line therapy for metastases[3]*		
Tamoxifen	0.99	Standard CT	0.7	Before second line therapy	0.59	0.56
Minor toxicity(CT)	0.9	Induction high dose CT	0.5	CR+PR	0.81	0.84
Major toxicity(CT)	0.7	Uncomplicated ABMT	0.3	Partial respose and severe peripheral oedema	0.75	0.78
1st recurrence	0.7	Complicated ABMT	0.1	Partial response and severe peripheral neuropathy	0.53	0.62
After 1st recurrence	0.85	CR	0.85	Stable disease	0.62	0.62
2nd recurrence	0.5	PR	0.6	Progressive disease	0.41	0.33
After 2nd recurrence	0.7	Stable	0.5	Sepsis	0.2	0.16
3rd recurrence	0.3	PD	0.4	Terminal disease	0.16	0.13

* Values in left column were derived from UK nurses only, values in the right column were derived from an international set of nurses.

[1] [Smith and Hillner, 1993], [2] [Hillner et al, 1992], [3] [Hutton et al, 1996]

Two papers were identified that collected primary estimates of utility values relating to some of the health states within the case study models. De Haes *et al* [1991] used 15 members of the Department of Public Health and Social Medicine and 12 health professionals to value 15 health states on a visual analogue scale (VAS) [de Haes et al, 1991]. Citing a function described by Loomes [1988], the median VAS scores were converted to time trade-off (TTO) scores that were taken as more accurate valuations of utility. Ashby *et al* [1994] described five health states, one year post-primary surgery combining alternative forms of surgery with two scenarios - coping well and coping poorly. TTO was used to value the states using 138 subjects, including health professionals, breast cancer patients and university staff. The data derived from these two studies are presented in table A7.26. A final study used patients with any type of metastatic cancer to compare three form of QoL or utility measurement - VAS, TTO and the Spitzer Quality of life Index (QLI). The mean scores for the former two measures were 0.41 and 0.63 [Perez et al, 1997].

Table A7.26 Utility values derived from primary studies

[de Haes et al, 1991]	VAS	TTO	[Ashby et al, 1994]	TTO
Initial surgery	62	0.867	Lumpectomy, occasional concern over relapse, normal activities, well supported	0.784
2mths-1yr after lump	71	0.914	Mastectomy+plastic surgery, occasional concern over relapse, normal activities, well supported	0.714
2mths-1yr after mast	65	0.844	Mastectomy, occasional concern over relapse, normal activities, well supported	0.703
Initial radiotherapy	60	0.803	Lumpectomy, swelling of arm, high anxiety and impact on normal life, not supported	0.284
Initial hormonal therapy	63	0.82	Mastectomy, swelling of arm, high anxiety and impact on normal life, not supported	0.257
Initial Chemotherapy	50	0.717		
DF>1yr after mastectomy	77	0.947		
DF>1yr after BCT	82	0.96		
Palliative+ Surgery	46	0.617		
Palliative+ Chemotherapy	36	0.531		
Palliative+ hormonal therapy	47	0.663		
Palliative+ Radiotherapy	43	0.591		
Terminal illness	19	0.288		

VAS=visual analogue scale, TTO=time trade-off
lump=lumpectomy, mast=mastectomy, plas=plastic surgery to make a new breast

The whole range of utility values used in the case study models is presented in Table A7.27.

Table A7.27 Utility values for all events included in the case study models

Health state	Minimum	Likeliest	Maximum
Disease free interval:			
First year	0.7	0.84	0.9
Subsequent years	0.8	0.95	1
Toxicity:			
Major	0.3	0.51	0.7
Grade ½	0.6	0.78	0.9
Grade ¾	0.55	0.65	0.85
Menopausal symptoms:			
Moderate	0.61	0.79	0.85
Severe	0.3	0.64	0.85
Locoregional relapse	0.3	0.5	0.7
Remission	0.6	0.84	0.9
Metastases	0.3	0.55	0.7

References

Abdool-Carrim T, Adler H, Becker P, Carides M, Ginsberg J, Golele R, Grobler G, Immelman E, Louwrens H, Lukhele M, Veller M, Watt K, Weber F (1997) The cost and benefit of prophylaxis against deep-vein thrombosis in elective hip replacement. South African Medical Journal 87: 594-600

Aberzik, W. J, Silver, B, Henderson, C, Cady, B, and Harris, J. R (1986) The use of radiotherapy for treatment of isolated locoregional recurrence of breast carcinoma after mastectomy. Cancer 58, 1214-8.

Aisner, J., Cirrincione, C., Perloff, M., Perry, M., Budman, D., Abrams, J. et al. (1995 Jun) Combination chemotherapy for metastatic or recurrent carcinoma of the breast--a randomized phase iii trial comparing caf versus vath versus vath alternating with cmfvp: cancer and leukemia group b study 8281. Journal of Clinical Oncology 13[6], 1443-52.

Alonso, M. C., Tabernero, J. M., Ojeda, B., Llanos, M., Sola, C., Climent, M. A. et al. (1995 Apr) A phase iii randomized trial of cyclophosphamide, mitoxantrone, and 5-fluorouracil (cnf) versus cyclophosphamide, adriamycin, and 5-fluorouracil (caf) in patients with metastatic breast cancer. Breast Cancer Research & Treatment 34[1], 15-24.

Anton, S. F. and Revivki, D. A (1995) The use of decision analysis in the pharmacoeconomic evaluation of antidepressant: a cost effectiveness study of nefazodone. Psychopharmacology Bulletin 31[2], 249-259.

Arai, Y., Sone, Y., Inaba, Y., Ariyoshi, Y., and Kido, C. (1994) Hepatic arterial infusion chemotherapy for liver metastases from breast cancer. Cancer Chemotherapy & Pharmacology. 33[suppl], S142-4.

Arriagada, R., Rutqvist, L. E., Kramar, A., and Johansson, H. (1992 Nov) Competing risks determining event-free survival in early breast cancer. British Journal of Cancer 66[5], 951-7.

Ashby, J., O'Hanlon, M., and Buxton, M. J. (1994 Aug) The time trade-off technique: how do the valuations of breast cancer patients compare to those of other groups? Quality of Life Research 3[4], 257-65.

Bagust, A., Place, M., and Posnett, J. W. (1999 Jul 17) Dynamics of bed use in accommodating emergency admissions: stochastic simulation model. British Medical Journal 319[7203], 155-8.

Baker, M. S., Kessler, L. G., Urban, N., and Smucker, R. C. (1991) Estimating the treatment costs of breast and lung cancer. Medical Care 29[1], 40-49.

Balmer DW, Paul RP (1986) CASM - The right environment for simulation. Journal of the Operational Research Society 37: 443-52

Barber, J. A. and Thompson, S. G. (1998) Analysis and interpretation of cost data in randomised controlled trials: review of published studies. British Medical Journal 317[7167], 1195-200.

Barnett, V. (1999) Comparative Statistical Inference, 3rd ed. Chichester: John Wiley & Sons

Barton P, Robinson S, Bryan S (forthcoming) The use of modelling in the economic evaluation of health care. Health Economics

Bastholt, L., Dalmark, M., Gjedde, S. B., Pfeiffer, P., Pedersen, D., Sandberg, E. et al. (1996 Apr) Dose-response relationship of epirubicin in the treatment of postmenopausal patients with metastatic breast cancer: a randomized study of epirubicin at four different dose levels performed by the Danish breast cancer cooperative group. Journal of Clinical Oncology 14[4], 1146-55.

Baum, M., Houghton, J., and Riley, D. (1992) Results of the cancer research campaign adjuvant trial for perioperative cyclophosphamide and long-term tamoxifen in early

breast cancer reported at the tenth year of follow-up. Cancer research campaign breast cancer trials group. Acta Oncologica 31[2], 251-7.

Beck JR, Kattan MW, Miles BJ (1994) A critique of the decision analysis for clinically localized prostate cancer. Journal of Urology 152: 1894-9

Beck JR, Pauker SG (1983) The Markov process in medical prognosis. Medical Decision Making 3: 419-458

Beck JR, Scardino PT (1994) How useful are models of natural history in clinical decision making and clinical research? [Editorial: comment]. Journal of Clinical Epidemiology 47: 1-2

Beck, J. R. and Pauker, S. G. (1983) The Markov process in medical prognosis. Medical Decision Making 3[4], 419-458.

Beck, T. M., Hart, N. E., Woodard, D. A., and Smith, C. E. (1983 Jun) Local or regionally recurrent carcinoma of the breast: results of therapy in 121 patients. Journal of Clinical Oncology 1[6], 400-5.

Bedwinek, J. M., Fineberg, B., Lee, J., and Ocwieza, M. (1981 May) Analysis of failures following local treatment of isolated local- regional recurrence of breast cancer. International Journal of Radiation Oncology Biology and Physics. 7[5], 581-5.

