ECONOMIC EVALUATION OF MINIMAL ACCESS SURGERY: THE CASE OF SURGICAL TREATMENT FOR MENORRHAGIA

A thesis submitted for the degree of Doctor of Philosophy

by

Mark John Sculpher

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

October 1996

List of Tables

List of Figures

List of Boxes

List of Appendices

Abbreviations

AH	Abdominal hysterectomy
CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
CUA	Cost-utility analysis
HRQL	Health-related quality of life
HYE	Healthy-years equivalent
LAVH	Laparoscopic-assisted vaginal hysterectomy
MAS	Minimal access surgery
MISTLETOE	Minimally invasive surgical techniques: laser, endothermal or resection
NHS	National Health Service
QALY	Quality-adjusted life year
RA QALY	Risk-adjusted quality-adjusted life year
RCT	Randomised controlled trial
RF	Radiofrequency
SG	Standard gamble
TCRE	Transcervical resection of the endometrium
TTO	Time trade-off
VH	Vaginal hysterectomy
WTP	Willingness to pay

Acknowledgements

Much of the research reported in this thesis was undertaken as part of a project grant awarded by the Department of Health.

I would like to acknowledge several colleagues for their assistance with this project: Dr Nuala Dwyer, Mr John Cullimore, Mr James Browning and Professor Gordon Stirrat for their willingness to allow their patients to be involved in the research; Ms Sarah Byford, Mrs Moira O'Hanlon and Mrs Sue Horsley for interviewing women involved in the health state valuation components of the project; Mrs Betty Samuel and Mrs Maggie Pragnall for their administrative support in the centres; and Mr Michael Maresh and Mr Chris Overton for giving me access to MISTLETOE survey data.

I am particularly grateful to Professor Martin Buxton, my supervisor, and to Professor Michael Drummond, the external consultant to my PhD, for their help and advice throughout the development of the thesis.

I would also like to thank my colleagues in the Health Economics Research Group for their comments on initial plans and subsequent drafts, and members of the Health Economists' Study Group who have provided valuable comments on earlier versions of chapters. For their tireless word processing services, I would like to thank Karen Arnold, Nicky Gillard and Jo Holland.

Finally, I would like to thank my wife, Clare, for her patience and encouragement.

Abstract

The purpose of this thesis is to explore the methodological and empirical issues relating to the economic evaluation of minimal access surgery (MAS). Given the likely increase in the utilisation of economic analysis in this area, it is crucial to explore whether the methods of economic evaluation have limitations in the face of the particular characteristics of MAS. The comparison of abdominal hysterectomy (AH) and transcervical endometrial resection (TCRE), for the treatment of menorrhagia, is used as a vehicle to develop methods in relation to MAS. Having reviewed the literature and issues relating to the economic analysis of this group of technologies, the empirical starting point of the thesis is the assessment of the limitations of economic evaluations alongside clinical trials, using a trial comparing AH and TCRE. Three major areas of weakness are identified, and alternative ways of addressing these weaknesses are explored in the remainder of the thesis. The first area of methodological development relates to the measure of benefit used in economic evaluation of MAS. In this clinical context, it is argued that the trade-offs that exist between MAS and conventional surgery, in terms of process characteristics and outcomes, should result in an important role for patients' preferences in the construction of a benefit measure. A cost-utility analysis using the standard quality-adjusted life year (QALY) is undertaken, using trial data augmented with valuation data from a further study. The lack of consistency between individual preferences and standard QALYs suggests a major weakness with this measure of benefit. The strengths and weaknesses of an alternative measure of benefit in cost-utility analysis - the *ex ante* healthy years equivalent (HYE) - are assessed based on a further valuation study. It is shown that it is feasible to elicit *ex ante* HYEs from patients and that this measure of benefit exhibits some consistency with other expressions of patients' preferences. However, the HYE is likely to impose a greater measurement burden than the standard QALY. The second area of methodological development in the thesis is the analysis of the generalisability of trial-based economic evaluation, given the limitations that often exist with the external validity of trials. A framework is developed within which trial and observational data can be synthesised. This facilitates the use of sensitivity analysis to explore the robustness of base-case (trial-generated) results to alternative sources of data, which may be more representative of routine practice. The third area of methodological development stems from the importance of patients' preferences in relation to MAS. This element relates to the use of methods to model and to evaluate management strategies which use patients' preferences to determine treatment allocation. It is concluded that preference-based decision making has the potential to be cost-effective in relation to TCRE and AH, and MAS applications more generally. The thesis demonstrates the importance of continued development in the detailed methods of economic evaluation.

Contents

Publications and Authorship

The research presented in Chapter 3 of this thesis has been published in two peer-reviewed journal articles [Sculpher *et al*, 1993; Sculpher *et al*, 1996A].

The research presented in Chapters 3, 4, 5, and 6 is the result of collaborative effort on the part of a multi-disciplinary research ream. Whilst acknowledging the contribution of my colleagues, it should be emphasised that for all the research presented in this thesis I have been the lead researcher generating the initial ideas, designing the main component studies and undertaking all analyses.

This thesis is dedicated to the memory of my father

Chapter 1

Introduction

1.1 The role of economic evaluation in health care

This thesis considers the methodological issues relating to the economic evaluation of minimal access surgery (MAS). It is now widely recognised that, in order to ensure that the health service maximises patient benefits from the limited resources available for health care, a systematic and formal assessment of the costs and consequences of health care interventions is necessary. The tools of economic evaluation have, therefore, become widely used to inform resource allocation in health care systems across the world. There is now an extensive literature on the basic methods of economic evaluation in health care [Drummond *et al*, 1987; Detsky and Naglie, 1990; Luce and Elixhauser, 1990; Petitti, 1994] and a growing literature relating to applied economic evaluation of health care interventions [Backhouse *et al*, 1992].

1

The increased demand for economic evaluation is clear in the UK. The largest component of the National Health Service (NHS) Research and Development strategy is a programme of health technology assessment where economic evaluation is a prominent research priority [Department of Health, 1993A and 1995]. The Health Services Research Board of the Medical Research Council now presumes that economic evaluation will normally be required as part of the clinical trials it funds. In addition to publicly funded economic evaluation, the pharmaceutical industry has increased its funding of economic evaluation of its products, both for regulatory purposes in some countries and for commercial reasons [Evans, 1995].

Despite the development of guidelines for economic evaluation in health care in several countries and for various purposes [Henry, 1992; Association of the British Pharmaceutical Industry, 1994; Ministry of Health, 1994; Canadian Coordinating Office for Health Technology Assessment, 1994; Drummond and Jefferson, 1996; Panel on Cost-Effectiveness in Health and Medicine, 1996], uncertainty still remains regarding the most appropriate detailed methods for economic analysis, necessitating further methodological development in this area [Drummond *et al*, 1993A].

1.2   Economic evaluation of MAS

The current limitations of the tools of economic evaluation are highlighted when they are applied to health care technologies with particular characteristics. One group of technologies which poses some particular problems for economic analysis is minimal access surgery. The need for a detailed assessment of the methods of economic evaluation in relation to MAS is emphasised by the growing importance of MAS procedures within the health service. In 1993, a committee formed to advise the UK government predicted that, by the year 2000, 70% to 80% of surgical practice will be based on MAS techniques [Cuschieri, 1993]. Given the large amount of health service resources that are, therefore, likely to be devoted to MAS during the next few years, formal

assessment of their costs and consequences relative to conventional treatment is essential.

Significant research resources are being devoted to the *clinical* evaluation of MAS procedures in the UK; for example, the MRC has recently funded a large multi-centre trial of laparoscopic surgery for colorectal cancer [MRC, personal communication]; and the NHS Executive Research and Development Programme has recently agreed to fund a multi-centre trial involving 1800 women to evaluate laparoscopic hysterectomy [Mr Ray Garry, personal communication].

Expenditure on the evaluation of MAS procedures may increase further in the UK if recent calls for the mandatory evaluation of surgical technologies prior to routine use are heeded [Advisory Council on Science and Technology (ACOST), 1993]. Although the Department of Health has so far rejected these calls [Department of Health, 1993B], it has agreed to fund a voluntary system for registering surgical procedures the effectiveness and safety of which have yet to be established, with procedures on this register feeding into the priority setting process for the NHS Health Technology Assessment Programme [Sheldon and Faulkner, 1996]. Although the register does not focus on the need for economic assessment, the likely increase in funding for clinical evaluation of surgical procedures will stimulate a demand for formal economic analysis, given that economic evaluation is a key part of the NHS Research and Development Programme. However, prior to increased funding of applied economic evaluations of MAS procedures, it is necessary to develop further the tools of analysis based on a careful consideration of the characteristics of this group of technologies.

1.3 Contribution of the thesis

The methodological problems related to the economic evaluation of MAS, and alternative approaches to overcoming them, are the focus of this thesis.

In part, the thesis reviews and applies some recent methodological developments, in particular relating to economic analysis alongside clinical trials, the role of the standard QALY in cost-utility analysis and the systematic presentation of uncertainty in economic evaluation. Although this does not represent new methodological work, these methods are being applied in detail to the area of MAS for the first time, providing new insight into their strengths and limitations in an important area of applied economic evaluation.

Much of the thesis, however, is devoted to the development of novel evaluative methods. MAS procedures tend to develop quickly, often with new versions of an application in a given clinical area diffusing prior to full evaluation of earlier versions. Related to this is the variation that often exists in how applications of MAS are actually used in routine clinical practice. These characteristics of MAS limit the external validity of many economic evaluations in this area, particularly those undertaken as part of, or alongside, a clinical trial. An important contribution of this thesis is the development of a framework to analyse the generalisability of an economic evaluation. Starting out with a core model based on trial data to provide high levels of internal validity, the framework involves the incorporation of experimental and observational data from a range of alternative sources within a series of sensitivity analyses. The aim is to assess the robustness of the base-case results to these alternative parameter estimates.

A second contribution relates to the multi-dimensionality of outcomes often associated with MAS and the limitations this imposes on cost-effectiveness analysis and cost-utility analysis based on the standard QALY. These problems stimulate a consideration of alternative benefit measures for economic analysis; in particular the role of healthy-years equivalents (HYEs) based on the time trade-off (TTO) valuation instrument. The thesis considers the differences in results of cost-utility analysis based on the standard QALY compared to those using the TTO-based HYE; the consistency of this measure of benefit with more descriptive measures of preference; and the practical issues related to the use of HYEs.

A third area in which this thesis contributes to the methods of economic evaluation relates to the role of patients' preferences in determining treatment allocation. The standard approach to economic evaluation in health care is to compare two or more interventions relating to a particular group of patients and to identify the (single) economically superior option, with the presumption that this will be provided to each patient in that group. If heterogeneity exists in the clinical characteristics of individuals, and these differences affect the extent of patient benefit, then it may represent good value for money to provide an intervention to a sub-group of patients which is not cost-effective for the whole group. Benefits will also depend on patients' preferences for particular treatment processes and prognoses and health states, and there seems to be a marked variation in patients' preferences relating to the range of different consequences of MAS procedures. There would seem to be a case, therefore, for allowing patients' preferences to play a formal part in determining treatment allocation, but this itself requires full economic evaluation and this approach to patient management has rarely been factored into the methods of economic evaluation. The thesis explores three alternative models of preference-based treatment allocation, considering for each the design of an economic analysis, data requirements and practicalities.

1.4   A case-study: surgical treatment for menorrhagia

As a way of highlighting the importance of these various areas of method, and as a vehicle for exploring the feasibility and implications of alternative approaches to the economic evaluation of MAS, the surgical treatment of menorrhagia is used as a case-study throughout the thesis. It is not the purpose of this section to provide a systematic and comprehensive review of existing evidence on the effectiveness and costs of alternative approaches to the management of menorrhagia. Rather, the purpose is briefly to summarise the key issues in the literature that impact on the economics of alternative treatments for menorrhagia, as well as to highlight the important areas of uncertainty relating to the use of surgical technologies in this clinical area.

1.4.1 The burden of menorrhagia

Menorrhagia, or 'excessive' menstrual bleeding, represents a major health care burden in the UK. This fact is illustrated by a range of statistics.

- The condition affects approximately 22% of otherwise healthy women [Gath et al, 1987].

- Some 822,000 prescriptions are issued for menorrhagia each year, costing £7 million [Effective Health Care, 1995].

- Menstrual problems account for some 12% of gynaecology referrals [Bradlow et al, 1992].

- Approximately 10,000 MAS procedures are undertaken each year for menorrhagia [RCOG Audit Unit, personal communication].

- The chance of a woman having undergone a hysterectomy by the age of 55 for menstrual problems is between 7% and 13% [Grant and Hussein, 1984; Vessey et al, 1992].

- Rates of surgical treatment for menorrhagia appear to be increasing in the UK [Vessey et al, 1992; Coulter et al, 1993; Bridgman, 1994; Coulter 1994].

Although the majority of women with menorrhagia exhibit no abnormal pathology (in these circumstances this condition is sometimes referred to as dysfunctional uterine bleeding), the condition clearly has a detrimental impact on women's health-related quality of life [Garratt et al, 1993; Jenkinson et al, 1994]. Furthermore, the burden in terms of the health care resources devoted to the care of women with menorrhagia is significant. Figure 1.1 breaks down the resource impact of menorrhagia in terms of the cost of visits to the GP, drugs, conservative surgery and hysterectomy. It can be seen that surgery represents 77% of the cost of menorrhagia.

1.4.2 Treatment options

The first-line therapy for most women with menorrhagia is medical treatment. However, there has been a widespread opinion that such drug treatment for this
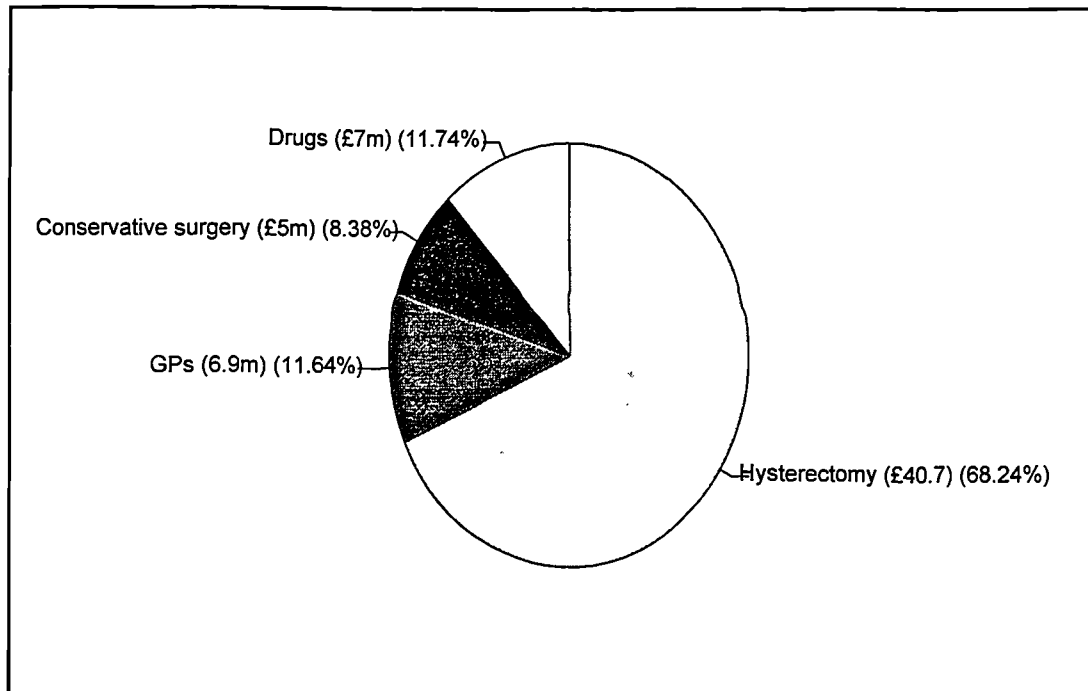
Figure 1.1   Resource costs of menorrhagia to the health service in England and Wales (1994 prices). Costs have been estimated as follows. GP costs are based on 868,000 consultations per year (31 per 1000 women [Coulter et al, 1991] applied to a female population of 28 million) at a unit cost of £8 per visit [Netten, 1994]. Drug costs are taken from Effective Health Care [1995]. The cost of hysterectomy assumes that 74,000 are undertaken each year [Effective Health Care, 1995], with 50% for menorrhagia, at a unit cost of £1,100 (see Chapter 3). The cost of conservative surgery is based on 10,000 procedures [RCOG Audit Unit, personal communication], with a unit cost of £500 (see Chapter 3).

condition is ineffective [Consumers' Association Ltd, 1990], with recent research demonstrating a mismatch between the prescribing policies of GPs and the most effective drug treatments [Coulter et al, 1995B].