Berger, J. O. (1980) Statistical Decision Theory and Bayesian Analysis, 2nd ed. New York: Springer-Verlag

Bergman, L., van Dongen, J. A., van Ooijen, B., and van Leeuwen, F. E. (1995 Apr) Should tamoxifen be a primary treatment choice for elderly breast cancer patients with locoregional disease?. Breast Cancer Research & Treatment 34[1], 77-83.

Bernhard, J., Hurny, C., Coates, A. S., Peterson, H. F., Castiglione-Gertsch, M., Gelber, R. D. et al. (1997 Sep) Quality of life assessment in patients receiving adjuvant therapy

for breast cancer: the IBCSG approach. The international breast cancer study group. Annals of Oncology 8[9], 825-35.

Berry, D. A. and Stangl, D. K. (1996) Bayesian methods in health-related research, In: Berry, D. A. and Stangl, D. K. Bayesian Biostatistics. New York: Marcel Dekker, Inc.

Blomqvist, C., Elomaa, I., Rissanen, P., Hietanen, P., Nevasaari, K., and Helle, L. (1993 Mar) Influence of treatment schedule on toxicity and efficacy of cyclophosphamide, epirubicin, and fluorouracil in metastatic breast cancer: a randomized trial comparing weekly and every-4-week administration. Journal of Clinical Oncology 11[3], 467-73.

Boccardo, F., Rubagotti, A., Amoroso, D., Perrotta, A., Sismondi, P., Farris, A. et al. (1997) Lack of effectiveness of adjuvant alternating chemotherapy in node-positive, estrogen-receptor-negative premenopausal breast cancer patients: results of a multicentric italian study. The breast cancer adjuvant chemo-hormone therapy cooperative group (grocta). Cancer Investigation 15[6], 505-12.

Bolger PG, Davies R (1992) Simulation model for planning renal services in a district health authority. British Medical Journal 305: 605-8

Bolger, P. G. and Davies, R. (1992 Sep 12) Simulation model for planning renal services in a district health authority. British Medical Journal 305[6854], 605-8.

Bonadonna, G., Zambetti, M., and Valagussa, P. (1995 Feb 15) Sequential or alternating doxorubicin and cmf regimens in breast cancer with more than three positive nodes. Ten-year results [see comments]. Journal of the American Medical Association 273[7], 542-7.

Borner, M. M., Bacchi, M., and Castiglione, M. (1996 Nov) Possible deleterious effect of tamoxifen in premenopausal women with locoregional recurrence of breast cancer. European Journal of Cancer 32A[12], 2173-6.

Borner, M., Bacchi, M., Goldhirsch, A., Greiner, R., Harder, F., Castiglione, M. et al. (1994 Oct) First isolated locoregional recurrence following mastectomy for breast

cancer: results of a phase iii multicenter study comparing systemic treatment with observation after excision and radiation. Swiss group for clinical cancer research. Journal of Clinical Oncology 12[10], 2071-7.

Bratley, P., Fox, B. L., and Schrage, L. E. (1987) A Guide to Simulation, 2nd ed. New York : Springer-Verlag

Briggs, A. and Fenn, P. (1998 Dec) Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Economics 7[8], 723-40.

Briggs, A. and Gray, A. (1998 Oct) The distribution of health care costs and their statistical analysis for economic evaluation. Journal of Health Services Research & Policy 3[4], 233-45.

Briggs, A. and Sculpher, M. (1995 Sep-Oct) Sensitivity analysis in economic evaluation: a review of published studies. [Review] [109 refs]. Health Economics 4[5], 355-71.

Briggs, A. and Sculpher, M. (1998) An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 13[4], 397-409.

Briggs, A. H. (1999 May) A bayesian approach to stochastic cost-effectiveness analysis. Health Economics 8[3], 257-61.

Briggs, A. H. (2000) Handling uncertainty in cost-effectiveness models. Pharmacoeconomics 17[5], 479-500.

Briggs, A. H. and Gray, A. M. (1999 Sep 4) Handling uncertainty in economic evaluations of healthcare interventions. British Medical Journal 319[7210], 635-8.

Briggs, A. H., Mooney, C. Z., and Wonderling, D. E. (1999) Constructing confidence intervals for cost-effectiveness ratios: an evaluation of parametric and non-parametric techniques using Monte Carlo simulation. Statistics in Medicine 18 [23], 3245-62.

Briggs, A., Sculpher, M., and Buxton, M. (1994 Mar-Apr) Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. Health Economics 3[2], 95-104.

Bryson, H. M. and Plosker, G. L. (1993) Tamoxifen. A review of pharmacoeconomic and quality-of-life considerations for its use as adjuvant therapy in women with breast cancer. Pharmacoeconomics 4[1], 40-66.

Budd, G. T., Green, S., O'Bryan, R. M., Martino, S., Abeloff, M. D., Rinehart, J. J. et al. (1995 Apr) Short-course fac-m versus 1 year of cmfvp in node-positive. hormone receptor-negative breast cancer: an intergroup study. Journal of Clinical Oncology 13[4], 831-9.

Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, Vray M (1997) Modelling in economic evaluation: an unavoidable fact of life. Health Economics 6: 217-27

Buxton, M. (1993) Scarce resources and informed choices, In: Ashton, D. Future Trends in Medicine.Royal Society of Medicine Services International Congress and Symposium Series No. 202.

Buxton, M. J. (1993) Health economics in the 1990's, In: Jonsson, B. and Rosenbaum, J. Health Economics of Depression, 1st ed.Wiley; 1-13.

Buxton, M. J., Drummond, M. F., Van Hout, B. A., Prince, R. L., Sheldon, T. A., Szucs, T. et al. (1997 May-Jun) Modelling in economic evaluation: an unavoidable fact of life. Health Economics 6[3], 217-27.

Campbell, H., Karnon, J., and Dowie, R. (2000) Cost analysis of hospital-at-home initiative using discrete event simulation. Journal of Health Services Research and Policy 5[4].

Cancer Research Campaign (1998) Fact sheet 1: Incidence. London: CRC Publications

Castiglione-Gertsch, M., Johnsen, C., Goldhirsch, A., Gelber, R. D., Rudenstam, C. M., Collins, J. et al. (1994 Oct) The international (ludwig) breast cancer study group trials i-iv: 15 years follow-up. Annals of Oncology 5[8], 717-24.

Castiglione-Gertsch, M., Tattersall, M., Hacking, A., Goldhirsch, A., Gudgeon, A., Gelber, R. D. et al. (1997 Dec) Retreating recurrent breast cancer with the same cmf-containing regimen used as adjuvant therapy. The international breast cancer study group. European Journal of Cancer 33[14], 2321-5.

Chacon, R., Romero Acuna, L., Blajman, C., Galvez, C., Bruno, M., Romero, A. et al. (1997) Less efficacy with alternating regimen as adjuvant chemotherapy in stage ii node-positive breast cancer: results at 8 years. British Journal of Cancer 76[4], 545-50.

Chaussalet, T., El-Darzi, E., and Millard, P. H. (July 1999) Markov and discrete event simulation models for the economic evaluation of alternative options for dementia services. www.wmin.ac.uk/~chausst .

Chu, L., Sutton, L. M., Peterson, B. L., Havlin, K. A., and Winer, E. P. (1996 Fall) Continuous infusion 5-fluorouracil as first-line therapy for metastatic breast cancer. Journal of Infusional Chemotherapy 6[4], 211-6.

CIPFA (1999) A Review of Financial Management and Performance Reporting within Health Authorities. London:

Clark, G. M., Sledge, G. W. Jr, Osborne, C. K., and McGuire, W. L. (1987 Jan) Survival from first recurrence: relative importance of prognostic factors in 1,015 breast cancer patients. Journal of Clinical Oncology 5[1], 55-61.

Claxton, K. (1999 Jun) The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. Journal of Health Economics 18[3], 341-64.

Claxton, K. and Posnett, J. (1996 Nov-Dec) An economic approach to clinical trial design and research priority-setting. Health Economics 5[6], 513-24.

Claxton, K., Neumann, P. J., Araki, S. S., and Weinstein, M. C. (1998) The value of information in the economic-evaluation of new health-care technologies - an application to a policy model of alzheimers-disease. Medical Decision Making 18[4], 484-484.

Cobau, C. D., Declercq, K., Neuberg, D., Ingle, J. N., and Tormey, D. C. (1996 Feb 1) A randomized trial of megestrol acetate with or without premarin in the treatment of potentially responsive metastatic breast cancer. A study of the eastern cooperative oncology group (e2185). Cancer 77[3], 483-9.

Cochrane, W. G. (1977) Sampling Techniques, 3rd ed. New York: John Wiley & Sons

Coleman, R. E. and Rubens, R. D. (1987) The clinical course of bone metastases from breast cancer. British Journal of Cancer. 55[1], 61-6.

Coombes, R. C., Bliss, J. M., Wils, J., Morvan, F., Espie, M., Amadori, D. et al. (1996 Jan) Adjuvant cyclophosphamide, methotrexate, and fluorouracil versus fluorouracil, epirubicin, and cyclophosphamide chemotherapy in premenopausal women with axillary node-positive operable breast cancer: results of a randomized trial. The international collaborative cancer group. Journal of Clinical Oncology 14[1], 35-45.

Cooper, H. M. (1998) Synthesizing research : a guide for literature reviews, 3rd ed. Thousand Oaks, Calif. ; London: Sage Publications

Cronin, K. A., Legler, J. M., and Etzioni, R. D. (1998) Assessing Uncertainty in Micro Simulation Modeling with Application to Cancer Screening Interventions. Statistics in Medicine 17[21], 2509-2523.

Crowe, J. P. Jr, Gordon, N. H., Antunez, A. R., Shenk, R. R., Hubay, C. A., and Shuck, J. M. (1991 Apr) Local-regional breast cancer recurrence following mastectomy. Archives of Surgery 126[4], 429-32.

Crystal Ball (2000) USA: Decisioneering, Inc;

Cummings, F. J., Gray, R., Tormey, D. C., Davis, T. E., Volk, H., Harris, J. et al. (1993 Jan) Adjuvant tamoxifen versus placebo in elderly women with node-positive breast cancer: long-term follow-up and causes of death [see comments]. Journal of Clinical Oncology 11[1], 29-35.

Daly, E., Gray, A., Barlow, D., McPherson, K., Roche, M., and Vessey, M. (1993 Oct 2) Measuring the impact of menopausal symptoms on quality of life. British Medical Journal 307[6908], 836-40.

Danese, M. D., Powe, N. R., Sawin, C. T, and Ladenson, P. W (1996) Screening for mild thyroid failure at the periodic health examination. Journal of the American Medical Association 276[4], 285-292.

DATA 3.5 (1999) Williamstown, USA: Treeage;

Davies A, Langley PC, Keks NA, Catts SV, Lambert T, Schweitzer I (1998) Risperiodone versus haloperidol: II. Cost effectiveness. Clinical Therapeutics 20: 196-213

Davies H, Davies R (1987) A stimulation model for planning services for renal patients in Europe. J. Operational Research Society 38: 639-700

Davies R and Roderick, P (1998) Planning resources for renal services throughout the UK using simulation. European Journal of Operational Research 105, 285-295.

de Haes, J. C., de Koning, H. J., van Oortmarssen, G. J., van Agt, H. M., de Bruyn, A. E., and van Der Maas, P. J. (1991 Oct 21) The impact of a breast cancer screening programme on quality-adjusted life-years. International Journal of Cancer 49[4], 538-44.

de Takats, P. G., Dunn, J. A., Kerr, D. J., and Morrison, J. M. (1993) Impact of adjuvant chemotherapy in breast cancer on response to tamoxifen at relapse. Cancer Treatment Reviews. 19, 11-9.

Delong, E. R. and Simons, T. (1999) Statistical Issues in Cost-Effectiveness. American Heart Journal 137[5], S47-S50.

Desch, C. E., Hillner, B. E., Smith, T. J., and Retchin, S. M. (1993 Apr) Should the elderly receive chemotherapy for node-negative breast cancer? A cost-effectiveness analysis examining total and active life-expectancy outcomes. Journal of Clinical Oncology 11[4], 777-82.

Detsky, A. S. (1989 Oct 6) Are clinical trials a cost-effective investment? [See comments]. Journal of the American Medical Association 262[13], 1795-800.

Detsky, A. S., Naglie, G., Krahn, M. D., Redelmeier, D. A., and Naimark, D. (1997 Apr-Jun) Primer on medical decision analysis: part 2--building a tree [see comments]. Medical Decision Making 17[2], 126-35.

Doubilet, P., Begg, C. B., Weinstein, M. C., Braun, P., and NcNeil, B. J. (1985) Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. Medical Decision Making 5[2], 157-77.

Dowie, R. (May 1996) Guidance for purchasers of cardiac services: Using a decision analytic framework. Health Information Utilisation Project, Thames Postgraduate Medical and Dental Education, University of London.

Drummond, M. (1992) Cost-effectiveness guidelines for reimbursement of pharmaceuticals: Is economic evaluation ready for its enhanced status? Health Economics 1, 85-92.

Drummond, M. F., Davies, L. M., and Ferris III, F. L. (1992) Assessing the costs and benefits of medical research: the diabetic retinopathy study. Social Science and Medicine 34[9], 973-981.

Drummond, M. F., O'Brien, B., Stoddart, G. L., and Torrance, G. W. (1997) Methods for the economic evaluation of health care programmes, 2nd ed. Oxford University Press

Earle, C. C., Chapman, R. H., Baker, C. S., Bell, C. M., Stone, P. W., Sandberg, E. A. et al. (2000) Systematic overview of cost-utility assessments in oncology. Journal of Clinical Oncology 18[18], 3302-17.

Earle, C. G. and Wells, G. A. (2000) An assessment of methods to combine published survival curves. Medical Decision Making 20, 104-11.

EBCTCG (1992 Jan 11) Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. The Lancet 339, 71-85.

EBCTCG (1998 May 16) Tamoxifen for early breast cancer: an overview of the randomised trials. The Lancet 351[9114], 1451-67.

EBCTCG (1998 Sep 19) Polychemotherapy for early breast cancer: an overview of the randomised trials. The Lancet 352[9132], 930-42.

Eckman, M. H., Greenfield, S., Mackey, W. C., Wong, J. B., Kaplan, S., Sullivan, L. et al. (1995) Foot infections in diabetic patients: decision and cost-effectiveness analyses. Journal of the American Medical Association 273[9], 712-720.

Eckman, M. H., Levine, H. J., Salem, D. N., and Pauker, S. G. (1998 Nov) Making decisions about antithrombotic therapy in heart disease: decision analytic and cost-effectiveness issues. Chest 114[5 Suppl], 699S-714S.

Eddy, D. M. (1985) Technology assessment: the role of mathematical modelling, In: Mosteller, F. Assessing Medical Technologies. Washington, DC: National Academy Press; 144-53.

Eddy, D. M. (1989) Selecting technologies for assessment. International Journal of Technology Assesment in Health Care 5, 485-501.

Eddy, D. M., Hasselblad, V., and Shachter, R. (1990) A Bayesian method for synthesing evidence: The confidence profile approach. International Journal of Technology Assessment in Health Care 6, 31-55.

Eddy, D. M., Hasselblad, V., and Shachter, R. (1990) An introduction to a Bayesian method for meta-analysis: The confidence profile method. Medical Decision Making 10, 15-23.

Fairclough, D. L. (1997 Jun 15) Summary measures and statistics for comparison of quality of life in a clinical trial of cancer therapy. Statistics in Medicine 16[11], 1197-209.

Falkson, G., Holcroft, C., Gelman, R. S., Tormey, D. C., Wolter, J. M., and Cummings, F. J. (1995 Jun) Ten-year follow-up study of premenopausal women with metastatic breast cancer: an eastern cooperative oncology group study. Journal of Clinical Oncology 13[6], 1453-8.

Felli, J. C. and Hazen, G. B. (1998 Jan-Mar) Sensitivity analysis and the expected value of perfect information. Medical Decision Making 18[1], 95-109.

Felli, J. C. and Hazen, G. B. (1999 May) A bayesian approach to sensitivity analysis. Health Economics 8[3], 263-8.

Fenwick, E., Claxton, K., Sculpher, M., and Briggs, A. (2000) Improving the efficiency and relevance of health technology assessment: The role of iterative decision analytic modelling. CHE discussion paper, University of York .

Fisher, B. J., Perera, F. E., Cooke, A. L., Opeitum, A., Venkatesan, V., Dar, A. R. et al. (1997 Jun 1) Long-term follow-up of axillary node-positive breast cancer patients receiving adjuvant systemic therapy alone: patterns of recurrence. International Journal of Radiation Oncology, Biology, and Physics 38[3], 541-50.

Fisher, B., Dignam, J., Bryant, J., DeCillis, A., Wickerham, D. L., Wolmark, N. et al. (1996 Nov 6) Five versus more than five years of tamoxifen therapy for breast cancer

patients with negative lymph nodes and estrogen receptor-positive tumors [see comments]. Journal of the National Cancer Institute 88[21], 1529-1542.

Fisher, B., Dignam, J., Mamounas, E. P., Costantino, J. P., Wickerham, D. L., Redmond, C. et al. (1996 Jul) Sequential methotrexate and fluorouracil for the treatment of node-negative breast cancer patients with estrogen receptor-negative tumors: eight-year results from national surgical adjuvant breast and bowel project (nsabp) b-13 and first report of findings from nsabp b-19 comparing methotrexate and fluorouracil with conventional cyclophosphamide, methotrexate, and fluorouracil. Journal of Clinical Oncology 14[7], 1982-92.