The perceived ineffectiveness of medical therapies has resulted in a large proportion of women being referred to hospital for possible surgical intervention. Traditionally, the only widespread surgical intervention was hysterectomy: a survey undertaken in 1988-89 found that 60% of women underwent hysterectomy within five years of being referred to hospital with menorrhagia [Coulter et al, 1991]. In the UK, the abdominal form of the operation has traditionally represented about 90% of hysterectomies [Vessey et al, 1992].

Recently, there has been a major change in surgical treatment in this area. In particular, the advent of therapeutic endoscopy in a range of specialties has stimulated hysteroscopic therapies for menorrhagia. Typically, these treatments involve the endometrium being resected using electro-diathermy with either a loop or rollerball device (transcervical resection of the endometrium (TCRE)) [Magos *et al*, 1989], or ablated using a Nd:YAG laser [Goldrath, 1981]. In principle, the attraction of these various forms of MAS alternatives to hysterectomy, from a clinical point of view, is their association with a shorter stay in hospital and a shorter period of convalescence for the patient. The former also explains the expectation of lower health service costs with these treatments, in comparison with hysterectomy.

The endoscopic revolution has also had important implications for hysterectomy. Some centres are now using laparoscopic hysterectomy or laparoscopic-assisted vaginal hysterectomy [Hunter and McCarthey, 1993]. In essence, these developments represent an attempt to retain the conventional therapeutic effects of hysterectomy (ie. amenorrhoea), whilst reducing the severity and length of convalescence, and duration of hospital stay.

To date, three randomised controlled trials (RCTs) have been published comparing non-hysterectomy forms of MAS (TCRE or laser ablation) with abdominal hysterectomy (AH), based in Reading [Gannon *et al*, 1991]; Bristol [Dwyer *et al*, 1993] and Aberdeen [Pinion *et al*, 1994]. Overall, they show a clear trade-off between the effects of the two forms of surgery: TCRE and laser ablation result in fewer complications, a shorter convalescence and less peri-operative pain, but these treatments often fail to ameliorate women's symptoms adequately; whereas AH provides a once-and-for-all solution to heavy menstrual bleeding. The cohort of women in the Bristol trial forms the basis of much of the detailed analysis described in this report. On the basis of the results published so far, however, the *long-term* effects of TCRE and laser ablation remain unclear.

The cost-effectiveness of alternative surgical treatments for menorrhagia, relative both to each other and to medical therapy, has been subject to little detailed analysis. There have been several cost analyses of alternative forms of surgery [Manyonda and Varma, 1991; East *et al*, 1994; Nezhat *et al*, 1994; Messina *et al*, 1995; Brumsted *et al*, 1996], but these have generally been incomplete in their coverage of costs and based on small sample sizes without appropriate controls. Some of the generated data for cost analyses [Gannon *et al*, 1991; Summitt *et al*, 1992; Raju and Auld, 1994], but their methods have been unclear. To date, no data on the cost-effectiveness of surgical procedures have been published.

1.5   Structure of the thesis

The thesis is made up of eight further chapters, the details of which are summarised below.

Chapter 2 is a review of the issues and literature relating to the economics of MAS. The chapter considers the economic characteristics of MAS in terms of resource and non-resource consequences of these interventions relative to therapeutic baselines. A systematic review of published economic evaluations of MAS applications is presented, focusing on the particular evaluative methods employed in studies. Based on the economic characteristics of MAS and the review of published studies, the chapter identifies the methodological problems likely to be faced in economic evaluation in this clinical area.

Chapter 3 is the starting point of the empirical analysis, and takes the form of a cost-effectiveness analysis of AH versus TCRE alongside the RCT undertaken in Bristol. Taking a health service perspective, the costs of the two forms of management until two years follow-up are reported, and these are related to differential effectiveness in terms of women's satisfaction rates with treatment. Although the analysis in the chapter provides the firm evidential basis

for further development in subsequent chapters, it raises some specific areas of uncertainty about the relative cost-effectiveness of the two treatments. In particular, the need for a generic measure of benefit, the importance of reflecting women's preferences within the evaluative analysis and the limited external validity of trial-based evaluations are emphasised.

Chapter 4 details an analysis of women's descriptive preferences concerning the treatment of menorrhagia. Using a survey of 221 women referred to hospital for possible surgery for menorrhagia, the chapter explores women's attitudes to, and preferences for, the characteristics of treatment and the trade-offs between them. It is concluded that women are heterogenous in their treatment-related preferences in this area, and that their preferences for the characteristics of treatment often conflict with each other when a specific therapy has to be identified.

Chapter 5 details a cost-utility analysis of AH versus TCRE, which has the aim of assisting in resource allocation between specialties and disease areas by expressing cost-effectiveness in terms of a generic measure of benefit which partly reflects patients' preferences - the quality-adjusted life year (QALY). The analysis is based on resource and non-resource consequence data taken from the Bristol trial, augmented by health state valuation data elicited from a sample of 60 women with menorrhagia. The chapter indicates that, even with core data taken from a RCT, a range of uncertainties remains when conclusions are being sought about relative cost-effectiveness. Methods for the systematic handling of this uncertainty are considered which identify some robust conclusions, but areas of analytical uncertainty are highlighted, in particular in relation to the extent to which the QALY truly reflects patients' preferences.

Chapter 6 considers the issues related to the generalisability of economic evaluation of MAS. A framework for the analysis of the generalisability of economic analysis is offered. This takes the form of a series of sensitivity analyses of base-case models, which are based on trial evidence, by the

incorporation of alternative parameter estimates from data sources more closely related to routine clinical practice. Five specific analyses of generalisability are presented, focusing on specific alternative data sources: alternative RCTs; a survey of national clinical practice; the extremes of clinical practice in terms of resource use; hospital-specific unit costs; and developments in the technologies under evaluation. The chapter concludes that incorporating the results of routine practice into evaluations based initially on RCTs can generate major variations in cost and benefit estimates.

Chapter 7 explores the shortcomings of QALYs as a generic measure of benefit in the economic analysis of MAS. The feasibility of using an alternative measure - the *ex ante* HYE based on the TTO valuation instrument - is considered, and estimates of this benefit measure are elicited from a sample of 63 women referred to hospital with menorrhagia. The chapter concludes that there is a trade-off between the use of QALYs and *ex ante* HYEs, with the former likely to be easier to elicit but with a weak relationship to individuals' preferences, and the latter more likely to reflect preferences but imposing a considerable measurement burden.

Chapter 8 considers the novel concept of preference-driven treatment allocation in the area of minimal access surgery. Three models of preference-based management are evaluated in relation to abdominal hysterectomy and transcervical endometrial resection. It is concluded that these approaches have scope to be cost-effective forms of management, but may prove difficult to incorporate into routine practice.

Chapter 9 draws together the conclusions of the thesis, focusing on the contribution of the thesis to the methods of economic evaluation, with particular reference to minimal access surgery, and considering the implications for the relative cost-effectiveness of AH and TCRE.

Chapter 2

Minimal Access Surgery: Economic Characteristics and Implications for Evaluative Methods

_____

2.1   Introduction

This chapter reviews the actual and potential economic characteristics of MAS
and issues of method associated with the economic evaluation of this
technology. Firstly, it considers the types of technology which can usefully be
grouped under the headings of 'minimally invasive therapy' and 'minimal access
surgery'; secondly, it reviews the characteristics of the technologies as they
impact on the health benefits patients are likely to experience; thirdly, it
considers the implications of MAS for health service and societal resource use;
fourthly, it reviews a sample of economic assessments of MAS interventions;
and finally, it explores the methodological difficulties involved with the evaluation
of these technologies.

Medical science continues to generate a wide variety of new technologies. In the broadest sense, health care technologies include "the drugs, devices, and medical and surgical procedures used in medical care, and the organisational and supportive systems within which such care is provided" [Office of Technology Assessment, 1982, p200-1]. New technologies can impose a cost on health services [Neumann and Weinstein, 1991]. This cost includes not only any capital costs but also the costs of such things as staff, in-patient care, the operating cost of any hardware, any building space, consumables, drugs, training and any complications experienced by patients. Two crucial issues need, therefore, to be considered: how does the cost of new technologies compare with that of existing technologies and what additional benefits are being generated by these new technologies. It is becoming increasingly recognised that these two issues need to be formally addressed before new technologies diffuse widely within the health service [Advisory Group on Health Technology Assessment, 1992].

One set of therapeutic technologies which is beginning to diffuse within the UK National Health Service (NHS) is characterised, clinically, by a reduction in the physical trauma imposed upon patients, as a result either of avoiding penetration of the body or of using an endoscope or catheter to facilitate a therapeutic procedure. To date, the terminology associated with these technologies has been variable, but they have been generically referred to as 'minimally invasive therapy' or 'minimal access surgery'. Perhaps the most important example of this form of technology is endoscopic therapy, where treatment is undertaken at the end of a telescope which permits internal areas of the body to be viewed without the need for large openings to facilitate access. Laparoscopic therapy is one type of endoscopic treatment which has quite recently found a role in a number of high volume surgical procedures. Often referred to as 'keyhole surgery', laparoscopic therapy is centred around a number of small incisions through which instruments are inserted to facilitate the viewing and treatment of a given body cavity. The absence of the large incisions associated with conventional open surgery usually results in less post-operative trauma to the patient and a shorter period of convalescence.

It is clear that these new forms of treatment have important implications for the NHS, largely because of the likely extent of their use. Clinicians have predicted their rapid diffusion, to the extent that they will replace virtually all open surgery [Wickham, 1987]. A working party convened by the Department of Health and the Scottish Home and Health Department forecast that, by the year 2000, 70% to 80% of surgical practice will be based on MAS techniques [Cuschieri, 1993]; MAS is currently used in about 20% to 30% of surgical interventions in NHS hospitals [Welsh Health Planning Forum, 1994]. It remains true, however, that few applications of MAS have been subject to thorough clinical and economic evaluation, despite their enormous potential scope [Banta, 1993A; Lancet, 1993; Pearson, 1994]. By radically altering the process by which much surgery is delivered, there is likely to be an important impact on the costs and benefits of the relevant therapeutic procedures. Whether this results in an overall improvement or deterioration in the *efficiency* of health service delivery is, as yet, unclear.

2.2   Defining the relevant set of technologies

2.2.1 Minimally invasive therapies

Over many years, health care has become progressively less physically traumatic to the patient. Technological developments have continued to reduce the disbenefits associated with the process and short-term outcomes of health care interventions. For example, hysterectomy is now only rarely used to manage cervical intra-epithelial neoplasia, which can be treated at a woman's first visit to hospital following an abnormal smear, using loop diathermy resection [Giles and Gafar, 1991]; and developments in pharmaceuticals have resulted in many patients with peptic ulcer avoiding surgery [Paimela *et al*, 1991].

The term 'minimally invasive therapy' (MIT) can, in principle, refer to any treatment which has replaced a more invasive alternative. As such, the membership of this group of therapies is enormous, encompassing a significant proportion of modern therapeutic clinical practice. It would include non-invasive

```
┌─────────────────────────────────────────────────────────────────────────────┐
│  ┌──────────────────────────────────────────────────────────────────────┐   │
│  │                   Minimally Invasive Therapies                         │   │
│  │                                                                        │   │
│  │  Those therapies which seek to replace a more invasive alternative. For example, pharmacotherapy and │
│  │  endoscopic therapy.                                                   │   │
│  └──────────────────────────────────────────────────────────────────────┘   │
│                                                                               │
│       ┌────────────────────────────────┐    ┌─────────────────────────────┐  │
│       │    Minimal Access Surgery (MAS)│    │     Non-Invasive Therapy    │  │
│       │                                │    │                             │  │
│       │  Those minimally invasive therapies which are invasive. │ Those therapies which do not require penetration of │
│       │  For example, endoscopic and percutaneous therapy. │ the body. For example, pharmacotherapy and │
│       │                                │    │  external radiotherapy.     │  │
│       └────────────────────────────────┘    └─────────────────────────────┘  │
│                                                                               │
│   ┌──────────────────┐  ┌──────────────────┐                                 │
│   │  New therapies   │  │  New Processes   │                                 │
│   │                  │  │                  │                                 │
│   │ MAS which represents │ MAS which represents a │                          │
│   │ a new way of treating a │ different way of doing a │                     │
│   │ condition. For example │ standard procedure. for │                       │
│   │ TCRE.            │  │ example laparoscopic │                             │
│   │                  │  │ hysterectomy.    │                                 │
│   └──────────────────┘  └──────────────────┘                                 │
│                                                                               │
│   ┌──────────────────┐  ┌──────────────────┐                                 │
│   │ Existing Clinicans │  │  New Clinicians │                                │
│   │                  │  │                  │                                 │
│   │ New procedures   │  │ New procedures   │                                 │
│   │ undertaken by same │  │ undertaken by new clinical │                     │
│   │ clinical professionals as │ professional groupings. │                    │
│   │ conventional surgery. │ For example, PTCA by │                           │
│   │ For example, TCRE by │ cardiologists    │                                │
│   │ gynaecologists   │  │                  │                                 │
│   └──────────────────┘  └──────────────────┘                                 │
└─────────────────────────────────────────────────────────────────────────────┘
```

Figure 2.1   Sub-categorisation of minimal invasive therapies. TCRE = transcervical resection of the endometruim; PTCA = percutaneous transluminal coronary angioplasty.

treatments such as pharmacotherapy, radiotherapy and shockwave lithotripsy, as well as invasive therapies which have replaced more radical procedures, such as percutaneous and endoscopic therapies. If the purpose of organising technologies into categories is to discuss the nature and implications of their

shared characteristics, it can be argued that the interventions which logically fall within the group labelled MIT are too heterogenous to facilitate useful analysis.

2.2.2 Minimal access surgery

A possible movement from this broad grouping of technologies to a more selective one is illustrated in Figure 2.1. The term 'minimal access surgery' (MAS), which excludes *non-invasive* procedures, has been defined to include five minimal access surgical approaches: laparoscopy (eg. laparoscopic cholecystectomy), endoluminal endoscopy (eg. endometrial ablation), perivisceral endoscopy (eg. nephrectomy), thoracoscopy (eg. pleurectomy) and intra-articular (eg. menisectomy) [Cuschieri, 1991]. Treatments within this more limited group share some key characteristics in that they are all endoscopic therapies and, therefore, all invasive and they use similar equipment. There would seem to be good reason also to include percutaneous therapies (eg. percutaneous transluminal coronary angioplasty (PTCA)) amongst the MAS group: although the catheter and the x-ray replace the endoscope, these treatments too are invasive but have replaced forms of open surgery.

A further level of categorisation is shown in Figure 2.1. MAS can be divided into those therapies which represent not only a less invasive means of treating a given condition than was previously the case, but also treat in a quite different way; and those therapies which are just a less invasive way of doing an established procedure. Examples of the first category include TCRE which, unlike established surgical treatment for menorrhagia using hysterectomy, leaves the uterus in place [Dwyer *et al*, 1993]; and PTCA, where arterial stenoses and occlusions are re-vascularised using a catheter, guidewire and balloon rather than by coronary artery bypass graft (CABG) [Goodman, 1992]. Examples of the second category of MAS include laparoscopic cholecystectomy, which removes the diseased gallbladder without the need for a large surgical opening [Wolfe *et al*, 1991A], and laparoscopic-assisted vaginal hysterectomy, where the removal of the uterus is made easier by the use of endoscopic methods [Olsson *et al*, 1996].

16

A further way of categorising entirely new forms of treatment is to distinguish between those that are undertaken by the same professional clinical group as the conventional form of surgery and those that are undertaken by a new clinical group. An example of the first type of procedure is TCRE which is undertaken by the same gynaecologists who undertake hysterectomy. Examples of the second include gastroenterologists who now treat a range of gastrointestinal tract conditions using endoscopy, where treatment would previously have been the preserve largely of the general surgeon; another example is the use of PTCA by cardiologists as a form of coronary artery revascularisation which would previously have been undertaken by cardiac surgeons using CABG.

MAS is a set of therapeutic procedures rather than a homogeneous technology. It is clear, however, that these therapies share some important characteristics which have implications for economic efficiency and for evaluative methods. The remainder of this chapter, therefore, concentrates on MAS, although many of the points made also apply to the broader category of MIT.

2.3   Implications of MAS: patients' benefits

2.3.1 Short-term benefits, long-term uncertainties

Conventional open surgery has some very clear implications for patients' experiences of therapy. These arise not only from the fact that a large opening is required through which the surgeon gains access to the operating field, but also from the fact that retraction, handling and instrument-related trauma will cause tissue damage and the exposure, cooling and drying of internal structures [Hirsch and Hailey, 1992]. Inevitably this will result in patients' health status deteriorating following open surgery: post-operative pain and limitations in physical functioning often require patients to remain in hospital for several days following their operation and limit their ability to return to usual activities for some weeks.

These 'short-term disbenefits' of conventional surgery will cause more distress to some groups of patients than to others. Those particularly affected will be individuals with dependent children and with friends and relatives requiring long-term care; patients who live on their own with nobody to help them during their convalescence; patients for whom a prolonged period away from work has a high personal opportunity cost; and individuals who have jobs requiring strenuous physical activity who may be unable to return to work for particularly long periods.

The process of conventional surgery may represent more than a short-term disbenefit to some. Individuals who are frail due to old age or concomitant illness will be at risk of serious post-operative morbidity or of mortality, to the extent that open surgery is often avoided altogether. Moreover, open surgery has some clear cosmetic disbenefits; for many, the existence of a large scar - which will fade over time but which may not altogether disappear - will be considered more than just a short-term disadvantage.

The disbenefits associated with the process and short-term outcomes of conventional surgery have contributed to the clinical development and diffusion of MAS. In the case of those forms of MAS which are simply less invasive means of undertaking a standard therapy, the advantages seem obvious. However, a randomised trial comparing laparoscopic and open appendicectomy found that there were no statistically significant differences between patients undergoing the two treatments in terms of postoperative pain, analgesic requirements and the proportions having returned to normal activity after three weeks [Tate *et al*, 1993].

The rapid diffusion of laparoscopic cholecystectomy is likely to reflect the perception that it reduces pain and disability, relative to the open form of the procedure, without an apparent increased mortality or overall morbidity [NIH Consensus Development Panel on Gallstones and Laparoscopic Cholecystectomy, 1993]. There has, however, been little thorough evaluation of the technique and there have been reports of complications in some patients

18

[Smith, 1991; Wolfe *et al*, 1991B; Shanahan and Knight, 1992; Messahel, 1995] and suggestions that minilaparotomy cholecystectomy may be a preferable technique [Baxter and O'Dwyer, 1992].

As regards those types of MAS which are not only less invasive than open surgery but also characterised by a new therapeutic approach, it would appear that the perception of improvement in process and short-term outcomes is again the key to their development. For example, TCRE was being used by almost 50% of gynaecologists to treat menorrhagia by as early as 1991 [Royal College of Obstetricians and Gynaecologists Medical Audit Unit, 1991], prior to data from randomised controlled trials (RCTs). Subsequent trial data indicated that, despite a clear reduction in post-operative pain and time until return to usual activities, it is not as effective as hysterectomy in relieving menstrual symptoms [Gannon *et al*, 1991; Dwyer *et al*, 1993; Pinion *et al*, 1994]. Similarly, PTCA is associated with increased post-operative incidence of angina and of repeat hospitalisations relative to CABG [Pocock *et al*, 1995], but is an established means of treating some forms of ischaemic heart disease.

The ability of MAS procedures to allow patients to return to their usual activities faster than would conventional surgery has been a particularly strong 'selling point' of the technology. Most of the clinical papers on laparoscopic cholecystectomy have focused much greater attention on patients' duration of convalescence than is the norm in clinical journals [Reddick and Olsen, 1989; Barkun *et al*, 1992; Stoker *et al*, 1992; McMahon *et al*, 1994; Stoker *et al*, 1994]. There has, however, been a concentration on the time patients take until they actually return to *work* rather than are able to resume *usual activities*. The *American Journal of Surgery* published a paper devoted to the consideration of the extent of interruption of professional and home activity in a sample of patients in the US and France following laparoscopic cholecystectomy [Vitale *et al* 1991]. The authors reported that 63% of Americans in their sample, compared with 25% of French, returned to work within two weeks of their operation, emphasising that there is an important distinction between returning to work and being able to go to work.

19

A key characteristic of MAS is the extent to which particular examples of the technology have diffused widely as a result of these types of perceived short-term benefit without a great deal of consideration of longer term outcomes. An illustration of the potential dangers that may result from this is the treatment of benign prostatic hypertrophy using transurethral resection of the prostrate (TURP) - a minimal access alternative to open prostatectomy introduced in the 1970s and used in 95% of prostatectomies undertaken in non-federal, short-stay hospitals in the US by 1986 [Concato *et al*, 1992]. This rapid diffusion occurred without formal evaluation, but reflects the perceived advantages of TURP in terms of it being less invasive relative to open prostatectomy [Wennberg, 1990]. Several retrospective studies using large administrative databases in several countries have, however, claimed that patients undergoing TURP have a higher incidence of stricture and of re-operation, and an elevated risk of death, relative to open prostatectomy [Wennberg *et al*, 1987; Wennberg *et al*, 1988; Roos *et al*, 1989; Malenka *et al*, 1990; Andersen *et al*, 1990; Sidney *et al*, 1992]. Although it is possible to question the ability of these studies fully to control for differences in case-mix [Concato *et al*, 1992], the case of TURP emphasises that beneficial changes in the process of care, resulting in less immediate post-operative pain and a shorter convalescence, may not automatically result in an overall improvement in effectiveness.

It is possible to identify a trade-off facing patients when deciding whether to choose some forms of MAS rather than conventional surgery. The short-term benefits for MAS, which are largely related to the process of treatment, may be accompanied by longer term disbenefits. These disbenefits can sometimes be defined (ie. *risks*): for instance, a significant proportion of women undergoing TCRE require repeat treatment to alleviate their menstrual symptoms; there may be an excess risk of mortality in men undergoing TURP. Occasionally, there is widespread ignorance amongst clinicians about longer-term outcomes because patients either have not been followed up for a sufficiently long period, or follow-up has not been undertaken systematically (ie. disbenefits cannot be defined; there is *uncertainty*). It is by no means clear whether patients are fully aware of

the risks and uncertainties associated with longer term outcomes when they agree to undergo these types of MAS.

2.3.2 The dynamic impact of MAS: changing thresholds for surgical treatment

As well as facing trade-offs when deciding whether or not to choose MAS rather than conventional surgery, there are trade-offs associated with the choice of MAS rather than non-surgical treatment, and the advent of MAS may alter the *thresholds* relating to decisions about whether or not to treat patients surgically. Such a change may fundamentally alter the *mix* of patients undergoing surgical intervention. The pressures that alter surgical thresholds can originate from the patient and/or from the clinician, depending upon the nature of the decision making process.

Adopting the 'patient as consumer' perspective, it is possible to see how MAS may influence surgical thresholds. Prior to the advent of MAS, some types of patient, whose symptoms were not sufficiently serious to justify what they perceived as the short-term disbenefits of surgery, postponed such treatment indefinitely. These patients, therefore, considered that the private costs of surgery (the private opportunity cost of the time spent in hospital and convalescing rather than at work, at leisure or fulfilling family commitments; and any direct costs such as those associated with child minding) were greater than the expected *net* benefits in terms of relief of symptoms. These benefits might be modest as a result of the underlying disease and symptoms not being particularly serious or because of the risks and short-run disbenefits of surgery being high. Unlike some forms of MAS, open surgery is usually undertaken under general anaesthetic, and the perceived mortality risk associated with this may reduce expected net benefits further. The availability of MAS is likely to influence that private trade-off, with the *expectation* of equivalent (or greater) net benefits at reduced (or equivalent) private cost.

The 'doctor as agent' perspective would suggest a similar weighing up of the costs and benefits of conventional surgery, but from the viewpoint of the health

service or the individual clinician: the net benefits anticipated for the patient would not justify the health service inputs required or the clinician's own time commitment. The availability of MAS may again alter this cost-benefit trade-off if it requires fewer health service resources.

It would appear that the first of these perspectives might have more relevance in relation to some forms of MAS than other areas of health care. The availability of minimal access alternatives to hysterectomy, for example, has been widely publicised in the lay press, and gynaecologists have claimed that patient demand has been an important factor behind the rapid diffusion of these procedures [Sutton, 1993]. Moody [1992] has argued that the public became aware of laparoscopic cholecystectomy, through the media, before most surgeons; and Banta *et al* [1993B] suggested that patient demand influenced the diffusion of PTCA and laparoscopic cholecystectomy.

Some evidence does exist to support the hypothesis that the availability of MAS alters surgical thresholds. Using health maintenance organisation (HMO) claims data, Legorreta *et al* [1993] found that cholecystectomy rates had increased since the introduction of the laparoscopic form of the procedure. Although the rate of open procedures declined between 1988 and 1992, the *total* cholecystectomy rate increased from 1.37 per 1000 enrollees to 2.15 per 1000. Furthermore, as a direct result of this increased cholecystectomy rate, total HMO medical expenditures for cholecystectomy increased over this period by 17.8%. This occurred despite a reduction of 25.1% in the unit cost of a cholecystectomy. In an attempt to exclude the possibility that a change in the patient or clinician population was influencing the findings, the authors compared these results with those for appendectomy and inguinal herniorrhaphy, where the diffusion of MAS methods had hitherto been limited. They found no significant increases in operation rates or medical expenditures for these comparator procedures. The findings of the study contrast with the generally stable rates of cholecystectomy in the US during most of the 1980s prior to the introduction of MAS [Diehl, 1993]. Similar evidence on a change in surgical thresholds for cholecystectomy has been generated in other US studies [Steiner *et al*, 1994;

Fendrick *et al*, 1994], as well as studies in Canada and Australia [Marshall *et al*, 1994] and the UK [Lam *et al*, 1996].

A second possible example of a changing surgical threshold is in the area of the management of peripheral vascular disease. Studies have shown that, since the late 1970s, the use of percutaneous transluminal angioplasty (PTA) has increased significantly, but without a reduction - often with an *increase* - in the use of bypass surgery [Anderson *et al*, 1986; Jeans *et al*, 1986; Veith *et al*, 1990; Tunis *et al*, 1991]. The situation relating to peripheral vascular disease appears to be mirrored in *coronary* vascular disease where the utilisation of bypass surgery has failed to decrease despite the rapid expansion of PTCA [Anderson and Lomas, 1988; Feinleib *et al*, 1989; Weintraub *et al*, 1990].

Evidence also exists to indicate that the diffusion of MAS techniques to treat menorrhagia has shifted the referral threshold for surgery in that area. Bridgman [1994] found that, for the Mersey region of the NHS, the standardised operation ratio for dysfunctional uterine bleeding rose to 135 [95% confidence intervals 129-141], compared to a 1990-91 baseline. Similar results have been found for the Oxford region of the NHS [Coulter, 1994].

It has been argued that surgical thresholds have been shifted so significantly in some areas that MAS procedures are being used inappropriately. Spiro suggested that "diagnostic laparoscopic cholecystectomy" is being carried out in some centres where, because the procedure is straightforward relative to open surgery, it is undertaken on patients with abdominal pain who have gallstones *incidentally* uncovered by ultrasound but who do not have symptoms of biliary colic [Spiro 1992]. Spiro characterises the surgeons perspective as "Well you've got gallstones, and we can't do very much about your indigestion. Let's take out your gallbladder, since this is now so easy to do, and see if you get better. Those stones aren't doing you any good, and they could be the cause of your trouble" [p.167]. It has been suggested that there is evidence of similar developments outside general surgery. On the basis of a series of 171 patients referred for a second opinion regarding the need for PTCA, it has been estimated

that 50% of PTCAs undertaken in the US are unnecessary, or at least could be postponed [Graboys *et al*, 1992].

2.4   Implications of MAS: resource use

It is likely that the development and diffusion of MAS will have important implications for health service resource use. It is useful, however, to draw a distinction between short- and long-term implications.

2.4.1 Short-term impact on health service resource use

In the short-term, MAS will continue to be adopted across most surgical specialities, and this will impact on two key categories of resource use.

Hospital stay. Commentators have frequently indicated the significant effect of MAS on duration of hospital in-patient stay [Hoare, 1992; Banta, 1992; Wickham, 1993]. Although some applications of MAS appear to provide little reduction in hospital length of stay [Tate *et al*, 1993; Stoker *et al*, 1994], this claim has evidence to support it for many forms of MAS. For example, studies have shown a significantly shorter initial hospital stay, relative to conventional surgery, in laparoscopic colectomy [Falk *et al*, 1993], PTCA [Sculpher *et al*, 1994], TCRE [Dwyer *et al*, 1993] and laparoscopic cholecystectomy [Kesteloot and Penninckx, 1993]. Although this implication of MAS will probably have a positive impact on both patient benefits and health service costs, it is worth noting the following caveats.

The first caveat is that a reduction in the demand for hospital beds may not automatically result in a realisation of cash savings. In the short-term, it is unlikely that MAS will facilitate the closure of surgical wards or redeployment of ward nursing staff, which are necessary to reduce cost markedly, because MAS procedures currently represent only a proportion of total surgical procedures, and there remains a need for conventional surgical backup facilities for those MAS procedures that can result in complications [Wilson *et al*, 1986]. Moreover, it is

unlikely that reduced demand for hospital beds will permit a significant number of patients waiting for elective surgical procedures (MAS or conventional) to undergo their treatment earlier than otherwise would be the case, because MAS rarely frees up significant amounts of other resources used for surgical patients, such as theatre time. Another way of viewing this is that MAS has a different *production function* to conventional surgery, requiring fewer inputs of in-patient bed days. Ideally, the health service needs to alter the mix of inputs it has available for the production process (by reducing bed days) to free-up cash for other inputs. In the short-run at least, however, the health service faces constraints when trying to change this mix.

A second factor that may limit the cost impact of a reduced demand for hospital beds is the importance of the *baseline* therapy. Some MAS procedures are replacing non-invasive therapies rather than open surgery. Occasionally this type of therapeutic shift has the potential to happen for the bulk of patients with a given condition. For example, trials in the US have indicated that argon laser trabeculoplasty may be more effective than medical management for primary open angle glaucoma [Glaucoma Laser Trial Research Group, 1990]; having largely replaced intra-ocular surgery for the condition in patients for whom medical therapy is ineffective [Glaucoma Laser Trial Research Group, 1989]. The possible increase in the utilisation of primary PTCA, instead of the conventional use of intravenous administration of thrombolytic agents, to achieve coronary artery recanalisation after acute myocardial infarction, is another example [Grech and Ramsdale, 1993]. A further example is the use of laser laparoscopy rather than expectant management to treat pelvic pain associated with endometriosis [Sutton *et al*, 1994]. More frequently, MAS is replacing a non-invasive therapy for a *sub-group* of patients because of the shift in surgical thresholds discussed above. It may be the case, therefore, that the use of some types of MAS *increases* demands on hospital-resources, including in-patient beds, because they replace medical management (or no intervention at all) rather than the conventional surgical intervention usually assumed.

A third caveat with regard to the impact of MAS on in-patient days is that a general reduction in post operative lengths of hospital stay had begun, in most developed countries including the UK, prior to the significant diffusion of MAS [Department of Health, 1990]. A long term trend can be identified in specialties where MAS has only recently began to diffuse: in one English region, the average length of stay per episode in gynaecology fell from 5.1 days in 1975 to 3.0 days in 1985 [Ferguson *et al*, 1991]. Some of this general trend is doubtless associated with the growth of day-case surgery which can be used for some non-MAS surgical procedures, reflecting a change of attitude in the health service [Audit Commission, 1992; Royal College of Surgeons of England, 1992]. For example, the length of stay for hernia repair fell from 10 days to 24 hours without a pronounced change in surgical technique [Johnson, 1994]; laparoscopic hernia repair has no advantage over open repair in terms of length of stay [Stoker *et al*, 1994]. It may be argued, therefore, that the more widespread use of MAS is being superimposed onto a system already displaying a reduced need for hospital beds and that the *additional* savings resulting from MAS may not be as significant as is frequently claimed.