Fisher, B., Dignam, J., Wolmark, N., DeCillis, A., Emir, B., Wickerham, D. L. et al. (1997 Nov 19) Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. Journal of the National Cancer Institute 89[22], 1673-82.

Fletcher, J., Hicks, N. R., Kay, J. D. S., and Boyd, P. A (1995) Using decision analysis to compare policies for antenatal screening for down's syndrome. British Medical Journal 311, 351-357.

Fowble, B. (1991) Staging of breast cancer, In: Fowble, B., Goodman, R. L., Glick, J. H., and Rosato, E. F. Breast Cancer Treatment. A Comprehensive Guide to Management. St Louis: Mosby Year Book.

Fox, K. R. (1991) Adjuvant therapy of node-positive operable breast cancer, In: Fowble, B., Goodman, R. L., Glick, J. H., and Rosato, E. F. Breast Cancer Treatment. A Comprehensive Guide to Management. St Louis: Mosby Year Book.

Fukutomi, T., Akashi, S., Nanasawa, T., and Yamamoto, H. (1995 Dec) Adjuvant six cycles of high-dose adriamycin, cyclophosphamide, methotrexate, 5-fluorouracil (acmf) vs. 12 cycles of low-dose acmf with tamoxifen for premenopausal, node-positive breast cancer patients: results of a prospective randomized study. Journal of Surgical Oncology 60[4], 242-6.

Gabra, H., Cameron, D. A., Lee, L. E., Mackay, J., and Leonard, R. C. (1996 Dec) Weekly doxorubicin and continuous infusional 5-fluorouracil for advanced breast cancer. British Journal of Cancer 74[12], 2008-12.

Gelber, R. D., Cole, B. F., and Goldhirsch, A. (1993) How to compare quality of life of breast cancer patients in clinical trials. International breast cancer study group. Recent Results in Cancer Research. 221-33. 127.

Gelber, R. D., Goldhirsch, A., and Cavalli, F. (1991) Quality of life adjusted evaluation of adjuvant therapies for operable breast cancer. Annals of Internal Medicine 114, 620-28.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995) Bayesian Data Analysis. London: Chapman & Hall

Gerard, J. P., Hery, M., Gedouin, D., Monnier, A., Goudier, M. J., Jacquin, J. P. et al. (1993) Postmenopausal patients with node-positive resectable breast cancer. Tamoxifen vs fec 50 (6 cycles) vs fec 50 (6 cycles) plus tamoxifen vs control--preliminary results of a 4-arm randomised trial. The french adjuvant study group. Drugs. 45, 60-7.

Gessner BD, Teutsch SM, Shaffer PA (1996) A cost-effectiveness evaluation of newborn hemoglobinopthy screening from the perspective of state health care systems. Early Human Development 45: 257-75

Glasziou, P. and Haas, M. (1994) An economic evaluation of the use of tamoxifen in the treatment of early breast cancer. CHERE Discussion Paper Series, Number 25 .

Glick, J. H. (1991) Adjuvant therapy for node-negative breast cancer, In: Fowble, B., Goodman, R. L., Glick, J. H., and Rosato, E. F. Breast Cancer Treatment. A Comprehensive Guide to Management. St Louis: Mosby Year Book.

Goldhirsch, A., Gelber, R. D., and Castiglione, M. (1988 Jan) Relapse of breast cancer after adjuvant treatment in premenopausal and perimenopausal women: patterns and prognoses. Journal of Clinical Oncology 6[1], 89-97.

Goldhirsch, A., Gelber, R. D., Price, K. N., Castiglione, M., Coates, A. S., Rudenstam, C. M. et al. (1994 Feb 12) Effect of systemic adjuvant treatment on first sites of breast cancer relapse [see comments]. The Lancet 343[8894], 377-81.

Goodwin, P. and Wright, G. (1998) Decision Analysis for Management Judgement , 2nd ed. Chichester : John Wiley and Sons

Gottfried, B. S. (1984) Elements of Stochastic Process Simulation. Englewood Cliffs : London: Prentice-Hall

Gregoire, G., Derderian, F., and Le Lorier, J. (1995) Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? Journal of Clinical Epidemiology 48[1], 159-63.

Gundersen, S., Hannisdal, E., Soreide, J. A., Skarstein, A., and Varhaug, J. E. (1995) Adjuvant tamoxifen for pre- and postmenopausal women with estrogen receptor positive, node positive breast cancer: a randomized study. Breast Cancer Research & Treatment 36[1], 49-53.

Habbema, J. D. F., Lubbe, J. Th. N., van Oortmarssen, G. J., and van der Maas, P. J. (1987) A simulation approach to cost-effectiveness and cost-benefit calculations of screening for the early detection of disease. European Journal of Operational Research 29, 159-166.

Habbema, J. D. F., van Oortmarssen, G. J., Lubbe, J. T. N., and van der Maas, P. J. (1984) The MISCAN simulation program for the evaluation of screening for disease. Computer Methods and Programs in Biomedicine 20, 79-93.

Hagard, S., Carter, F., and Milne, R. G. (1976) Screening for spina bifida cystica. A cost-benefit analysis. British Journal prev. soc. Medicine 30, 40-53.

Halpern, E. F., Weinstein, M. C., Hunink, M. G. M., and Scott Gazelle, G. (2000) Representing both first- and second-order uncertainties by Monte Carlo simulation for groups of patients. Medical Decision Making 20, 314-22.

Halpern, M. T, Luce, B. R, Brown, R. E, and Geneste, B (1998) Health and economic outcomes modeling practices: a suggested framework. Value in Health 1[2], 131-147.

Halpern, M. T., McKenna, M., and Hutton, J. (1998 Dec) Modeling in economic evaluation: an unavoidable fact of life. Health Economics 7[8], 741-2.

Harper, G., Townsend, J., and Buxton, M. (1998 Fall) The preliminary economic evaluation of health technologies for the prioritization of health technology assessments. A discussion. International Journal of Technology Assessment in Health Care 14[4], 652-62.

Harrold, E. V., Turner, B. C., Matloff, E. T., Pathare, P., Beinfield, M., McKhann, C. et al. (1998 Sep-Oct) Local recurrence in the conservatively treated breast cancer patient: a correlation with age and family history [see comments]. Cancer Journal From Scientific American 4[5], 302-7.

Hart, W. M., Espinosa, C., and Rovira, J. (1997) A simulation model of the cost of the incidence of IDDM in Spain. Diabetologia 40, 311-318.

Hasselblad, V. and McCrory, D. C. (1995 Jan-Mar) Meta-analytic tools for medical decision making: a practical guide. Medical Decision Making 15[1], 81-96.

Hayes, D. F., Van Zyl, J. A., Hacking, A., Goedhals, L., Bezwoda, W. R., Mailliard, J. A. et al. (1995 Oct) Randomized comparison of tamoxifen and two separate doses of toremifene in postmenopausal patients with metastatic breast cancer. Journal of Clinical Oncology 13[10], 2556-66.

Heitjan, D. F., Moskowitz, A. J., and Whang, W. (1999 Jan-Mar) Problems with interval estimates of the incremental cost-effectiveness ratio [see comments]. Medical Decision Making 19[1], 9-15.

Hillier, F. S. and Lieberman, G. J. (1995) Introduction to operations research, 6th ed. New York: McGraw-Hill

Hillner BE, Smith TJ, and Desch CE (1993) Assessing the cost-effectiveness of adjuvant therapies in early breast cancer using a decision analysis model. Breast Cancer Research & Treatment 24, 97-105.

Hillner, B. E. and Smith, T. J. (1991 Jan 17) Efficacy and cost effectiveness of adjuvant chemotherapy in women with node-negative breast cancer. A decision-analysis model. New England Journal of Medicine 324[3], 160-8.

Hillner, B. E. and Smith, T. J. (1992) Should women with node-negative breast cancer receive adjuvant chemotherapy?--Insights from a decision analysis model. Breast Cancer Research & Treatment 23[1-2], 17-27.

Hillner, B. E. and Smith, T. J. (March 1992) Estimating the efficacy and cost-effectiveness of tamoxifen versus tamoxifen plus adjuvant chemotherapy in post-menopausal node positive breast cancer: a decision analysis model. Proceedings of ASCO 11, 55.

Hillner, B. E., Smith, T. J., and Desch, C. E. (1992 Apr 15) Efficacy and cost-effectiveness of autologous bone marrow transplantation in metastatic breast cancer. Estimates using decision analysis while awaiting clinical trial results. Journal of the American Medical Association 267[15], 2055-61.

HM Treasury (1997) Appraisal and evaluation in central government. London HMSO .

Holloway, R. G., Witter, D. M. Jr, Lawton, K. B., Lipscomb, J., and Samsa, G. (1996 Mar) Inpatient costs of specific cerebrovascular events at five academic medical centers [see comments]. Neurology 46[3], 854-60.

Hortobagyi, G. N. (1998 Oct 1) Treatment of breast cancer. [Review] [106 refs]. New England Journal of Medicine 339[14], 974-84.