A final caveat regarding the link between MAS and reduced lengths of hospital stay is that most commentators who emphasise this link concentrate on *initial* hospitalisations. Some forms of MAS require patients to return for subsequent treatment more frequently than they would have had they undergone conventional surgery. For example, the Bypass Angioplasty Revascularization Investigation (BARI) trial comparing PTCA with CABG in patients with multi-vessel coronary artery disease found that, after a mean follow-up of 5.4 years, 20.5% of the 915 patients randomised to PTCA required at least one subsequent CABG, 23.2% required at least one further PTCA and, in addition, 10.8% needed at repeat PTCA and a CABG; of the 914 patients randomised to CABG, the respective rates were 0.7%, 6.9% and 0.4% [BARI Investigators, 1996].

Therapeutic resources. The advent of MAS may influence the demand for resources such as theatre/treatment room time, staff number and mix, anaesthetic and therapeutic equipment, consumables, gases and drugs. There is,

however, no clear link between the use of MAS procedures and a *reduced*
demand for these types of therapeutic resource. Some forms of MAS offer
savings in some categories but place additional demands on others. For
example, the use of the neodymium yttrium aluminum garnet (Nd:YAG) laser to
palliate advanced colorectal cancer, rather then palliative surgical resection,
substitutes less expensive endoscopy suite time for theatre time, and requires
less anaesthetic and staff support, but needs investment in expensive laser
equipment [Vondeling, Mathus-Vliegen and Banta 1991; Sculpher, 1993].
Furthermore, laparoscopic procedures are undertaken in theatre, and it has been
suggested that they require up to 75% *additional* time to be completed compared
to open surgery [McCloy, 1992]. For example, laparoscopic cholecystectomy
has been shown to take longer than the open procedure in a number of countries
[Hirsch and Hailey, 1992; Kesteloot and Pennickx, 1993]. TCRE, on the other
hand, requires less time in theatre, fewer staff and puts less demand upon
anaesthetic resources; even allowing for additional equipment costs, overall
operation costs are less than those associated with abdominal hysterectomy (see
Chapter 2).

The use of MAS procedures can, therefore, either increase or decrease the
demands on therapeutic resources. Indeed, generalisation about therapeutic
resource use is difficult even in relation to a specific form of MAS due to
differences in how they are applied in clinical practice and the speed of
development over time. For example, there is considerable variation between
clinicians undertaking laparoscopic surgery in whether they use disposable
consumables or re-usable equipment, with significant cost implications (see
Chapter 6). Therapeutic resource use is also influenced by the different versions
of operations which exist. For example, the use of laparoscopic techniques in
hysterectomy has resulted in a number of specific procedures developing
including laparoscopic-assisted vaginal hysterectomy, total laparoscopic
hysterectomy and laparoscopic-assisted doderlein hysterectomy [Garry *et al*,
1994].

Furthermore, the caveats outlined above regarding the reduced demand for in-patient beds associated with MAS need to be born in mind here too. In short, the effect of MAS on short-run hospital costs depends very much on the MAS application and the baseline against which it is compared. Although a number of studies indicate advantages in these sorts for costs for MAS technologies [Kesteloot and Pennickx, 1993], others suggest that some applications of MAS may be more costly than their comparators [Cuckow, 1994; McMahon *et al*, 1994].

2.4.2 The costs to the health service of transition

McKinlay [1981] identifies seven stages in the career of a medical innovation, running from "promising report" to "erosion and discreditation". The *process* of shifting MAS procedures from the "promising report" to the "standard procedure" is not costless [Gelijns and Fendrick, 1993]. For many MAS procedures, there is a need to invest in new capital equipment: endoscopes, videos, monitors, lasers. The 'learning curve' related to new ways of undertaking therapy also imposes costs. The development of a new MAS procedure from use in a limited number of clinical centres and by enthusiasts into routine clinical practice and widespread use requires organisation and training. In the NHS there is no systematic process to achieve this. The *initial* diffusion of skills often comes through workshops and seminars facilitated by the funding of companies manufacturing or marketing the equipment used as part of MAS procedures. Sometimes enterprising clinical enthusiasts take the initiative by running courses, 'selling' new procedures in terms of the financial advantages to the health service and perceived patient and professional benefits [Bryan, 1993]. Either way, resources are expended in this process: both the time participating clinicians spend away from their practice and the time that goes into organising and providing the seminars, workshops and courses.

Costs are also associated with the transition of clinicians to being accomplished practitioners of new MAS procedures; that is, with the movement up the learning curve. Inexperienced practitioners are likely to require more resources to complete a procedure than experienced ones. For example, one study reported a

mean operating time for laparoscopic cholecystectomy of 2.4 hours for the first 50 patients receiving treatment, against a mean of 1.8 hours for the next 65 patients [Hirsch and Hailey, 1992]. Moreover, it is likely that a clinician at the foot of the learning curve will not be so effective as a more experienced clinician in undertaking a new form of MAS [Still and Walsh, 1992]. This may result in additional health service costs due to an increased likelihood of complications or technical failure [See et al, 1993]. There have been some well-published examples of complications with MAS procedures which have apparently been due to a lack of experience on the part of the clinician [eg. Times, 1992]. Indeed, some clinicians may never fully climb the learning curve associated with a new MAS procedure. Evidence exists to support the view that clinicians need to undertake a procedure frequently to become accomplished practitioners [Showstack et al, 1987; Luft et al, 1987; Cromwell et al, 1990; Luft et al, 1990; Hannan et al, 1991; Woods et al, 1992; Farley and Ozminkowski, 1992], although the quality of these observational studies has been questioned [NHS Centre for Reviews and Dissemination, 1995]. If volumes are insufficient for this purpose, costs are likely to be imposed on the health service and patients.

It is important to emphasise, however, that the costs associated with clinicians moving up learning curves are in no way unique to MAS procedures. Any new therapeutic practice will impose transition costs; indeed fully established surgical procedures may impose 'learning curve costs' if clinicians undertake too few to become proficient. Furthermore, the cost of a new MAS procedure undertaken by an inexperienced clinician may still be lower than that of a conventional surgical procedure carried out by an experienced surgeon. The overall benefits for patients too may be higher; hence patients may be prepared to accept the additional risk of complications or technical failure related to a 'sub-optimal' MAS procedure, relative to conventional surgery undertaken by an experienced practitioner.

2.4.3 Long-term impact on health service resource use

The foregoing discussion would suggest that the assumption that the development and use of MAS procedures automatically reduces the cost of

health service resource use in the short-run is possibly unfounded. Uncertainty surrounds this issue, however, because of the absence of good quality economic evaluations of new MAS procedures compared to conventional therapy. Knowledge about the effects of the greater use of MAS on resource use in the *longer term* is even more sketchy, although various commentators have made predictions [Wickham, 1993; Wickham, 1994; Banta *et al*, 1993A; Banta, 1993B]. For instance, Wickham [1993] predicted that, in the longer term, the diffusion of MAS will have the following sorts of organisational effect.

(a)   The large general hospital of over 350 beds will be replaced by "single storey 'stand alone production units'" [p12] because there will be fewer in-patients at any given time.

(b)   There will be a greater need for well trained nursing staff, experienced in the various types of MAS.

(c)   Greater cooperation will be required between providers of care in the hospital, clinic and community.

(d)   There will be a radical change in the nature of therapeutic facilities, with a movement away from the traditional operating theatre towards purpose-built therapy suites containing radiological, ultrasonic and endoscopic facilities, and multiple-monitor displays.

(e)   There will be less need for the type of methods traditionally required to achieve asepsis during an operation such as 'gowning up'; there will also be a reduced demand for sterile operating clothing. More sophisticated ways of sterilising modern equipment such as endoscopes may be required, however.

(f)   Changes will continue in the type of doctor who undertakes interventional therapy; for example, recent changes have resulted in an increased role for interventional radiologists and physician gastroenterologists in areas previously the preserve of vascular and general surgeons, respectively. Wickham sees a declining role for the surgeon: "it would seem that surgeons are going to have to accept the changing status of being only one member of a group of interventionalists and not the present 'leader of the pack'" [p13].

(g)   The changing type of doctor will alter specialty boundaries. Wickham
       predicts further specialisation but with a declining role for the surgeon.
       "It may well be that organ specific physicians will after diagnosis direct
       the patient to the most efficient 'sub-contractor interventionalist'" [p13].
       Similarly the surgeon's role in pre- and post-operative care may well give
       way to more involvement by anaesthetists and intensive care physicians.

However, in general, these longer-term implications of MAS are extrapolations
from developments which have been taking place in medicine for some years.
For various reasons, hospitals are less pre-occupied with open surgery than
previously was the case and there has been an increased role for non-invasive
therapies such as drugs, more sophisticated forms (and hence often greater
utilisation) of diagnostic technologies and the development of hospital-based
screening and assessment programmes. Although probably a key explanatory
factor, innovations in surgical techniques are only partly responsible for the
changing face of medicine, and of the hospital and its staff; the specific
development of MAS has an even more limited explanatory role in these
developments.

It is reasonable to suggest that the diffusion of new forms of MAS may
accelerate the sorts of changes Wickham outlined, but it is important to
remember that MAS techniques are unlikely totally to replace conventional
surgery. Hirsch and Hailey referred to studies which report between 1.8% and
10% of laparoscopic cholecystectomies being converted to open procedures
[Hirsch and Hailey, 1992]; Tate *et al* [1993] report that 20% of laparoscopic
appendicectomies require conversion to open surgery. Although these types of
result may well be a reflection of the process of climbing the learning curve
referred to above, the need to convert from MAS to open techniques *at short
notice* may limit the speed and extent of hospital change.

It is unlikely that the sorts of long-term change in health care in general and the
hospital in particular identified by Wickham will reduce the total amount (and
value) of resources society allocates to health care. Although there may be

31

some release of resources following the possible closure of traditional in-patient wards and surgical theatres, new forms of infrastructural investment in such things as interventional suites and therapeutic and diagnostic equipment may well *at least* absorb these savings. Furthermore, the perceived advantages of greater *throughput* achieved by hospitals as a result of declining patient lengths of stay may be cancelled out by a re-discovery of the post-war hidden-iceberg of illness: the changing thresholds to surgical intervention discussed above is a specific example of the more general phenomenon that new breakthroughs in medicine generate new 'needs' which eventually lead to increased utilisation of health care resources. As such, MAS may represent in the 1990s what renal dialysis was in the 1960s and plastic hip replacement surgery was in the 1970s [Klein, 1989].

2.4.4 The impact on resource use outside the hospital

In addition to the effects that MAS may have on hospital-based resource use, there is likely to be an impact on the resource use consequences of health care outside the hospital.

Community-based health services. It is possible that any cost reductions related to the shorter lengths of in-patient hospital stay associated with MAS, relative to conventional surgery, will be offset by a greater burden on community-based health services. The post-operative care which was previously provided by hospitals is now more frequently the responsibility of primary health care teams and district nurses. However, as noted above, shorter lengths of stay and the growth of day-case surgery have been generally evident over recent years, and are not solely related to increased use of MAS. To the extent, therefore, that it results in less post-operative morbidity for the patient, some forms of MAS may *reduce* the burden on community-based health services.

A study from the Welsh Health Planning Forum has attempted to quantify the organisational changes that are likely to take place as a result of developments in MAS in the area of gastroenterology [Warner *et al*, 1993]. Using a Delphi panel of clinical and technological experts and site visits to a number of clinical centres

in the US, Sweden, the Netherlands and Canada, the study explored the following hypothesis: "that as a result of emerging technology in the fields of diagnostics, treatment and communications, the role and function of the acute hospital, and its relationship with community services and general practice, could by the year 2002 alter considerably" [p.5]. The authors report that "in general the results of the Delphi meeting strongly support this hypothesis" [p.29].

On the basis of the Delphi meeting, the study reached some additional conclusions regarding gastroenterology:

(a)   there will be an increased use of day-case surgery in gastroenterology because of greater use of endoscopic and other activities;

(b)   there will be a greater use of community-based minimally invasive techniques where GPs initiate (and sometime directly undertake) endoscopic and other forms of diagnosis (which will be open-access) and initiate and monitor therapy;

(c)   there may, therefore, be an increased role for GP decision making but, due to the growth of specialist radiologists and endoscopists, this may be more apparent than real; and

(d)   the community-based approach would be more "clinically efficient" [p.29].

Patients' private cost. Increased use of MAS procedures may also affect patients' private costs. As noted above, it is likely that patients who undergo MAS rather than conventional surgery will incur fewer direct costs such as those associated with child minding. Perhaps more importantly, the shorter lengths of stay associated with the shift from conventional surgery to MAS will reduce the time patients are required to allocate to the process of health care and the period of convalescence. This time is of value to individuals, in terms of either work-related income which for some individuals is reduced or removed as a result of illness, or the enjoyment of leisure time. The link between MAS procedures and reduced private costs may not, however, be automatic for the reasons discussed above in relation to reduced lengths of hospital stay.

33

Production losses. It has been argued that, even if MAS procedures are more expensive than conventional surgery in terms of direct health service costs, these costs may be offset by the production gains resulting from patients' earlier return to normal activities [McCloy, 1992]. As noted above, the impact of MAS on the period until return to usual activities has been of considerable interest in *clinical* evaluations. Patients' length of convalescence will undoubtedly have an important effect on their health-related quality of life (HRQL), which is a key outcome of these interventions, although part of this impact will be a reflection of improvements in other key domains of HRQL such as post-operative pain. Whether patients' early return to normal activities should also be valued in monetary terms is an area of methodological controversy considered in more detail in Section 2.6.2.

One of the important conclusions of the foregoing discussion is that MAS is not *synonymous* with improved benefits to patients, or with reduced costs to the health service or to society more widely. Although it may well be true that many MAS procedures reduce the level of trauma patients experience with surgery, some major assumptions are required to conclude, without empirical support, that these procedures improve health. Furthermore, even if a net benefit to patients is proven, it is not necessarily the case that these benefits will be considered of sufficient value to justify any additional cost.

2.5   A review of published economic evaluations of MAS

2.5.1 Introduction

There is now a general acceptance amongst clinicians and health service managers of two 'health care truisms': the amount of resources available for society to devote to maintaining and improving health is, and always will be, finite; and the opportunities available to the health service to attempt to influence health are continuing to grow rapidly. It is, therefore, incumbent upon society to assess the extent to which particular health care technologies generate the sorts of benefits that justify the resources devoted to them.

There now exists an established set of general methods to evaluate the value for money offered by health care technologies [Drummond *et al*, 1987; Eisenberg, 1989; Luce and Elixhauser, 1990; Pettiti, 1994; Sloan, 1995], although some of the specific analytical techniques required to undertake an economic evaluation remain controversial [Drummond *et al*, 1993A]. Most groups of health care technology raise particular problems and issues as regards their economic evaluation, where this may be related to the clinical area to which they are relevant or to the nature of the technologies themselves. This section of the chapter considers the methodological issues related to the economic evaluation of MAS in more detail.

Much can be learnt about the methods of economic evaluation from reviewing published studies in the area. A systematic review of published economic analyses has been undertaken with he following objectives:

(a)   to describe the economic evaluation methods which tend to be adopted
        for the economic evaluation of MAS procedures;

(b)   to assess whether the methods that generally prevail are adequate given
        the characteristics of the technologies under evaluation and, if not, where
        methodological developments are most urgently required; and

(c)   to identify any novel methods which have been developed in this area.

2.5.2 Review methods

The focus of the review is economic evaluations of MAS undertaken since 1985. The choice of year in which to begin the review is based on the fact that the majority of developments of MAS have taken place over the last decade.

Economic evaluation is taken to mean a full analysis, as defined by Drummond *et al* [1987], which includes cost-effectiveness analysis, cost-utility analysis, cost-minimisation analysis and cost-benefit analysis. This definition excludes studies which have looked in detail at costs and outcomes within a cost-consequence framework because they do not incorporate decision rules to assess technical or

allocative efficiency. MAS is defined as in Section 2.2.2 above; namely, an application of endoscopic or percutaneous therapy.

The sample of studies in the review has been identified using the following strategy.

(a)   A search of publications on the Medline and Health Planning and Administration on-line databases (US National Library of Medicine) as at July 1996. The search employed a mixture of index terms and free text searches, details of which are provided in Appendix 2.1. The abstracts of each article retrieved were read and the full articles of apparently appropriate studies acquired.