Hupperets, P. S., Wils, J., Volovics, L., Schouten, L., Fickers, M., Bron, H. et al. (1993 Apr) Adjuvant chemohormonal therapy with cyclophosphamide, doxorubicin and 5-fluorouracil (caf) with or without medroxyprogesterone acetate for node-positive breast cancer patients. Annals of Oncology 4[4], 295-301.

Hurley SF, Huggins RM, Snyder RD, and Bishop JF (1992) The cost of breast cancer recurrences. British Journal of Cancer 65, 449-455.

Hurny, C., Bernhard, J., Coates, A. S., Castiglione-Gertsch, M., Peterson, H. F., Gelber, R. D. et al. (1996 May 11) Impact of adjuvant therapy on quality of life in women with node-positive operable breast cancer. International breast cancer study group [published erratum appears in lancet 1997 jul 26;350(9073):298]. The Lancet 347[9011], 1279-84.

Hutton, J., Brown, R., Borowitz, M., Abrams, K., Rothman, M., and Shakespeare, A. (1996) A new decision model for cost-utility comparisons of chemotherapy in recurrent metastatic breast cancer. Pharmacoeconomics 9[suppl 2], 8-22.

Ibrahim, N. K., Frye, D. K., Buzdar, A. U., Walters, R. S., and Hortobagyi, G. N. (1996 Apr 22) Doxorubicin-based chemotherapy in elderly patients with metastatic breast cancer. Tolerance and outcome. Archives of Internal Medicine 156[8], 882-8.

Ingle, J. N., Foley, J. F., Mailliard, J. A., Krook, J. E., Hartmann, L. C., Jung, S. H. et al. (1994 May 1) Randomized trial of cyclophosphamide, methotrexate, and 5-fluorouracil with or without estrogenic recruitment in women with metastatic breast cancer. Cancer 73[9], 2337-43.

International breast cancer study group (1996 Jun) Duration and reintroduction of adjuvant chemotherapy for node-positive premenopausal breast cancer patients. Journal of Clinical Oncology 14[6], 1885-94.

Iverson, G. R. (1984) Bayesian Statistical Inference.Sage

Jaiyesimi, I. A., Buzdar, A. U., Decker, D. A., and Hortobagyi, G. N. (1995 Feb) Use of tamoxifen for breast cancer: twenty-eight years later. Journal of Clinical Oncology 13[2], 513-29.

Janjan, N. A., McNeese, M. D., Buzdar, A. U., Montague, E. D., and Oswald, M. J. (1986) Management of locoregional recurrent breast cancer. Cancer 58, 1552-6.

Jansen, R. B., Burrell, A., Nuijten, M. J. C., and Hardens, M (1996) An economic evaluation of meloxicam 7.5 mg versus diclofenac 100 mg retard in the treatment of osteoarthritis in the UK: a decision analysis model based on gastrointestinal complications. British Journal of Medical Economics 10, 247-262.

Jansen, S. J. T., Stiggelbout, A. M., Wakker, P. P., Vliet Vlieland, T. P. M., Leer, J-W H., Nooy, M. A. et al. (1998) Patients' utilities for cancer treatments: A study of the chained procedure for the standard gamble and time tradeoff. Medical Decision Making 18, 391-99.

Jardines, L., Callans, L. S., and Torosian, M. H. (1993 Oct) Recurrent breast cancer: presentation, diagnosis, and treatment. Seminars in Oncology 20[5], 538-47.

Johannesson, M., Hedbrant, J., and Jonsson, B. (1991) A computer simulation model for cost-effectiveness analysis of cardiovascular disease prevention. Medical Informatics 16[4], 355-362.

Johnston, K., Gerard, K., and Brown, J. (1998) Generalising costs from trials. Analyzing center selection bias in a breast screening trial. International Journal of Technology Assessment in Health Care 14[3], 494-504.

Jonat, W., Howell, A., Blomqvist, C., Eiermann, W., Winblad, G., Tyrrell, C. et al. (1996 Mar) A randomised trial comparing two doses of the new selective aromatase inhibitor anastrozole (arimidex) with megestrol acetate in postmenopausal patients with advanced breast cancer [see comments]. European Journal of Cancer 32A[3], 404-12.

Jonsson, B., Christiansen, C., Johnell, O., and Hedbrandt, J (1995) Cost-effectiveness of fracture prevention in established osteoporosis. Osteoporosis Inernational 5[136-142], 136-142.

Jonsson, B., Hedbrant, J., and Juhnell, O. (1993) A computer simulation model to analyse the cost-effectiveness of fracture prevention of osteoporosis. The Economic Research Institute, Stockholm School of Economics 6525.

Kalish, S. C., Gurwitz, J. H., Krumhoiz, H. M., and Avorn, J (1995) A cost-effectiveness model of thrombolytic therapy for acute myocardial infarction. Journal of General International Medicine 10, 321-330.

Kamby, C. and Sengelov, L. (1997 Sep) Pattern of dissemination and survival following isolated locoregional recurrence of breast cancer. A prospective study with more than 10 years of follow up. Breast Cancer Research & Treatment 45[2], 181-92.

Kamby, C., Rose, C., Ejlertsen, B., Andersen, J., Birkler, N. E., Rytter, L. et al. (1988 Mar) Adjuvant systemic treatment and the pattern of recurrences in patients with breast cancer. European Journal of Cancer & Clinical Oncology 24[3], 439-47.

Karnon J, Zeuner D, Ades A, Brown J, and Yardumian A (2001) The effects of neonatal screening for sickle cell disorders on lifetime treatment costs and early deaths avoided: a modelling approach. Journal of Public Health Medicine 22: 500-11

Karnon, J and Brown, J (1998) Selecting a decision model for economic evaluation: a case study and review. Health Care Management Science 1, 133-140.

Kassirer, J. P. (1976 May) The principles of clinical decision making: an introduction to decision analysis. Yale Journal of Biological Medicine 49[2], 149-64.

Kassirer, J. P. and Angell, M. (1994) The Journal's policy on cost-effectiveness analyses. New England Journal of Medicine 331, 669-70.

Kassirer, J. P., Moskowitz, A. J., Lau, J., and Pauker, S. G. (1987 Feb) Decision analysis: a progress report. Annals of Internal Medicine 106[2], 275-91.

Kattan, M. W., Miles, B. J., Beck, P. T., and Scardino, P. T. (1995) Second Order Monte Carlo simulation in the decision to treat clinically localized prostate cancer. Medical Decision Making 15[4], 429.

Kelton, W. D., Sadowski, R. P., and Sadowski, D. A. (1998) Simulation with Arena. Boston: McGraw-Hill

Klein RW, Dittus RS, Roberts SD, Wilson JR (1993) Simulation modeling and health-care decision making. Medical Decision Making 13: 347-54

Kobelt, G. and Jonsson, L. (1999) Modelling cost of treatments with new topical treatments for glaucoma: Results from France and the United Kingdom. International Journal of Technology Assessment in Health Care 15[1], 207-19.

Koenders, P. G., Beex, L. V., Kloppenborg, P. W., Smals, A. G., and Benraad, T. J. (1992) Human breast cancer: survival from first metastasis. Breast cancer study group. Breast Cancer Research & Treatment 21[3], 173-80.

Koopmanschap, M. A., van Ineveld, B. M., and Miltenburg, T. E. (1992 Oct) Costs of home care for advanced breast and cervical cancer in relation to cost-effectiveness of screening. Social Science & Medicine 35[8], 979-85.

Kuntz, K. M., Fleischmann, K. E., Hunink, M. G., and Douglas, P. S. (1999 May 4) Cost-effectiveness of diagnostic strategies for patients with chest pain. Annals of Internal Medicine 130[9], 709-18.

Leese, B. (1993) The costs of treating febrile neutropenia in six u.k. Hospitals. European Journal of Cancer. 29A[suppl 7], S15-8.

Legorreta, A. P., Brooks, R. J., Leibowitz, A. N., and Solin, L. J. (1996 Oct 28) Cost of breast cancer treatment. A 4-year longitudinal study. Archives of Internal Medicine 156[19], 2197-201.

Lehaney, B. and Paul, R. P. (1994) Using soft systems methodology to develop a simulation of out-patient services. Journal of the Royal Society of Health. 248-51.

Leonard, R. C., Rodger, A., and Dixon, J. M. (1994 Dec 3) Abc of breast diseases. Metastatic breast cancer. British Medical Journal 309[6967], 1501-4.

Liberati, A. (1995) The givio trial on the impact of follow-up care on survival and quality of life in breast cancer patients. Interdisciplinary group for cancer care evaluation. Annals of Oncology. 6[suppl 2], 41-6.

Lieu TA, Watson SE, Washington AE (1994) The cost-effectiveness of prenatal carrier screening for cystic fibrosis. Obstetrics and Gynaecology 84: 903-12

Lindeman, G. J., Boyages, J., Driessen, C., and Langlands, A. O. (1992 Jul) Intravenous or oral adjuvant cmf for node-positive breast cancer. Australian & New Zealand Journal of Surgery 62[7], 556-62.

Lipscomb, J., Weinstein, M. C., and Torrance, G. W. (1996) Time preference, In: Gold, M. R., Siegel, J. E., Russell, L. B., and Weinstein, M. C. Cost-Effectiveness in Health and Medicine .Oxford University Press.

Lipton, J., Shaw, W. D., Holmes, J., and Patterson, A. (1995) Short communication: Selecting input distributions for use in Monte Carlo simulations. Regulatory Toxicology and Pharmacology 21, 192-8.

Lloyd, A., Aitken, J. A., Hoffmeyer, U. K. O., Kelso, E. J., Wakerly. E. C., and Barber, N. D. (1997 Oct) Economic evaluation of the use of nadroparin calcium in the prophylaxis of deep vein thrombosis and pulmonary embolism in surgical patients in Italy. Pharmacoeconomics 12[4], 475-485.

Lober, J., Sogaard, J., Mouridsen, H. T., and Jorgensen, J. (1988) Treatment costs of adjuvant cytotoxic therapy in premenopausal breast cancer patients. Acta Oncologica 27[6A], 767-71.

Lokich, J. J., Moore, C. L., and Anderson, N. R. (1996 Jul 15) Comparison of costs for infusion versus bolus chemotherapy administration: analysis of five standard chemotherapy regimens in three common tumors--part one. Model projections for cost based on charges. Cancer 78[2], 294-9.

Loomes, G. (1988) Disparaties between health state measures: an explanation and some implications. Centre for Health Economics, University of York .

Lord, J. and Asante, M. A. (1999 Jun) Estimating uncertainty ranges for costs by the bootstrap procedure combined with probabilistic sensitivity analysis. Health Economics 8[4], 323-33.

Luce, B. R (1995) Policy implications of modeling the cost-effectiveness of health care technologies . Drug Information Journal 29, 1469-1475.

Macquart-Moulin, G., Viens, P., Genre, D., Bouscary, M. L., Resbeut, M., Gravis, G. et al. (1999 May 15) Concomitant chemoradiotherapy for patients with nonmetastatic breast carcinoma: side effects, quality of life, and organization. Cancer 85[10], 2190-9.

Mandelblatt JS, Fryback DG, Weinstein MC (1996) Assessing the effectiveness of health interventions. In Cost-Effectiveness in Health and Medicine, Gold MR, Siegel JE, Russell LB, Weinstein MC (eds) Oxford University Press:

Manning, W. G., Fryback, D. G., and Weinstein, M. C. (1996) Reflecting uncertainty in cost-effectiveness analysis, In: Gold, M. R., Siegel, J. E., and Russell. L. B. Cost-effectiveness in Health and Medicine. New York: Oxford University Press.

Marini, G., Murray, S., Goldhirsch, A., Gelber, R. D., Castiglione-Gertsch, M., Price, K. N. et al. (1996) The effect of adjuvant prednisone combined with cmf on patterns of relapse and occurrence of second malignancies in patients with breast cancer.

International (ludwig) breast cancer study group [see comments]. Annals of Oncology 7[3], 245-50.

Marshall and Oliver (1995) Decision Making and Forecasting.McGraw-Hill, Inc.

Martelli, G., DePalo, G., Rossi, N., Coradini, D., Boracchi, P., Galante, E. et al. (1995 Nov) Long-term follow-up of elderly patients with operable breast cancer treated with surgery without axillary dissection plus adjuvant tamoxifen. British Journal of Cancer 72[5], 1251-5.

McCabe, C. and Dixon, S. (2000) Testing the validity of cost-effectiveness models. Pharmacoeconomics 17[5], 501-13.

McHaney, R. (1991) Computer Simulation : A Practical Perspective. San Diego; London: Academic Press

McNeil, B. J. and Pauker, S. G. (1984) Decision analysis for public health: principles and illustrations. Annual Review of Public Health 5, 135-61.

Medical Research Council (1994) Developing high quality proposals in health services research.

Messori, A., Becagli, P., Trippoli, S., and Tendi, E. (1996) Cost-effectiveness of adjuvant chemotherapy with cyclophosphamide+methotrexate+fluorouracil in patients with node-positive breast cancer. European Journal of Clinical Pharmacology 51[2], 111-6.

Michel, B. C, van Staveren, R. J. E, Geven, W. B, and van Hout, B. A (1996) Simulation models in the planning of health care facilities: an application in the case of neonatal extracorporeal membrane oxygenation. Journal of Health Services Research & Policy 1[4], 198-204.

Midgette, A. S., Wong, J. B., Beshansky, J. R., Porath, A., Fleming, C., and Pauker, S. G (1994) Cost-effectiveness of streptokinase for acute myocardial infarction: A

combined meta-analysis and decision analysis of the effects of infarct location and of likelihood of infarction. Medical Decision Making 14[2], 108-167.

Miller, D. K. and Homan, S. M. (1994) Determining transition probabilities: Confusion and suggestions. Medical Decision Making 14[1], 52-8.

Misset, J. L., di Palma, M., Delgado, M., Plagne, R., Chollet, P., Fumoleau, P. et al. (1996 Apr) Adjuvant treatment of node-positive breast cancer with cyclophosphamide, doxorubicin, fluorouracil, and vincristine versus cyclophosphamide, methotrexate, and fluorouracil: final report after a 16-year median follow-up duration. Journal of Clinical Oncology 14[4], 1136-45.

Murphy, J. F., Jain, N. l., Romero, C. A., and Kahn, M. G. (1994) Construction of a literature database and its use to provide probabilities for decision-analytic models of thrombolytic therapy. The Annual Symposium on Computer Applications in Medical Care , 819-23.

Mustacchi, G., Milani, S., Pluchinotta, A., De Matteis, A., Rubagotti, A., and Perrota, A. (1994 Sep-Oct) Tamoxifen or surgery plus tamoxifen as primary treatment for elderly patients with operable breast cancer: the g.r.e.t.a. Trial. Group for research on endocrine therapy in the elderly. Anticancer Research 14[5B], 2197-200.

National Cancer Institute (1996) Questions and answers about adjuvant therapy for breast cancer. gopher.ncc.go.jp:/70/CNET/All_files/600720 .

NHS Centre for Reviews and Dissemination (1996) Undertaking Systematic Reviews of Research on Effectiveness. University of York:

NHS Economic Evaluation Database University of York: http://agatha.york.ac.uk/nhsdhp.htm

NHS Executive (1997) The annual report of the NHS Health Technology Assessment programme 1997: Identifying questions, finding answers. National Co-ordinating Centre for Health Technology Assessment .

Nuijten MJC, Hardens M, Souetre E (1995) A Markov process analysis comparing the cost effectiveness of maintenance therapy with citalopram versus standard therapy in major depression. Pharmacoeconomics 8: 159-168

Office of National Statistics (2000) Deaths: selected causes (International Classification) and sex, 1971 Onwards (England and Wales). Health Statistics Quarterly 06 .

OHE (July 2000) Health economic evaluations database. OHE-IFPMA Database Ltd, 12 Whitehall, London .

Paltiel, A. D., Scharfstein, J. A., Seage, G. R. 3rd, Losina, E., Goldie, S. J., Weinstein, M. C. et al. (1998 Apr-Jun) A monte carlo simulation of advanced hiv disease: application to prevention of cmv infection. Medical Decision Making 18[2 Suppl], S93-105.

Paridaens, R., Heuson, J. C., Julien, J. P., Veyret, C., Van Zyl, J., Klijn, J. G. et al. (1993 Sep) Assessment of estrogenic recruitment before chemotherapy in advanced breast cancer: a double-blind randomized study. European organization for research and treatment of cancer breast cancer cooperative group. Journal of Clinical Oncology 11[9], 1723-8.

Parmigiani, G., Ancukiewicz, M., and Matchar, D. (1996) Decision models in clinical recommendations development: The stroke prevention policy model, In: Berry, D. A. and Stangl, D. K. Bayesian Biostatistics. New York: Marcel Dekker, Inc.

Pasta, D. J., Taylor, J. L., and Henning, J. M. (1999 Jul-Sep) Probabilistic sensitivity analysis incorporating the bootstrap: an example comparing treatments for the eradication of helicobacter pylori. Medical Decision Making 19[3], 353-63.

Paul, R. and Balmer, D. (1993) Simulation Modelling. Bromley: Chartwell-Bratt

Pavesi, L., Preti, P., Da Prada, G., Pedrazzoli, P., Poggi, G., and Robustelli della= Cuna G (1995 Mar-Apr) Epirubicin versus mitoxantrone in combination chemotherapy for metastatic breast cancer. Anticancer Research 15[2], 495-501.