(b)   Several comprehensive reviews of MAS have been published [Hirsch and Hailey, 1992; Banta, 1991; Banta, 1993A; Hirsch, 1994; Pearson, 1994]. The references cited in each were manually searched and full copies of articles apparently fulfilling the inclusion criteria for the review acquired.

(c)   Each article was read carefully and a decision taken as to whether or not it fulfilled the inclusion criteria for the review.

2.5.3 Results of the review

The search of the on-line databases generated 181 possible full economic evaluations of MAS applications. On the basis of the abstracts, 155 articles were excluded for the following reasons: 23 were cost or cost-consequence analyses; 2 looked at benefits only and not at costs; 40 were reviews rather than evaluation studies; and 90 did not relate to MAS. The manual search of the references of the MAS reviews identified a further three likely full economic evaluations. Hence, full articles were acquired for 29 studies.

After studying the full articles, a further 13 studies were excluded because they were only cost or cost-consequence analyses (9), were not evaluation studies (2) or because they were not evaluations of MAS (2). This left a total of 16 studies for detailed review, the results of which are described in Table 2.1.

One observation from the review is the paucity of rigorous economic evaluations of MAS applications. If predictions that up to 80% of surgery will be undertaken on a minimal access basis by 2000 are accurate, it should be seen as a major research priority to increase the number of economic assessments in this area. Contrasting with the dearth of full economic evaluations is the large number of cost analyses and cost-consequence analyses. By systematically describing the various clinical benefits and disbenefits, as well as the costs, of MAS alongside an appropriate comparator, some cost-consequence analyses would have provided useful information to health service decision makers. However, the majority of these partial economic evaluations fail to provide any clear indication of the value for money offered by MAS: they are either clinical evaluations which include a modest amount of cost (or charge) data almost as an after-thought, or they are detailed costings without data on outcomes.

2.6   Issues of method in the economic evaluation of MAS

A range of methodological issues exist in relation to the economic evaluation of MAS, many of which are brought out by assessment of the economic characteristics of MAS and by the review of published studies. The key issues are discussed below, some of which provide the focus of subsequent chapters.

2.6.1 Sources of data

Good quality data on the resource and non-resources consequences of interventions are a requirement whatever the technology being evaluated. However, there are some particular issues of method associated with measuring the effects of surgical procedures in general, and MAS in particular, that deserve specific attention.

The randomised controlled trial. Amongst the hierarchy of methods of *clinical* evaluation, it is widely accepted that the properly designed RCT, with an adequate sample size, is the preferred method for assessing the health effects of therapeutic interventions [Pocock, 1983]. In addition, the RCT is increasingly seen as a vehicle for the collection of resource use data [Drummond and Stoddart, 1984; Eisenberg *et al*, 1989; Drummond and Davies, 1991; Drummond

Table 2.1   Results of the review of economic evaluations of minimal access surgery

Study	Type of analysis	Diagnosis	Interventions evaluated	Sample size	Clinical data sources	Primary outcome measure	Costs included	Period of analysis	Indirect costs considered?	Key results	Assessment of generalisability?
England et al [1987]	CEA	Renal artery stenosis	Percutaneous transluminal angioplasty versus renal artery bypass surgery	52	Retrospective comparative study and published data, synthesised into model	Life-years	Hospital charges	50 years	No	Angioplasty is more cost-effective in patients with fibro-muscular disease and those less than 50 years of age	Yes. Clinical results from a range of centres considered. No analysis of costs
Labelle et al [1987]	CEA	Renal and ureteric calculi	Extracorporeal shock wave lithotripsy [ESWL], percutaneous ultrasonic lithotripsy [PUL] and standard surgery	N\A	Model incorporating data from various sources	Disability days	Hospital costs	5 years	No	Surgery is dominated by both ESWL and PUL. At 500 procedures per year and over, ESWL dominates PUL	Limited. Sensitivity analysis looks at plausible variation in some parameters likely to be influenced by location and context
Laffel et al [1987]	CEA	Acute myocardial infarction	Various strategies for coronary thrombolysis/ reperfusion therapy involving intravenous and intracoronary streptokinase, and percutaneous transluminal coronary angioplasty	N/A	Model incorporating data from a range of sources	Number of survivors	Hospital costs	1 year	No	Strategies involving intravenous administration of thrombolytic agents are consistently more cost-effective than those using intracoronary thrombolytic agents or primary PTCA	Limited. Sensitivity analysis looks at plausable variation in some parameters likely to be influenced by location and context

38

Table 2.1   Results of the review of economic evaluations of minimal access surgery (continued)

Study	Type of analysis	Diagnosis	Interventions evaluated	Sample size	Clinical data sources	Primary outcome measure	Cost included	Period of analysis	Indirect costs considered?	Key results	Assessment of generalisability?
Nishioka and Richter [1987]	CEA	Gastro-intestinal haemorr-hage	Alternative forms of endoscopic therapy [eg. laser] and usual management	N/A	Model incorporating probability data from a range of sources including randomised trials	Efficacy [haemo-stasis]	Hospital	Unclear - short-term	No	Nd:YAG laser can be cost-effective if used on patients at high risk of recurrent bleeding	None
Carlsson et al [1989]	CEA	Renal calculi	Percutaneous nephrolithtomy [PCN], extracorporeal shockwave lithotripsy [ESWL]	600	Cohort study	Succes-sfully treated patients	Hospital	1 month [ESWL] or 3 months [PCN]	No	ESWL is more cost-effective with stones up to diameter of 2-3cm; PCN may be more cost-effective for larger stones	None except a descriptive discussion of some recent technological developments
Wong et al [1990]	CUA	Chronic stable angina	Conservative therapy, percutaneous transluminal coronary angioplasty [PTCA], coronary artery bypass surgery [CABG]	N/A	Model synthesising data from a range of sources	Quality-adjusted life years	Hospital charges	Lifetime	No	Both PTCA and CABG have a cost-effective role depending on the nature of the disease	None

Table 2.1   Results of the review of economic evaluations of minimal access surgery (continued)

Study	Type of analysis	Diagnosis	Interventions evaluated	Sample size	Clinical data sources	Primary outcome measure	Cost included	Period of analysis	Indirect costs considered?	Key results	Assessment of generalisability?
Mays [1991]	CEA	Renal and ureteric stones	Extracorporeal shock wave lithotripsy [ESWL] and percutaneous nethrolithotomy [PCN]	1934	Cohort study	Stone-free rates	Hospital [capital, operating]	2 years	No	PCN likely to be less costly but this conclusion is sensitive to throughput with ESWL. PCN more cost-effective means of rendering a patient free of stones	Yes. A range of scenarios reflecting practice in other centres incorporated into the sensitivity analysis
Carlsson et al [1992]	CEA	Kidney stones	Extracorporeal shock wave lithotripsy [ESWL]; percutaneous nephrolithotomy [PCN]	55	Randomised controlled trial	Success rates	Hospital	12 months	No	ESWL less costly than PCN because of high ESWL throughput in centre and shorter length of stay with ESWL. Cost per treatment success lower with ESWL	Very limited at a descriptive level in discussion
Bass et al [1993]	CUA	Symptomatic gallstone disease	Open [OC] and laparoscopic cholecystectomy [LC]	N/A	Model synthesising data from a range of sources	Quality-adjusted life months	Hospital charges	5 years	No	For the primary analysis, LC dominated OC for women and men of all ages. However, results are sensitive to choice of preoperative procedures and level of professional charges	Very limited at a descriptive level in discussion

40

Table 2.1   Results of the review of economic evaluations of minimal access surgery (continued)

Study	Type of analysis	Diagnosis	Interventions evaluated	Sample size	Clinical data sources	Primary outcome measure	Costs included	Period of analysis	Indirect costs considered?	Key results	Assessment of generalisability?
Bisonni et al [1993]	CUA	Benign prostatic hyperplasia	Transurthral prostatectomy [TURP] and transurethral dilatation of the prostatic urethra [TUDP]	N/A	Model incorporating data from various sources	Quality-adjusted life years	Procedure costs based on hospital charges	Patients' lifetimes	No	TUDP dominated TURP despite using the 'best case' scenario for TURP	Limited. Sensitivity analysis looks at plausible variation in some parameters likely to be influenced by location and context
Cohen et al [1994]	CUA	Symptomatic single-vessel coronary disease	Percutaneous transluminal coronary angioplasty [PTCA], primary coronary stenting and PTCA followed by coronary stenting for symptomatic restensois [secondary stenting]	N/A	Model incorporating data from a range of sources	Quality-adjusted life years	Hospital costs	Patients' lifetimes	No	Coronary stenting is cost adding but more effective than PTCA, and is reasonably cost-effective. Secondary stenting is less cost-effective than primary stenting	Limited. Sensitivity analysis looks at plausible variation in some parameters likely to be influenced by location and context
Cook et al [1994]	CUA	Gallstone disease	Open [OC], laparoscopic cholecystectomy [LC], extracorporeal shock wave lithotripsy [ESWL]	653	Cohort study of ESWL versus LC; historical OC controls	Quality-adjusted life years	Hospital, patient and indirect costs	15 years	Yes - results presented with and without	LC found to be generally superior to OC and ESWL. However, this conclusion is sensitive to perspective and inclusion of indirect costs	Limited. Sensitivity analysis looks at plausible variation in some parameters likely to be influenced by location and context

Table 2.1   Results of the review of economic evaluations of minimal access surgery (continued)

Study	Type of analysis	Diagnosis	Interventions evaluated	Sample size	Clinical data sources	Primary outcome measure	Costs included	Period of analysis	Indirect costs considered?	Key results	Assessment of generalisability?
de Boer et al [1995]	CEA	Acute myocaridal infarction	Primary coronary angioplasty versus thrombolysis	301	Randomised controlled trial	Event-free survival after one year	Hospital costs	1 year	No	Angioplasty found to be more expensive, with an incremental ratio of $3,010 per additional event-free survivor	Very limited. Some variation in unit costs explored
Gray et al [1995]	CEA	Ectopic pregnancy	Laparoscopic versus laparotomic treatment	109	Randomised controlled trial	Successful treatment (ie. elimination of trophoblastic activity)	Hospital costs	3 months	No	The two treatments found to be equally successful, but laparoscopy is less costly	Very limited. Some descriptive discussion
Vroegin-deweij et al [1995]	CEA	Femoro-popliteal artery disease	Endartectomy (EA) versus balloon angioplasty (BA)	103	Retrospective	Patency month	Hospital costs	Median: 27 mths (EA), 23 mths (BA)	No	EA more costly, but with higher patency rates. *Average* cost per patency month ratios were not significantly different	None
Hunink et al [1995]	CUA	Femoro-popliteal artery disease	Percutaneous transluminal angioplasty (PTA) versus bypass surgery (BS)	N/A	Model synthesising data from a range of sources	Quality-adjusted life years	Hospital costs	Patients' lifetimes	No	PTA more cost-effective in patients with disabling claudication and stenoses or occlusion; BS more cost-effective in critical ischaemia and occlusions	None

42

1995]. Of the 16 economic evaluations of MAS procedures reviewed here, 3 were based largely on data collected in a RCT (Table 2.1). Chapter 3 of this thesis uses RCT data as a starting point for the economic evaluation of AH versus TCRE.

The primacy of RCTs follows from the link between random allocation - usually of patients to alternative therapeutic interventions - and *internal validity* [Jaeschke and Sackett, 1989]. That is, randomisation facilitates control over potential but unknown confounding variables, which may influence the outcome of interest, because they should be symmetrically allocated across the arms of the trial, thus allowing differences in effectiveness and resource use to be safely attributed to the only factor known to vary systematically between treatments - the intervention itself.

The RCT is accepted as being the gold standard for the evaluation of drug therapies and such methods are required in many countries to establish the clinical efficacy and safety of new compounds prior to licensing [Dukes, 1986]. The process of 'blinding' both clinicians and patients to the exact details of the intervention to which patients have been randomised - thus minimising the chance of measurement bias - ensures that the advantages of the true experiment are maximised.

Although RCTs have successfully been carried out to evaluate surgery [Miller *et al*, 1989], their use in this area remains rare compared to trials employing historical controls [Sacks *et al*, 1982]. As an example of a set of new surgical procedures, some forms of MAS have entered into widespread clinical use with very little evaluation by RCT. For example, despite a survey indicating that 58% of British surgeons thought a trial was required and 45% were willing to be involved [McMahon *et al*, 1992], only six RCTs have been located in the published literature focusing on the clinical or patient-based outcomes of laparoscopic cholecystectomy [Barkun *et al*, 1992; Trondsen *et al*, 1993;

McMahon *et al*, 1994; Berggren *et al*, 1994; Makinen and Nordback, 1995;

Dauleh *et al*, 1995]. Furthermore, five of these were small, randomising

between 24 and 78 patients. It should be noted, however, that evaluation of

MAS by RCT is more common in some specialties other than in general surgery.

In gynaecology, for example, a range of MAS procedures has now been subject

to trial-based evaluation [Berget *et al*, 1987; Tulandi, 1986; Lundorff *et al*, 1991;

Summitt *et al*, 1992; Dwyer *et al*, 1993; Pinion *et al*, 1994].

A number of potential difficulties in using RCTs to assess the effects of surgical

technologies have been identified in the literature, many of which are particularly

relevant to new MAS procedures. The first problem relates to that fact that is it

usually felt to be virtually impossible to incorporate a placebo control and double

blinding into a RCT of a surgical procedure and that this represents a movement

away from the "gold standard for proper evaluation" [Stirrat *et al*, 1992, p.81].

However, there has been successful use of placebo controls and double blinding

in evaluations of MAS procedures; for example, a recent RCT compared

transurethral microwave treatment for benign prostatic hypertrophy with sham

treatment using double blinding [Bdesha *et al*, 1993]. It may also be possible to

blind the *evaluator* of the technology if they are independent of the clinical team.

Moreover, even if placebo control and blinding are not feasible, this might only

preclude the successful completion of an *explanatory* trial in surgery which

would seek to test a biological hypothesis under optimal conditions to establish

the relative *efficacy* of procedures [Schwartz and Lellouch, 1967]. If an RCT is

to be undertaken to inform *clinical policy*, however, it is more likely to exhibit a

pragmatic rather than an explanatory design, whereby alternative procedures are

compared under conditions which would normally apply in routine practice, in

order to identify the apparently more *effective* therapy [MacRae, 1989].

The second problem relates to the 'learning curve' in MAS, which has some

important implications for the use of RCTs [van der Linden, 1980; MacRae,

1989; Stirrat *et al*, 1992]. One implication is that it is likely that RCTs will be

undertaken by clinical enthusiasts: clinicians with high levels of skill, often based

in medical schools. Indeed, some of the few RCTs evaluating MAS techniques

specify a minimum level of experience for the clinicians carrying out treatment. Given that the surgeon (or endoscopist) is a fundamental part in the therapeutic technology of MAS (which is less the case with drug therapy), their characteristics directly influence resource and non-resource consequences. As has been noted above, studies have now highlighted the link between the experience of the therapist and clinical results. Hence it may be doubted that RCTs of MAS can ever be truly pragmatic if the trial therapists have characteristics which are dissimilar to those of the practitioners who will undertake the procedure following widespread diffusion. The *generalisability* or *external validity* of RCTs in this area may, therefore, be limited.

It is possible, however, to design RCTs of MAS which attempt to maximise external validity. One approach would be to undertake the trial in a number of centres which more fully represent the heterogeneity of clinical practice. The use of a multicentre design may bring additional advantages, moreover, by avoiding the tendency to under-power trials which frequently prevents the detection of relatively effective therapies [Frieman *et al*, 1978]. As either an addition or an alternative to increasing the number of centres, the generalisability of a RCT may be increased by ensuring that therapists within the trial occupy the full spectrum of clinical experience likely to be observed in routine practice: senior registrars as well as consultants, for example. If it is not feasible for each clinician taking part in the trial to undertake each of the therapies being evaluated, there should be an effort to ensure that the skill/experience mix of the clinicians is equivalent in the two arms of the trial. If this is not feasible - clinicians are often further up a learning curve for an established procedure than they are for a new one - statistical methods can be used to adjust results for asymmetry in clinicians' experience between the arms of the trial [MRC Health Services and Public Health Research Board, 1993].