Perez, D. J., McGee, R., Campbell, A. V., Christensen, E. A., and Williams, S. (1997 Mar) A comparison of time trade-off and quality of life measures in patients with advanced cancer. Quality of Life Research 6[2], 133-8.

Peto, R. (1987) Why do we need systematic overviews of randomised trials? Statistics in Medicine 6, 233-40.

Pidd, M. (1989) Computer Modelling for Discrete Simulation. Chichester: Wiley

Pidd, M. (1998) Computer Simulation in Management Science, 4th ed. Chichester: Wiley

Pocock, S. J. (1983) Clinical Trials. A Practical Approach. Chichester: John Wiley and sons

Polsky, D., Glick, H. A., Willke, R., and Schulman, K. (1997) Confidence intervals for cost-effectiveness ratios: a comparison of four methods. Health Economics 6[3], 243-52.

Popper, K. R. (1965) Conjectures and Refutations. London: Routledge

Powles, T. J., Hickish, T. F., Makris, A., Ashley, S. E., O'Brien, M. E., Tidy, V. A. et al. (1995 Mar) Randomized trial of chemoendocrine therapy started before or after surgery for treatment of primary breast cancer. Journal of Clinical Oncology 13[3], 547-52.

Price, A., O'Brien, M., Judson, I., and Klastersky, J. (1995) Handbook of adult cancer chemotherapy schedules. The Royal Marsden NHS Trust .

Pritchard, K. I., Paterson, A. H., Fine, S., Paul, N. A., Zee, B., Shepherd, L. E. et al. (1997 Jun) Randomized trial of cyclophosphamide, methotrexate, and fluorouracil

chemotherapy added to tamoxifen as adjuvant therapy in postmenopausal women with node-positive estrogen and/or progesterone receptor-positive breast cancer: a report of the national cancer institute of canada clinical trials group. Breast cancer site group. Journal of Clinical Oncology 15[6], 2302-11.

Pritchard, K. I., Paterson, A. H., Paul, N. A., Zee, B., Fine, S., and Pater, J. (1996 Oct) Increased thromboembolic complications with concurrent tamoxifen and chemotherapy in a randomized trial of adjuvant therapy for women with breast cancer. National cancer institute of canada clinical trials group breast cancer site group. Journal of Clinical Oncology 14[10], 2731-7.

Procite 5 (1999), www.procite.com

Pyrhonen, S., Valavaara, R., Modig, H., Pawlicki, M., Pienkowski, T., Gundersen, S. et al. (1997) Comparison of toremifene and tamoxifen in post-menopausal patients with advanced breast cancer: a randomized double-blind, the 'nordic' phase iii study. British Journal of Cancer 76[2], 270-7.

Quirk, D. M., Barry, M. J., Aserkoff, B., and Podolsky, D. K. (1997 Nov) Physician specialty and variations in the cost of treating patients with acute upper gastrointestinal bleeding [see comments]. Gastroenterology 113[5], 1443-8.

Rasmusson, B., Vejborg, I., Jensen, A. B., Andersson, M., Banning, A. M., Hoffmann, T. et al. (1995 Mar) Irradiation of bone metastases in breast cancer patients: a randomized study with 1 year follow-up. Radiotherapy & Oncology 34[3], 179-84.

Raudenbush, S. W. (1994) Random effects models, In: Cooper, H. and Hedges, L. V. The Handbook of Research Synthesis. New York: Russell Sage Foundation; 301-22.

Ravaioli, A., Bagli, L., Zucchini, A., and Monti, F. (1998 Jun-Aug) Prognosis and prediction of response in breast cancer: the current role of the main biological markers. Cell Proliferation 31[3-4], 113-26.

Read, G. (1994) Estimating the cost of radiotherapy. Clinical Oncology (Royal College of Radiologists) 6[1], 35-9.

Ribeiro, G. and Swindell, R. (1992) The christie hospital adjuvant tamoxifen trial. Journal of the National Cancer Institute Monographs. [11], 121-5.

Rice, J. A. (1995) Mathematical Statistics and Data Analysis, 2nd ed. Belmont, California: Duxberry Press

Richards MA, Braysher S, Gregory WM, and Rubens RD (1993) Advanced breast cancer: use of resources and cost implications. British Journal of Cancer 67, 856-860.

Richards, M. A., Hopwood, P., Ramirez, A. J., Twelves, C. J., Ferguson, J., Gregory, W. M. et al. (1992) Doxorubicin in advanced breast cancer: influence of schedule on response, survival and quality of life. European Journal of Cancer 28A[6-7], 1023-8.

Rivkin, S. E., Green, S., Metch, B., Cruz, A. B., Abeloff, M. D., Jewell, W. R. et al. (1994 Oct) Adjuvant cmfvp versus tamoxifen versus concurrent cmfvp and tamoxifen for postmenopausal, node-positive, and estrogen receptor-positive breast cancer patients: a southwest oncology group study. Journal of Clinical Oncology 12[10], 2078-85.

Rivkin, S. E., Green, S., O'Sullivan, J., Cruz, A. B., Abeloff, M. D., Jewell, W. R. et al. (1996 Jan) Adjuvant cmfvp versus adjuvant cmfvp plus ovariectomy for premenopausal, node-positive, and estrogen receptor-positive breast cancer patients: a southwest oncology group study. Journal of Clinical Oncology 14[1], 46-51.

Robertson, J. F., Whynes, D. K., Dixon, A., and Blamey, R. W. (1995 Jul) Potential for cost economies in guiding therapy in patients with metastatic breast cancer. British Journal of Cancer 72[1], 174-7.

Rosenthal, M. C. (1994) The fugitive literature. In: Cooper, H. and Hedges, L. V. The Handbook of Research Synthesis. New York: Russell Sage Foundation; 85-96.

Schumacher, M., Bastert, G., Bojar, H., Hubner, K., Olschewski, M., Sauerbrei, W. et al. (1994 Oct) Randomized 2 x 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. German breast cancer study group. Journal of Clinical Oncology 12[10], 2086-93.

Schwaibold, F, Fowble, B. L, Solin, L. J, Schultz, D. J, and Goodman, R. L (1991) The results of radiation therapy for isolated local regional recurrence after mastectomy. International Journal of Radiation, Oncology, Biology, and Physics. 21, 299-310.

Scottish cancer trials breast group (1993 May 22) Adjuvant ovarian ablation versus cmf chemotherapy in premenopausal women with pathological stage ii breast carcinoma: the Scottish trial. The Lancet 341[8856], 1293-8.

Sculpher, M., Fenwick, E., and Claxton, K. (2000) Assessing quality in decision analytic cost-effectiveness models: A suggested framework and example of application. Pharmacoeconomics 17[5], 461-77.

Selke, B., Lenne, X., Bercez, C., Bonneterre, M. E., Lebrun, T., and Bonneterre, J. (1998) Cost-Effectiveness Analysis of Adjuvant Treatments in the Management of Breast-Cancer. Medical Decision Making 18[4], 488-488.

Senn, H. J., Maibach, R., Castiglione, M., Jungi, W. F., Cavalli, F., Leyvraz, S. et al. (1997 Jul) Adjuvant chemotherapy in operable breast cancer: cyclophosphamide, methotrexate, and fluorouracil versus chlorambucil, methotrexate, and fluorouracil--11-year results of swiss group for clinical cancer research trial sakk 27/82. Journal of Clinical Oncology 15[7], 2502-9.

Sharples, L. D., Briggs, A., Caine, N., McKenna, M., and Buxton, M. (1996 Sep 15) A model for analyzing the cost of main clinical events after cardiac transplantation. Transplantation 62[5], 615-21.

Sharples, L. D., Taylor, G. J., Karnon, J., Caine, N., Buxton, M., McNeil, K. et al. (forthcoming) A model for analysing the cost of main clinical events after lung transplantation. Journal of Heart and Lung Transplantation .

Sheldon, T. A. (1996 Jan-Feb) Problems of using modelling in the economic evaluation of health care. Health Economics 5[1], 1-11.

Silva, O. E. and Zurrida, S. (1999) Breast Cancer: A Guide for Fellows. Amsterdam: Elsevier

Simul8, version6 (2000) Glasgow, Scotland: Visual Thinking International Ltd: http://www.SIMUL8.com/

Sisk, J. E., Moskowitz, A. J., Whang, W., Lin, J. D., Fedson, D. S., McBean, A. M. et al. (1997 Oct 22-29) Cost-effectiveness of vaccination against pneumococcal bacteremia among elderly people. Journal of the American Medical Association 278[16], 1333-9.

Smith, T. J. and Hillner, B. E. (1993 Apr) The efficacy and cost-effectiveness of adjuvant therapy of early breast cancer in premenopausal women. Journal of Clinical Oncology 11[4], 771-6.

Sonnenberg FA, Roberts MS, Tsevat J, Wong JB, Barry M, Kent DL (1994) Toward a peer review process for medical decision analysis models. Medical Care 32: JS52-64

Sonnenberg, F. A. and Beck, J. R. (1993 Oct-Dec) Markov models in medical decision making: a practical guide. Medical Decision Making 13[4], 322-38.

Sonnenberg, F. A. and Wong, J. B. (1993 Apr-Jun) Commentary: fine-tuning life-expectancy calculations using Markov processes. Medical Decision Making 13[2], 170-2.

Sonnenberg, F. A., Roberts, M. S., Tsevat, J., Wong, J. B., Barry, M., and Kent, D. L. (1994 Jul) Toward a peer review process for medical decision analysis models. Medical Care 32[7 Suppl], JS52-64.

Spector, T. D., and Thompson, S. G. (1991) Research methods in epidemiology. 5. The potential and limitations of meta-analysis. Journal of Epidemiology and Community Health 45, 89-92.

Stat::Fit (1996) USA: Geer Mountain Software Corporation;

Stewart, D. J., Evans, W. K., Shepherd, F. A., Wilson, K. S., Pritchard, K. I., Trudeau, M. E. et al. (1997 May) Cyclophosphamide and fluorouracil combined with mitoxantrone versus doxorubicin for breast cancer: superiority of doxorubicin. Journal of Clinical Oncology 15[5], 1897-905.

Stewart, H. J. (1992) The scottish trial of adjuvant tamoxifen in node-negative breast cancer. Scottish cancer trials breast group. Journal of the National Cancer Institute Monographs. [11], 117-20.

Stinnett, A. A. and Mullahy, J. (1998 Apr-Jun) Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical Decision Making 18[2 Suppl], S68-80.

Stinnett, A. A. and Paltiel, A. D. (1997 Oct-Dec) Estimating CE ratios under second-order uncertainty: the mean ratio versus the ratio of means. Medical Decision Making 17[4], 483-9.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., and Song, F. (1998) Systematic reviews of trials and other studies. Health Technology Assessment 2[19], 1-276.

Swain, S. M., Whaley, F. S., Gerber, M. C., Weisberg, S., York, M., Spicer, D. et al. (1997 Apr) Cardioprotection with dexrazoxane for doxorubicin-containing therapy in advanced breast cancer. Journal of Clinical Oncology 15[4], 1318-32.

Swedish breast cancer cooperative group (1996 Nov 6) Randomized trial of two versus five years of adjuvant tamoxifen for postmenopausal early stage breast cancer. Journal of the National Cancer Institute 88[21], 1543-9.

Szeto, L. L. and Devlin, N. J. (1996) The cost-effectiveness of mammography screening: Evidence from a microsimulation model for New Zealand. Health Policy 38, 101-15.

Tambour, M., Zethraeus, N., and Johannesson, M. (1998 Summer) A note on confidence intervals in cost-effectiveness analysis. International Journal of Technology Assessment in Health Care 14[3], 467-71.

The givio investigators (1994 May 25) Impact of follow-up testing on survival and health-related quality of life in breast cancer patients. A multicenter randomized controlled trial. Journal of the American Medical Association 271[20], 1587-92.

The new NHS - 1998 Reference Costs (1998) London: http://www.doh.gov.uk/nhsexec/refcosts/1998.htm

Thompson, S. G. and Pocock, S. J. (1991) Can meta-analyses be trusted? The Lancet 338, 1127-30.

Thorpe, S. M., Christensen, I. B., Rasmussen, B. B., and Rose, C. (1993) Short recurrence-free survival associated with high oestrogen receptor levels in the natural history of postmenopausal, primary breast cancer. European Journal of Cancer 29A[7], 971-7.

Thurlimann, B., Beretta, K., Bacchi, M., Castiglione-Gertsch, M., Goldhirsch, A., Jungi, W. F. et al. (1996 Jul) First-line fadrozole hci (cgs 16949a) versus tamoxifen in postmenopausal women with advanced breast cancer. Prospective randomised trial of the swiss group for clinical cancer research sakk 20/88 [see comments]. Annals of Oncology 7[5], 471-9.

Toi, M., Tanaka, S., Bando, M., Hayashi, K., and Tominaga, T. (1997 Jan) Outcome of surgical resection for chest wall recurrence in breast cancer patients. Journal of Surgical Oncology 64[1], 23-6.

Toonkel, L. M, Ivor, F, Jacobson, L. H, and Wallach, C. B (1983) The significance of local recurrence of carcinoma of the breast. International Journal of Radiation, Oncology, Biology, and Physics. 9[33-9].

Tormey, D. C., Gray, R., Abeloff, M. D., Roseman, D. L., Gilchrist, K. W., Barylak, E. J. et al. (1992 Dec) Adjuvant therapy with a doxorubicin regimen and long-term tamoxifen in premenopausal breast cancer patients: an eastern cooperative oncology group trial. Journal of Clinical Oncology 10[12], 1848-56.

Townsend, J. and Buxton, M. (1997 Mar) Cost effectiveness scenario analysis for a proposed trial of hormone replacement therapy. Health Policy 39[3], 181-94.

Tross, S., Herndon, J. 2nd, Korzun, A., Kornblith, A. B., Cella, D. F., Holland, J. F. et al. (1996 May 15) Psychological symptoms and disease-free and overall survival in women with stage ii breast cancer. Cancer and leukemia group B. Journal of the National Cancer Institute 88[10], 661-7.

UKCCR (1993) The Adjuvant Breast Cancer (ABC) Trial Protocol.

Urban, N., Drescher, C., Etzioni, R., and Colby, C (1997) Use of stochastic simulation model to identify an efficient protocol for ovarian cancer screening. Controlled Clinical Trials 18, 251-270.

van Hout, B. A., Al, M. J., Gordon, G. S., and Rutten, F. F. (1994 Sep-Oct) Costs, effects and c/e-ratios alongside a clinical trial. Health Economics 3[5], 309-19.

Velez-Garcia, E., Carpenter, J. T. Jr, Moore, M., Vogel, C. L., Marcial, V., Ketcham, A. et al. (1992) Postsurgical adjuvant chemotherapy with or without radiotherapy in women with breast cancer and positive axillary nodes: a south-eastern cancer study group (seg) trial. European Journal of Cancer 28A[11], 1833-7.

Venturini, M., Bruzzi, P., Del Mastro, L., Garrone, O., Bertelli, G., Guelfi, M. et al. (1996 Mar) Effect of adjuvant chemotherapy with or without anthracyclines on the

activity and efficacy of first-line cyclophosphamide, epidoxorubicin, and fluorouracil in patients with metastatic breast cancer. Journal of Clinical Oncology 14[3], 764-73.

Vogel, C. L., Azevedo, S., Hilsenbeck, S., East, D. R., and Ayub, J. (1992 Jul 1) Survival after first recurrence of breast cancer. The Miami experience. Cancer 70[1], 129-35.

Wakker, P. and Klaassen, M. (1995) Confidence intervals for cost-effectiveness ratios. Health Economics 4[5], 373-82.

Warde, P. and Murphy, T. (1996) Measuring the cost of palliative radiotherapy. Canadian Journal of Oncology. 6[suppl 1], 90-4.

Warner, K. E, Smith, R. J, Smith, D. G, and Fries, B. E (1996) Health and economic implications of a work-site smoking-cessation program: a simulation analysis . Journal of Occupational and Environmental Medicine 38[10], 981-992.

Weinstein, M. C. and Fineberg, H. V. (1980) Clinical Decision Analysis. Philadelphia ; London: Saunders

Will, B. P., Berthelot, J. M., Houle, C., Tomiak, E., Verma, S., and Evans, W. K. (1998) The economic impact of locoregional radiotherapy (LRRT) on all post-surgical stage 2 breast cancer patients in Canada. Statistics Canada/ Ottowa Regional Cancer Centre & Statistics Canada Ottowa, Ontario, Canada .

Willner, J., Kiricuta, I. C., and Kolbl, O. (1997 Mar 1) Locoregional recurrence of breast cancer following mastectomy: always a fatal event? Results of univariate and multivariate analysis. International Journal of Radiation Oncology, Biology, and Physics 37[4], 853-63.

Wolstenholme J and Whynes D (1998) The costs of treating breast cancer: implications for screening. International Journal of Technology Assessment in Health Care 14[2].

Wong, K. and Henderson, I. C. (1994 Jan-Feb) Management of metastatic breast cancer. World Journal of Surgery 18[1], 98-111.

Wood, W. C., Budman, D. R., Korzun, A. H., Cooper, M. R., Younger, J., Hart, R. D. et al. (1994 May 5) Dose and dose intensity of adjuvant chemotherapy for stage ii, node-positive breast carcinoma. New England Journal of Medicine 330[18], 1253-9.

Zambetti, M., Bonadonna, G., Valagussa, P., Daidone, M. G., Coradini, D., Bignami, P. et al. (1992) Adjuvant cmf for node-negative and estrogen receptor-negative breast cancer patients. Journal of the National Cancer Institute Monographs. [11], 77-83.