A third problem likely to be encountered when designing trials of MAS procedures is the variability that exists between centres in certain elements of the procedure. For example, some centres use a laser as an integral part of laparoscopic cholecystectomy, while others use diathermy [Voyles *et al*, 1990;

Hunter, 1991]. The problem can be overcome, as long as the variability in methods is known when the trial is being designed. In which case, if the alternative methods are considered important enough and sufficient patient numbers and resources are available, patients can be randomised to two or more forms of MAS as well as to conventional surgery. A similar problem exists concerning the potential speed of change of MAS technologies. If a trial is designed to evaluate a new MAS procedure and, midway through the study, the clinicians in the study feel that the procedure should alter in some way, the value of the results may be limited. The extent to which this is a significant problem will inevitably depend on the importance of the technological change and whether the trial can recruit sufficient additional patients to assess the new form of the technology within the same study.

A fourth and related problem is that trial protocols may impose atypical patterns of care upon unrepresentative samples of patients, which makes the resource use and outcomes observed difficult to generalise to routine clinical practice. This problem can be minimised by a making the RCT as pragmatic as possible: undertaking it in as many centres as is feasible; including the vast majority of patients to encompass the heterogeneity of the presenting condition; and developing a protocol which does not impose new or additional forms of clinical practice to the norm in a given centre [Simon *et al*, 1995].

A fifth difficulty likely to be encountered when RCTs are used to evaluate MAS procedures concerns patient recruitment. In view of the most apparent characteristics of MAS relating to the reduction in trauma and length of convalescence, potential study patients may not be attracted to the idea of entering a clinical trial which involves a chance of being randomised to an alternative form of therapy: the patient quite simply may not be indifferent between the options under evaluation. It has been argued that patient recruitment to RCTs designed to evaluate laparoscopic cholecystectomy has suffered in this way: "Well intentioned and well-designed protocols have drowned in the tidal wave of optimism among patients and physicians about the benefits of the laparoscopic approach" [Cotton, 1992, p1626]. Clinicians, too,

may be reluctant to randomise patients into trials; it has been shown that a common reason for this reluctance is the clinician's fear of admitting ignorance to patients [Taylor, Margolese and Soskolne, 1984].

Various suggestions have been made as to how recruitment problems related to RCTs could be overcome within RCTs. These include the education of the public regarding the social and moral arguments in favour of "doing their bit" for improving health [Baum, 1993], the provision of incentives to doctors to recruit patients such as emphasising the 'training' effect of trials [Macintyre, 1991], the concept of randomised consent designs for clinical trials where patients' preferences about treatment options are taken into account at randomisation [Brewin and Bradley, 1989; Zelen, 1990] and the use of pre-randomisation [Chang *et al*, 1990]. Despite calls for more RCTs in health care technology assessment [Advisory Group on Health Technology Assessment, 1992], it is likely that the use of trials to evaluate MAS procedures will be hampered by patient recruitment problems.

It is likely that an important focus for future evaluation of MAS procedures will be on the long-term implications of these technologies, and a sixth problem with RCTs lies here. In order to have detected the alleged shortcomings of TURP (see Section 2.3.1) as part of a RCT, such a study would have to have been both large in terms of patient numbers - to identify the suggested higher incidence amongst TURP patients of stricture, re-operation and death, which are rare events - and based on a very long-term period of follow up. The cost of running a RCT and the possible difficulties in recruiting patients and clinicians, however, may limit the scope for designing sufficiently large trials. Although trial patients can be 'flagged' for some key clinical events such as death or the development of cancer, the cost and logistical difficulties of organising trial-specific data collection over many years usually results in RCTs having a relatively short-term focus. For this reason, it has been argued that "case-control studies . . . constitute, at least for now, the only feasible method for studying rare and late adverse effects of drugs and other technologies. Randomised trials are simply

neither big enough nor long enough to detect important but rare adverse events"
[Jaeschke and Sackett, 1989, p.509].

A seventh problem concerns the sample size of RCTs used to collect resource
use data. Most trials are powered to detect differences in clinical end-points
such as survival rates. If data from RCTs are to be used to determine the cost-
effectiveness of MAS procedures, sample sizes will need to be sufficient to get
reasonably tight estimates of levels of resource use. This will require more
general evidence on the variability in these types of data amongst patients
[Drummond and O'Brien, 1993].

A final problem with RCTs relates to concerns that have been raised about the
ethics of RCTs in general [Passamani, 1991]. When they are used as a vehicle
for the collection of resource use data, additional concerns may be identified.
Perhaps the most intractable of these occurs when the *clinical* uncertainty
concerning two technologies has been resolved by a RCT, but it would be
necessary to continue randomising patients to achieve a sample size sufficient
for good estimates of resource use. Should randomisation continue despite the
fact that *clinical uncertainty* may have been removed [Buxton and Sculpher,
1994]? As such, this is a specific example of the general conflict between
ethics at an individual level and those at a societal level that frequently arises
within discussions of rationing and economic evaluation [Mooney and McGuire,
1988].

It is probably true that the relative rarity of the RCT in health care evaluation
[Fletcher and Fletcher, 1979] is more the result of the difficulties and cost of
undertaking them, and of a failure of clinicians to accept genuine clinical
uncertainty, than of a widespread concern about their potential limitations
regarding evaluative methods; this applies in particular to surgery. Although it is
widely accepted that, even with the sort of drawbacks described above, the
appropriately undertaken RCT offers the most rigorous source of evaluative data,
in many contexts such trials simply are not possible due to such factors as poor
patient recruitment, ethical concerns, the nature of the technology and the

urgency of getting results. Inevitably, therefore, the surgical evaluation literature is full of non-experimental evaluative designs and these have contributed to clinical policy decisions in the area. When assessing how to evaluate emerging MAS procedures, these methods need to be considered.

Observational studies. The non-experimental study can serve a useful purpose in establishing the *potential* effectiveness of new technologies and the cost-effectiveness of *undertaking* a RCT. In clinical areas where the prognosis is poor and no effective intervention has previously existed, positive evidence from such analyses may be sufficiently credible to prompt the diffusion of a new technology [Advisory Group on Health Technology Assessment, 1992]. Indeed, it is possible to identify by this route technologies which are self-evidently effective, such as dialysis in terminal renal failure [Guyatt *et al*, 1986]. Furthermore, the non-experimental study can serve a purpose in selecting those new technologies with a likelihood of improving outcomes, and these can then be subject to more rigorous evaluation [Sculpher *et al*, forthcoming]. Such studies can also provide data to plan certain aspects of RCTs, such the calculation of sample sizes, and inclusion and exclusion criteria [Roos, 1989].

Further advantages of non-experimental designs have been identified [Sechrest and Hannah, 1990]:

(a)   data are often readily available, and analyses can be undertaken quickly and at limited cost;

(b)   the results of non-experimental studies may have high levels of external validity because the unrealistic conditions of the experimental design are avoided; and

(c)   quasi-experiments (ie. when two or more groups are compared retrospectively or prospectively but are not generated by random allocation) may be inherent in administrative arrangements which routinely generate data.

Sechrest and Hannah argue that, faced with these possible advantages but also with the clear problems of non-experimental data, the researcher can either

ignore the information provided or seek to improve the weaknesses in design and to strengthen interpretations of the findings of non-experimental studies. Some researchers have accepted the latter approach and have sought ways to strengthen causal interpretations of non-experimental data. These include a consideration of a greater role for meta-analysis [Cordray, 1990] and *ex post* adjustment of the results of non-experimental studies based on the expected direction and size of their biases [Miller *et al*, 1989]. There have also been calls for greater use of Bayesianism in evaluative research, whereby the value of analysis can be independent of study design [Berry, 1993].

With a focus on determining the cost-effectiveness of a technology, Sculpher *et al* [forthcoming] suggest an iterative approach to economic evaluation, where early-stage assessment would involve the synthesis of available data to clarify the key uncertainties in resource use and outcomes to plug into subsequent designs. In certain circumstances, early-stage modelling is sufficient to identify the cost-effectiveness of a technology (eg. screening for diabetic retinopathy [Dasbach *et al*, 1990]).

2.6.2 Indirect costs

A methodological controversy which impinges on cost analyses involving MAS procedures is whether or not to include indirect costs. Of the 16 studies in the review of published economic evaluations in Section 2.5 , only one considered indirect costs. However, the fact that patients return to their normal activities more speedily with many types of MAS than with conventional surgery is fairly clear and, as discussed above, this has become a clear *clinical* 'selling point' for these sorts of technology [Stoker *et al*, 1994]. Whether these consequences should also be *valued in monetary terms* is less clear. Amongst health economists, certainly in the UK, there is little consensus about the incorporation of indirect cost savings into the calculus of economic evaluation [Drummond *et al*, 1993A]. Koopmanschap and Rutten [1994] found that decisions about whether or not to include indirect costs in an economic evaluation can have a major impact on results. This has also been concluded in the specific area of the economic evaluation of MAS: Cook *et al's* [1994] comparison of alternative

treatments for gallstone disease, found that the inclusion of indirect costs altered the results of the analysis markedly.

A number of concerns exist as regards the validity of valuing these effects as a form of cost. Firstly, invariably those studies which do value these consequences concentrate on earlier return to *work* and value this time using gross pay rates, although the time of patients not actively engaged in the labour market is also of value and a focus solely on the value of forgone working time may bias service provision against individuals not in employment [Ratcliffe, 1995].

A second concern relates to the implications of an effectively permanent pool of unemployed on such costs: if an individual is required to take time off work due to illness, the employer can, in principle, dip into this pool for temporary labour with little effect on total production. Models have, however, recently been developed to address this problem [Koopmanschap and van Ineveld, 1992].

A further concern about the appropriateness of incorporating indirect costs into economic evaluation is the risk of double-counting. If the outcomes of treatment have been valued in non-monetary terms on the benefit side of a CUA and if those outcomes include increased mobility and quicker return to usual activities, it would surely not be appropriate to value these particular outcomes again in monetary terms to incorporate on the cost side of the analysis.

2.6.3 Assessing generalisability

Section 2.6.1 argues that trial data on the resource and non-resource consequences of MAS may have limited external validity. Indeed, even if the economic analysis is based on observational, rather than experimental, data, there may be limits to external validity if the data are taken from only a small number of centres at a particular point in time. It is, therefore, important for an economic assessment of this form of technology to explore the generalisability of its results to data from other sources. The review of published economic evaluations of MAS in Section 2.5 above indicated that four studies considered

the generalisability of their results at a descriptive level within their discussions. For example, Carlsson *et al* [1989] discussed the development of anaesthesia-free extracorporeal shock wave lithotripsy in out-patients, indicating that this would probably reduce costs below that of the base-case analysis, but suggesting that the effect of the approach on effectiveness was as yet unknown.

Seven studies in the review provided some assessment of generalisability in the form of their sensitivity analyses. However, this typically involved some plausible one-way variation in parameters that were considered likely to vary by location or context. Using this approach, it is possible to state under which conditions the base-case conclusions remain robust. For example, Laffel *et al* [1987] looked at how variation in the time until presentation to the health care system influences the relative cost-effectiveness of alternative forms of thrombolysis and primary PTCA, and found that this was a crucial parameter. A more sophisticated approach was followed by Mays [1991] who assessed the impact on the relative cost-effectiveness of ESWL and percutaneous nephrolithotomy of alternative scenarios about such variables as the utilisation of capital equipment.

Most studies using this approach, however, simply varied assumptions within the analysis over a plausible range. Only one study undertook a rigorous assessment of generalisability by incorporating alternative sources on data into the analysis and assessing the extent to which the base-case conclusions were robust to the results of clinical practice in other centres. England *et al* [1987] generated base-case conclusions from a simulation model based on data taken from one centre, and then undertook a sensitivity analysis by incorporating data from several other sites.

If the resource and non-resource consequences of MAS procedures are likely to vary considerably by treatment location and over time, it is important for the conclusions of economic evaluations which are based on data (whether observational or experimental) taken from one or a small number of centres, to

be established as robust to other sources of data. Chapter 6 of this thesis explores these methods further in the context of the economic comparison of AH and TCRE.

2.6.4 Benefit measurement

In recent years, there has been a move to complement conventional clinical outcome measures, such as survival rates and technical success rates, with measures of patients' own assessment of their health status [McDowell and Newell, 1987; Streiner and Norman, 1989]; and some specific work has been undertaken in the area of surgery [Cleary *et al*, 1991A]. The terminology related to this broad group of outcome measures is variable; but the terms 'quality of life' and 'health-related quality of life [HRQL]' appear frequently in the literature [Mosteller and Falotico-Taylor, 1989; Spilker, 1990; Patrick and Erickson, 1993]. Furthermore, various types of measure exist under these sorts of general heading, including specific and generic instruments, profiles and indices [Guyatt and Jaeschke, 1990]. The increased use of these instruments is evident in evaluations of MAS procedures. The Nottingham Health Profile (NHP), a generic measure of perceived health [Hunt *et al*, 1986], has been used to evaluate PTCA [Henderson, 1989], TURP [Doll *et al*, 1993], percutaneous nephrolithotomy [Mays, 1991] and laparoscopic cholecystectomy [Barkun *et al*, 1992]. Another generic measure, the SF-36 [Ware, 1993], has been used in evaluations of laparoscopic cholecystectomy [McMahon *et al*, 1994].

Specific instruments have also been used, such as the General Health Questionnaire to assess psychological well-being in the evaluation of TCRE [Dwyer *et al*, 1993], and the McGill Pain Questionnaire and Hospital Anxiety and Depression Scale (HADS) in the assessment of laparoscopic cholecystectomy [Barkun *et al*, 1992; McMahon *et al*, 1994]. The use of HRQL measures provides a broader assessment of the outcomes of MAS than do clinical measures in isolation.

In order to assess the *cost-effectiveness* of MAS procedures, however, it is usually necessary to acquire outcome data which exhibit certain characteristics

[Cairns, 1996]. Unless the MAS procedure *dominates* its conventional comparator(s) - that is, it is at least as effective on all measures of outcome and is less costly - an assessment of cost-effectiveness will require the incremental cost of the more expensive technology to be related to its improved outcomes. The conventional way of doing this in economic evaluation is through the use of some form of incremental cost to effect ratio. In cost-effectiveness analysis (CEA), effectiveness is measured on a uni-dimensional scale in natural units which should embrace the key differences in outcomes between the technologies under comparison. In the review of published economic evaluations in Section 2.5, the majority of studies (10/16) were CEAs, with cost-effectiveness expressed using measures such as life-years gained [England *et al*, 1987], disability days averted [Labelle *et al*, 1987] and additional survivors [Laffel *et al*, 1987].

A shortcoming of CEA is that it limits the assessment of relative cost-effectiveness to a specific clinical area or programme. For example, in the review, Carlsson *et al's* [1989] comparison of ESWL and percutaneous nephrolithtomy for the removal of renal stones expressed cost-effectiveness by relating cost to treatment success rates, where the latter was defined as being stone at follow-up. Whilst this information may be useful to decision makers looking to allocate resources in this particular clinical area, it is unlikely that condition-specific outcome measures of this type would assist in cross-programme resource allocation because, outside the area of renal stones, this measure of outcome would have little meaning.

There has, therefore, been an increasing use of cost-utility analysis (CUA) to facilitate, in principle, a system-wide assessment of relative cost-effectiveness. With this type of evaluation, technologies are compared using generic measures of effectiveness or benefit which, although uni-dimensional, can more fully embrace the various outcomes of programmes. The most frequently used measure of benefit in CUA is the quality-adjusted life-year (QALY), which is a measure of the additional life-years generated by technologies, weighted by the perceived quality of those years [Loomes and McKenzie 1989; Mooney and

Olsen, 1991]. In principle, the relative cost-effectiveness of all uses of health care resources can be compared using the QALY.

Of the 16 published economic evaluations of MAS procedures reviewed in Section 2.5, six were CUAs. This proportion compares with 10/93 identified in a review of all economic evaluations published in 1992 [Briggs and Sculpher, 1994]. One factor explaining the greater use of CUA in the economic evaluation of MAS is the multi-dimensionality of outcomes of these technologies when they are compared to standard interventions. As discussed above, the comparison of MAS and open surgery frequently highlights a trade-off in outcomes: open surgery often has a higher chance of a good longer-term outcome, but the process of treatment and the shorter-term outcomes (eg. duration of convalescence) tend to be superior with MAS. It is not easy to express this trade-off using standard CEA, but CUA can encompass differences in the various dimensions of HRQL as well as in mortality. Chapter 5 of this thesis considers the methods and practice of CUA in detail, in relation to the surgical treatment of menorrhagia.

It is not clear, however, that the standard QALY fully reflects individuals' preferences about the trade-offs inherent in MAS compared to standard therapy. These individuals could be members of the public, as their values would be used to inform the process of allocating society's scarce resources. It could be argued, however, that patients' preferences over combinations of outcomes should have an important role in the decision making process. Chapter 4 considers the treatment-related preferences of a sample of 221 women with menorrhagia and their implications for economic evaluation.

It has recently been argued that, even when patients' values are used to measure QALYs, the assumptions underling the standard approach to QALY estimation will not generate a benefit measure which is necessarily consistent with their preferences [Gafni, 1989; Richardson *et al*, 1996]. Benefit measures have been suggested for use in CUA which seek to avoid some of these assumptions. These include the risk-adjusted QALY [Pliskin *et al*, 1980] and

various forms of healthy-years equivalent (HYE) [Mehrez and Gafni, 1989; Cook *et al*, 1994; Gafni *et al*, 1995], but there are issues about their practicality in applied evaluations. Chapter 7 focuses on alternative benefit measures to the standard QALY in CUA and the implications of one - the *ex ante* HYE - for the economic analysis of AH versus TCRE.

2.6.5 Preference-based treatment allocation

Patients' preferences about MAS and conventional therapy can be incorporated into economic evaluation in a more direct manner than through conventional CUA. It is usually the logic of economic evaluation that a single preferred intervention is identified from the two or more being compared and that this will be used in future clinical practice. Occasionally, however, this 'all or nothing' policy approach is abandoned when clinical heterogeneity exists and a treatment that is found to be cost-effective for a particular sub-group of patients is not considered to represent good value for money for all patients.

However, patients differ in ways other than their clinical characteristics, in particular in terms of their preferences. Given that the benefits patients derive from many therapies are often tied to their preferences about likely outcomes [Henshaw *et al*, 1993], it can be argued that the question that an economic evaluation should explicitly ask is what are the incremental costs and benefits of having all the comparators under evaluation available from which patients can choose. This approach seems particularly appropriate to the evaluation of MAS procedures, where patients may exhibit particularly strong preferences about the trade-offs in outcomes between interventions.

Although looking at the valuation of the outcomes provided by the average (or the median) patient or member of public might result in the MAS procedure being considered relatively cost-effective, many patients will have quite different preferences. An economic evaluation could then assess whether the incremental cost of retaining the conventional therapy, as well as having the MAS procedure available, could be justified by the additional benefits enjoyed by this 'atypical' group of patients. Chapter 8 considers preference-based treatment allocation in

detail and explores whether this form of management is potentially cost-effective in the area of surgical treatment of menorrhagia.

2.7   Conclusions

If the recent Cushieri Report on MAS is accurate, 70% to 80% of surgical procedures will be undertaken using endoscopic and percutaneous methods within 10 years [Cuschieri, 1993]. Although a large number of specific technologies falls into the MAS category, it is valuable to consider their common actual and potential characteristics, and the methodological issues that surround their evaluation.

It is argued in this chapter that there is no automatic link between the advent and diffusion of new MAS techniques and an increase in the cost-effectiveness of health care provision. It is, therefore, crucial to undertake appropriate economic evaluation of MAS procedures, and it is likely that this clinical area will absorb an increasing proportion of research and development resources in coming years. For this reason it is important to develop further the methods of economic evaluation and, in particular, to understand the methodological issues and difficulties which arise in assessing the cost-effectiveness of this type of technology.

A range of methodological issues relating to the economic analysis of MAS have been discussed in this chapter. Some of these have been discussed widely in the methods literature as they arise in relation to the economic evaluation of a range of health care technologies; others have not been fully considered in the literature. The key issues are:

(a)   the RCT has a number of limitations as a source of resource and non-resource consequence data for economic analysis;

(b)   generalisability needs to be assessed carefully in economic analysis of MAS;

(c)   the multi-dimensionality of the outcomes of MAS may mean that
        conventional CEA is not suitable for the evaluation of these technologies;

(d)   CUA will provide a means of handling multi-dimensional outcome and
        generate information to assist in cross-programme resource allocation, but
        the standard QALY may not reflect individuals' preferences about trade-
        offs between outcomes; and

(e)   if there is heterogeneity in patients' treatment-related preferences
        regarding MAS, it may be appropriate to assess the potential cost-
        effectiveness of preference-driven treatment allocation.

The remaining chapters of this thesis look in detail at particular approaches to
the economic evaluation of MAS and, in particular, these key methodological
issues.

Appendix 2.1   Details of the strategy used to search Medline and

Health Planning and Administration databases

Strategy to identify full economic evaluations

The following index terms were used:

• cost-benefit analysis.

The following free text searches were made on titles:

• economic evaluation;

• cost effectiveness;

• cost utility;

• cost minimisation/minimization;

• cost-benefit.

Strategy to identify applications of minimal access surgery.

The following generic term was used:

• minimal invasive surgery.

The following index terms were used related to types of technology:

• endoscopy;

• laparoscopy;

• angioplasty;

• hysteroscopy;

• gastroscopy;

• sigmoidoscopy;

• thoracoscopy;

• angioscopy.

The following free text searches were done on titles:

• minimal(ly) invasive;

• minimal access.

Documents were not considered if they were not English language; were animal studies; or were classified as the following document types: review, comment, letter or editorial.

Chapter 3

AH versus TCRE: A Cost-Effectiveness Analysis Alongside a Randomised Controlled Trial

3.1 Introduction

Using the case-study of abdominal hysterectomy (AH) versus transcervical resection of the endometrium (TCRE), this chapter details an economic evaluation of MAS versus conventional surgery alongside a RCT. The analysis provides an opportunity to explore the methodological implications of running an economic evaluation alongside a trial which has been designed largely from the perspective of clinical evaluation. The chapter provides a starting point for the methodological developments introduced in subsequent chapters, by highlighting the areas of uncertainty likely to exist in trial-based analysis.

The economic evaluation is undertaken alongside a RCT which was set up largely in response to the perceived need for applications of MAS to be subjected to

experimental evaluation [Stirrat *et al*, 1990].   The trial took place in Bristol and included 200 women and, based on an initial follow-up of four months, the clinical evaluation showed that TCRE reduces post-operative morbidity, permitting patients to return more quickly to their usual activities, but is not so effective as AH at relieving menstrual symptoms [Dwyer *et al*, 1993]. Furthermore, a sub-group of resection patients requires additional surgical treatment.

The trial provided an ideal opportunity to begin the economic assessment of the two treatments. Other studies have considered the relative costs of AH and TCRE [Rutherford and Glass, 1990; Manyonda and Varma, 1991; Gannon *et al*, 1991; Vilos *et al*, 1996], but none of these incorporated the cost of complications or of re-treatment of resection failures. Furthermore, no study attempted to relate the differential cost of the procedures to differential effectiveness to assess their relative cost-effectiveness.

Based on data collected in the Bristol RCT, this chapter reports the resource costs of the two procedures from a health service perspective. In addition, alternative measures of outcome are presented including menstrual symptoms, patient satisfaction and descriptive health-related quality of life (HRQL). Relative cost-effectiveness is expressed using patient satisfaction to define treatment success. Two points of follow-up were used in the analysis: four months post-operation and an average of 2.2 years post-operation.

3.2   Methods

3.2.1 Design of the trial
The clinical trial was a parallel group, randomised controlled trial based at the gynaecology department of a teaching hospital. A total of 200 women were recruited to the trial between January 1990 and June 1991. Required sample size was estimated with reference to the anticipated rate of patient satisfaction with the two treatments. After withdrawals, 97 patients underwent AH and 99

TCRE. A detailed description of the design of the clinical trial has been reported elsewhere [Dwyer et al, 1993].

3.2.2 Short-term analysis: four months post-operation

Resource use. Data were collected prospectively on the health service resources used in the treatment of each woman in the trial during the period from randomisation to four months after her operation. These resource use data can be described under seven general headings: pre-operative, operative, post-operative, in-patient hotel, complications, re-treatment and general practice. Pre-operative resources included pathology tests, drugs and blood transfusions. Operative resources were the staff present, equipment, consumables, drugs, theatre time and histology tests. Anaesthetic resource use was estimated from a separate study based at the hospital in 1990 which looked in detail at a sample of five TCREs and 13 AHs [Dr D. Wilkins, personal communication]. Operative resource use data also included those resources consumed as a result of related procedures undertaken, as part of the theatre episode, in addition to the main procedure of AH or TCRE. In-patient hotel resource use was represented by the number of in-patient nights each woman spent in hospital. If patients spent any time in the intensive therapy unit (ITU), the number of hours was recorded. Immediate post-operative resource use included tests, drugs and blood transfusions.

Data were collected on the additional resource use resulting from all operative and post-operative complications, together with late complications requiring readmission to hospital: additional operative procedures, drugs, tests, and in-patient stay. Re-treatment resource use was that resulting from repeat TCRE or AH on women for whom initial TCRE was considered a failure at four months follow-up. Data were also collected on general practice resources: four months after their operation, each woman was asked how many visits they had made to their GP since their operation.

Valuing resource use. Health service resources used by the patients in the two arms of the trial until four months follow-up are valued in 1994 prices. Where available, the unit costs used are those generated routinely for Bristol General Hospital, where the patients were treated. Where such unit costs are not available, a variety of published sources is used.

All consumables used during the treatment episode have been costed using market prices including value added tax (VAT). The unit cost of tests undertaken during the treatment episode are those estimated routinely by the hospital. Drugs have been costed based on British National Formulary prices plus VAT; to represent Pharmacy Department overhead costs an uprate of 20% is added, which was the department's administrative costs as a percentage of its total drug budget in 1990-1. AH patients who had both ovaries removed during their operation were advised to use hormone replacement therapy (HRT): an estimate of the present value of the cost of using HRT for four months is calculated. The cost of blood transfusions is estimated using the contract price of the red cell (operative additive solution) product for the South Western Regional Health Authority plus an allowance for hospital handling costs, based on the estimate of each patient receiving four units of blood.

Operative staff time is costed using the mid-range salary for each relevant member of staff, uprated by 11% to allow for employers' costs. Although surgical procedures were largely undertaken by a clinical research fellow, the time of surgeons, as well as that of anaesthetists, is costed using the consultant pay scale.

The cost of surgical equipment which is *not* considered usually available in a gynaecological theatre (non-routine) is estimated separately. For this equipment, per patient costs are based on the estimation of an annual equivalent cost using a discount rate of 6% and assuming the equipment has a useful life of five years. Annual rates of utilisation have been estimated assuming that all surgically-treated menorrhagia would be treated using the modality for which the

equipment is relevant. For all other routine theatre equipment, costs are taken from the Bevan Report [Bevan, 1989], in the form of a cost per minute of theatre time, adjusted to a 1994 price base.

An anaesthetic cost per minute is used based on the separate study at the hospital, with the costs adjusted to a 1994 price base. Theatre overhead costs have been taken from the Bevan Report [Bevan, 1989], in the form of a cost per minute of theatre time, also adjusted to a 1994 price base, and these are taken to represent the cost of all resources not directly allocated to a given patient. Total operative costs per patient are, therefore, made up of a fixed cost per patient (non-routine equipment, some staff, drugs and consumables) and a variable cost per minute of time in theatre (routine equipment, most staff, anaesthetic and overheads).

The ward cost of a patient's hospital stay is estimated by multiplying the number of nights each patient was in hospital by the daily ward unit cost, where the latter is the sum of the average hospital ward and general service cost per in-patient day estimated for the hospital's 1991-2 cost returns. The cost of any patient stay in ITU is estimated by multiplying the relevant length of stay by the average daily cost of ITU for a spontaneously breathing patient estimated by Ridley *et al* [1991] and adjusted to a 1994 price base.

The cost of re-treatment of patients for whom TCRE was considered a failure at four months follow-up, with either repeat resection or AH, is estimated using mean pre-operative resource use for the relevant procedure together with the actual length of hospital stay for each re-treated patient.

The cost of a visit to a general practitioner is estimated assuming a 8.25 minute consultation [DHSS, 1987]. The cost of a GP's time is based on average net remuneration, allowing for superannuation and national insurance and assuming a 38 hour week and a 46 week year [DHSS, 1987]. An uprate of 66% on the cost

of GP time is used to represent direct and indirect practice expenses [Special Medical Development Project Team, 1990].

Outcome measures. The RCT collected a range of outcome measures at or before four months follow-up. These included pain during the week following surgery; improvement in menstrual symptoms; time away from usual activities; and rates of satisfaction with treatment where women indicated whether they were 'very satisfied', 'quite satisfied', 'not very satisfied' or 'very dissatisfied'.

3.2.3 Longer-term analysis: two years post-operation

The clinical and economic characteristics of these two surgical treatments are time dependent: resource use and effects are not confined to the peri-operative period and convalescence but extend, in principle, throughout a woman's life. The costs and consequences of TCRE, in particular, are likely to vary markedly with time because the treatment has been shown to fail in a proportion of women, who often require further surgery. Both treatments may have longer-term consequences that impact on patient benefits and health service resource use. It is crucial, therefore, to follow-up those women in the Bristol RCT for as long a period as is feasible.

Postal questionnaire. For the longer-term follow-up, all women in the trial were sent a questionnaire in the post which sought information on menstrual symptoms, HRQL, satisfaction with treatment for their menstrual problems and health service resource consumption. The questionnaires were sent out in two batches: the first 100 patients receiving surgery were sent a questionnaire in February 1993; the remaining patients received one in October 1993. Women were asked to complete the questionnaire and to return it to the study team in a pre-paid envelope.

Menstrual symptoms. The first section of the questionnaire sought information on women's menstrual symptoms. Women randomised to TCRE were asked about continued menstrual bleeding. All women were asked about

pain, either associated with their period (for TCRE patients) or similar to that they previously experienced before a period (for AH patients). For those women experiencing pain, they were asked to rate its severity compared to the pain they experienced before surgery. Similar information was also requested on premenstrual symptoms described as bloating, breast tenderness or headache. Women were also asked whether they had taken time off work during the last year as a result of menstrual problems.

Health-related quality of life. In the second part of the questionnaire, women were asked to complete the Short Form 36 (SF36), a generic measure of subjective health in the form of a profile with 36 items [Ware *et al*,1993]. The instrument has eight multi-item dimensions covering physical functioning, social functioning, role limitations due to physical problems, role limitations due to emotional problems, mental health, vitality, pain and general health perception. Based on responses, a scoring algorithm produces a scale from 0 (for poor health) to 100 (good health) for each dimension. Although the SF36 was developed in the USA, it has been shown to be internally consistent, valid and acceptable to individuals in the UK [Brazier *et al*,1992; Jenkinson *et al*,1993; Garratt *et al*,1993]; it has also been used with women with menorrhagia, and been found to be acceptable and sensitive to change [Jenkinson *et al*, 1994; Coulter *et al*, 1994A; Ruta *et al*, 1995]. In the analysis reported here, if a woman failed to respond to one or more items on a particular dimension, her overall score for that dimension is taken as missing data.

Resource use and costs. The final section of the questionnaire asked women whether they had received any hospital treatment since their original surgery. The purpose of this section was twofold. Firstly, to establish which women, having originally been randomised to TCRE, had received additional surgical treatment for their menstrual problems since the initial four month follow-up. Information on the re-treatment of some women not responding to the postal questionnaire was available from other sources, such as hospital

records. Life table methods are used to estimate the cumulative probability of re-treatment by two years after initial surgery [Kahn and Sempos 1989].

The second purpose of the section of the questionnaire on resource use was to assess what other hospital resources women had used as a result of their menstrual problems. Women were asked to detail any visits to hospital they had made since their four-month follow-up. Based on the judgement of a clinical collaborator (Dr Nuala Dwyer), resource use likely to be related to menstrual problems was identified. A separate question in this section of the questionnaire asked women whether they were using HRT.

The total treatment costs estimated on the basis of 4 months follow-up are revised using the resource use data collected within the questionnaire. For resource use related to re-treatment, the unit cost of a repeat procedure is based on the average cost estimated in the earlier analysis and includes pre-operative, operative, post-operative and ward costs, and costs related to average complications. The cost of one out-patient visit per additional surgical procedure is added to this. Re-treatment costs relating to women who did not respond to the postal questionnaire and whose re-treatment details were not known from other sources are counted as missing data.

The unit costs associated with other related resource use, together with that of hormone replacement therapy, are based on those used in the short-term analysis and those routinely estimated by United Bristol Health Trust. Unit costs additional to those in the four-month analysis are the cost of an ultrasound (£50), the cost of a urinary dynamics test (£200) and the cost of an out-patient visit (£80). It is established practice in economic evaluation to discount resource costs that occur in future years [Drummond *et al*, 1987]. Therefore, all re-treatment and other related costs are discounted according to the time that elapsed since randomisation using a 6% annual discount rate [HM Treasury 1991]. All costs are expressed at a 1994 price level.

3.2.4 Cost-effectiveness analysis

Based on the two episodes of data collection, the Bristol RCT provided data, four months and 2.2 years after surgery, on the health service resource costs of the two procedures and a range of outcome measures consisting of symptoms, satisfaction rates and descriptive measures of HRQL. Given the aim to use these data to inform resource allocation decisions in this area, costs are related to outcomes to assess the relative value for money of the two surgical therapies.

Cost-effectiveness analysis (CEA) relates the differential cost of two or more treatments to their differential effectiveness. In order to achieve this, however, it is necessary for effectiveness to be measured on a uni-dimensional scale which adequately reflects the overall impact of the treatment options on patient benefits [Drummond *et al*, 1987]. As surgical treatments for menorrhagia, AH and TCRE have a range of different effects. A woman's satisfaction with treatment, however, can be considered a relevant uni-dimensional measure of outcome in this context. Although the measurement of patient satisfaction is relatively under-developed [Fitzpatrick, 1993], a woman's reaction to this sort of question is likely to be influenced by her overall experience with the treatment. In rating her satisfaction, she needs to trade-off any elements of the process and short- and longer-term outcomes of treatment which she did not like against those she did, and hence arrive at a global assessment of the intervention from her perspective. The fact that satisfaction rates reflect patients' perceptions, and also that they have more intuitive meaning than some other measures of effectiveness used in CEA, are the main reasons for their use here.

For the CEA, it is necessary to dichotomise the four-level satisfaction response offered to women. If a women indicated in the questionnaire that she was 'very satisfied' or 'quite satisfied', her treatment has been defined as successful. In contrast, if she answered that she was 'not very satisfied' or 'very dissatisfied', her treatment is taken as being unsuccessful. Using this measure of treatment success, the differential cost of AH and TCRE are related to their differential success rates using cost and outcome data up to four months and 2.2 years.

Table 3.1   **Short-term analysis: resource use from randomisation until 4 months follow-up. Figures are numbers of patients (%) or means (95% CI)**

Resource Use	TCRE (n = 99)	AH (n = 97)
Pre-operative		
Full blood count (FBC) (n(%))	99 (100)	97 (100)
Blood grouped and saved (n(%))	99 (100)	97 (100)
Blood transfusion and additional FBC (n(%))	4 (4)	4 (4)
Antibiotics (n(%))	99 (100)	97 (100)
Operative		
Theatre time mins (mean 95% CI)	51.2 (49.2 to 53.2)	2.9 (60.6 to 65.2)
Histology test (n(%))	99 (100)	97 (100)
Post-operative		
FBC (n(%))	0 (0)	97 (100)
Blood transfusion and additional FBC (n(%))	2 (2)	2(2)
Urea and electrolytes (n(%))	3 (3)	0 (0)
Analgesia (n(%))	49 (49)	97 (100)
HRT as a result of removal of both ovaries (n(%))	0 (0)	5 (5)
In-patient nights (mean (95% CI))	2.1 (1.9 to 2.2)	6.4 (6.16 to 6.58)
GP visits (mean (95% CI))	0.7 (0.5 to 0.9)	2.2 (1.8 to 2.6)

3.2.5 Statistical analysis

All analyses are undertaken on an intention-to-treat basis. That is, despite the fact that a proportion of the group which was initially randomised to TCRE subsequently underwent a hysterectomy, all comparisons leave women in the

group to which they were originally randomised. Variation in differential costs is shown using 95% confidence intervals. Unless otherwise stated, all statistical hypothesis tests are Wilcoxon Rank Sum tests and a 5% significance level is used [Armitage and Berry 1989].

3.3   Results

3.3.1 Short-term analysis: four months post-operation

Resource use. Table 3.1 details the main elements of resource use until four months follow-up under the headings introduced in Section 3.2.1 above. The staff present in theatre for TCRE were a trolley nurse (it has been assumed they were present for the first 10 minutes only), a surgeon, an anaesthetist, an anaesthetic nurse and a circulating nurse; for AH a senior house officer and an instrument nurse were also present. Non-routine operative equipment included a camera, colour monitor, xenon light source, resectoscope and telescope. The consumables used during TCRE were a loop (which was changed, on average, once every seven patients), irrigation tubing, catheter and gloves; a mean of 7.4 litres (95% CI 6.84 to 7.96) of glycine irrigation fluid was also used.

Due to the fact that TCRE does not ensure infertility, 10 patients (10%) in this group requested laparoscopic sterilisation at the time of their resection. Additional operating time is reflected in the average times detailed in Table 3.1; Filshie clips were also required. For AH, the list of consumables used included suturing materials, blades, dressings, catheter and gloves. Anaesthetic resource use, for both patient groups, included a range of drugs and gases.

Table 3.2 provides details of the operative, post operative and late complications occurring in study patients where additional resource use resulted. The Table also describes those additional resources. A total of 12 patients in the TCRE arm of the trial required re-treatment after their initial operation, due either to dissatisfaction with results at four months follow-up (11 patients) or to abnormal histology (one patient).

Table 3.2   Short-term analysis: operative, post-operative and late complications resulting from procedures and details of additional resource use*

Complication	TCRE	AH	Additional resource use
Operative complicationst			
Uterine perforation	4	0	Antibiotics 3-9 days
Fluid overload	1	0	3 hours in ITU; urea and electrolytes x 4; diuretics 1 day
Haemorrhage (>500ml)	1	3	Transfusion
Bladder perforation	0	1	Supra-pubic catheter; antibiotics 8 days
Inadvertent removal of ovaries	0	1	HRT until menopause
Post-operative complicationst			
Urinary retention	0	4	'Foley' catheter and bag
Urinary infection	0	12	Antibiotics 5 days
Pelvic haematoma	1	8	Antibiotics 5 days; pelvic ultrasound
Pelvic infection	2	1	Antibiotics 5 days
Haemoglobin < 10	0	10	Ferrous sulphate 3 months
Wound haematoma	0	1	Return to theatre for 20 minutes
Suspected haemorrhage	0	1	Return to theatre for laparotomy for 40 minutes; transfusion
Late complicationstt			
Intractible diarrhoea	0	1	Readmitted for 10 nights; antibiotics 10 days
Pelvic haematoma	0	1	Readmitted for 3 nights; antibiotics 5 days
Bleeding plus pelvic haematoma	0	1	Readmitted for 1 night
Suspected deep vein thrombosis	0	1	Readmitted for 2 nights
Vaginal bleeding	1	0	Readmitted for 2 nights
Retreatment**			
Using endometrial resection	7	0	Additional procedure; 2 nights in hospital
Using hysterectomy	5	0	Additional procedure: 5 or 6 nights in hospital

* Only complications resulting in additional resource use detailed
t Any extra stay in hospital included in results for overall stay
tt Complications occurring up to four months after operation
** As a result of assessment at four months

Table 3.3   Short-term analysis: costs per patient until four months after surgery*

Cost component	TCRE		AH	
	Mean (SD)	Median (range)	Mean (SD)	Median (range)
Pre-operative	16.63 (24.65)	11.61 (11.61-136.16)	16.74 (24.89)	11.61 (11.61-136.11)
Operative	234.32 (29.59)	231.29 (194.54-329.13)	292.00 (34.79)	283.00 (236.57-422.28)
Post-operative	2.78 (17.61)	0.13 (0.13-124.68)	6.81 (21.69)	2.58 (2.58-127.13)
Ward	250.56 (78.98)	239.56 (119.78-958.24)	763.27 (121.68)	718.68 (598.90-1197.80)
Complications	5.09 (28.24)	0.00 (0.00-239.56)	30.11 (136.56)	0.00 (0.00-1226.69)
Re-treatment	85.30 (247.71)	0.00 (0.00-1029.30)	0.00 (-)	0.00 (0.00-0.00)
General practice	4.05 (5.72)	0.00 (0.00-22.05)	11.83 (10.82)	11.02 (0.00-55.12)
TOTAL	593.65 (276.89)	487.36 (445.83-1792.48)	1123.21 (209.92)	1053.39 (876.15-2414.27)

*   The summations of the mean and of the median columns do not equate exactly with the stated totals because data were not available for a small number of patients for certain cost components, and total costs are based only on those patients for whom data were available on *all* cost components (92 TCRE, 93 AH)

73

Table 3.4   Short-term analysis: outcome measures collected until four months after surgery*

Outcome measure	AH (n = 97)	TCRE (n = 99)
Post-operative pain		
Number (%) experiencing no pain at[t]:		
Day 1	5 (5)	6 (6)
Day 2	4 (4)	19 (19)
Day 3	5 (5)	48 (48)
Day 4	6 (6)	61 (62)
Day 5	9 (9)	74 (75)
Day 6	13 (13)	78 (79)
Day 7	14 (14)	81 (82)
Pre-menstrual symptoms		
Number of women whose pre-menstrual		
symptoms improved after surgery[‡]:		
Dysmenorrhoea	81/86	53/85
Bloating	35/77	14/78
Breast tenderness	21/50	13/60
Mood	16/41	11/36
Post-operative recovery		
Median (range) time off work (weeks)	11 (1-24)	2 (<1-8)
Median (range) time until return to daily activities (weeks)	4 (1-10)	1 (<1-8)
Satisfaction with treatment		
Number (%):		
Very satisfied	77 (80)	66 (67)
Quite satisfied	13 (13)	17 (17)
Not very satisfied	6 (6)	13 (13)
Very dissatisfied	0 (0)	2 (2)

*	See Dwyer *et al* [1993] for further details
t	Women judged subjectively
‡	Denominator is the number of women complaining on symptom before surgery

Resource costs per patient. Appendix 3.1 details the unit costs used to value the resource use of patients in the trial. Table 3.3 shows the resource costs per patient, for each major cost component and in total, up to four months follow-up. The total mean cost of surgically treating menorrhagia using TCRE is

statistically significantly lower than that using AH: the mean difference is -
£529.66 (95% CI £458 to £601).

Outcome measures. The range of outcome measures collected over the
period between surgery and four months follow-up is detailed in Table 3.4. The
table shows a clear trade-off in these outcomes: the proportion of women
experiencing pain during the week following surgery is higher in the AH arm and
women undergoing TCRE return to work and to their usual activities earlier;
however, the improvement in pre-menstrual symptoms is more pronounced in
the AH group as is women's satisfaction with treatment. If satisfaction rates are
used to define a treatment success as defined in Section 3.2.4, 90% of women
randomised to AH were treated successfully compared to 83% randomised to
TCRE (Chi squared, p = 0.04).

3.3.2 Longer-term analysis: two years post-operation

As part of the longer-term assessment, questionnaires were posted to all 196
women who underwent surgery in the trial. Three questionnaires were returned
incomplete, these women having moved and provided no forwarding address. A
total of 155 women returned a completed questionnaire - a response rate of
79%. Of these, 82 and 73 had been randomised to TCRE and to AH,
respectively. The mean period of time that had elapsed since surgery for those
women responding to the questionnaire was 2.8 years (range 1.8 to 3.8). The
mean follow-up period overall, including that for women not responding, was 2.2
years (range 0.3 years to 3.8 years).

Menstrual symptoms. Table 3.5 presents details of women's responses
to questions on menstrual symptoms. Of women randomised to resection, and
including those women who eventually had an AH in the denominator, 70% were
still experiencing some bleeding. If the denominator is adjusted to remove those
women who had a AH (ie. departing from the intention-to-treat analysis), this
rate increases to 87% (54/62). More women who had been randomised to TCRE
still experienced pain, but fewer of these women considered their pre-menstrual

Table 3.5   Longer-term analysis: menstrual symptoms reported in the postal
                 questionnaire. Unless stated, details are the number of women reporting
                 the symptom over the number of women answering the question, with
                 percentages in parentheses

Symptom	TCRE	AH
Women experiencing bleeding	54/77 (70)*	-
Of those experiencing bleeding:		
Mean (SD) days bleeding per month	4.5 (3.1)	-
Women reporting clots	11/54 (20)	-
Women reporting flooding episodes	6/54 (11)	-
Women experiencing pain	46/72 (64)	13/70 (19)
Compared to before operation for those		
Not as bad	29/46 (63)	10/13 (77)
About the same	7/46 (15)	2/13 (15)
Worse	10/46 (22)	1/13 (8)
Pre-menstrual symptoms compared to before surgery		
Not as bad	30/71 (42)	52/68 (76)
About the same	28/71 (39)	13/68 (19)
Worse	13/71 (18)	3/68 (4)
Women taking time off work due to menstrual problems	18/80 (23)	3/71 (4)

*   Denominator includes 15 women who responded to the questionnaire and who had
         undergone AH after their initial TCRE

symptoms to have improved since surgery. More women who had been
randomised to TCRE had taken time off work during the previous year as a result
of menstrual problems.

Health-related quality of life. Table 3.6 provides details of the SF36 scores
which have been calculated on the basis of women's responses to the
questionnaire. The scores are calculated on a scale from 0 (poor health) to 100
(good health), and results are presented for each of the eight dimensions, by
treatment group. Details of means with standard deviations and medians with
ranges are provided, as well as mean differences with 95% confidence intervals.
The mean score on seven of the eight dimensions favours women randomised to
AH. With a mean difference of nearly 10 points, the most marked difference

Table 3.6   Longer-term analysis: SF-36 scores from postal questionnaire responses.
Mean differences relate to AH scores minus TCRE scores.

| SF-36 dimension | TCRE | | AH | | |
	Mean (SD)	Median (Range)	Mean (SD)	Median (Range)	Mean differences (95% CI)
Physical function	89.6 (17.8)	95 (5-100)	92.1 (14.3)	100 (10-100)	2.5 (-2.82 to 7.82)
Role limitations (physical)	82.7 (33.1)	100 (0-100)	82.0 (33.6)	100 (0-100)	-0.7 (-11.5 to 10.1)
Role limitations (emotion)	80.0 (31.6)	100 (0-100)	86.2 (29.8)	100 (0-100)	6.2 (-3.8 to 16.2)
Social function	84.4 (22.5)	100 (0-100)	90.4 (16.1)	100 (22-100)	6.0 (-0.5 to 12.5)
Mental health	74.1 (15.7)	76 (36-100)	76.4 (17.1)	82 (28-100)	2.3 (-3.1 to 7.7)
Energy	60.8 (20.6)	60 (5-100)	62.3 (21.7)	70 (10-95)	1.5 (-5.4 to 8.4)
Pain	73.2 (26.2)	77.8 (11-100)	83.1 (22.9)	100 (22-100)	9.9 (1.9 to 17.9)
Health perceptions	74.4 (21.7)	77 (10-100)	79.7 (20.2)	87 (15-100)	5.3 (-1.5 to 12.1)

between the groups in favour of AH patients is in bodily pain (p = 0.01). Quite large differences in mean scores are evident in role limitations due to emotional problems (6 points; p = 0.12), social functioning (6 points; p = 0.12) and general health perceptions (5 points; p = 0.09), but none of these reflects an overall statistically significant difference.

Satisfaction. There are clear differences in women's levels of satisfaction with treatment between the two treatment groups. Amongst women randomised to TCRE, 46 (57%) were 'very satisfied' with treatment, 18 (22%) were 'quite satisfied', 12 (15%) were 'not very satisfied' and 5 (6%) were 'very dissatisfied'. The relevant numbers for women randomised to AH were 61 (85%), 8 (11%), 2 (3%) and 1 (1%), respectively. If these rates are used to define a treatment success, 79% of the women randomised to TCRE and 96% of women randomised to AH were treated successfully (Chi squared, p = 0.002) at longer-term follow-up.

In response to the question about whether they would have the same operation if they had the choice again, 19 (24%) of women randomised to TCRE said they
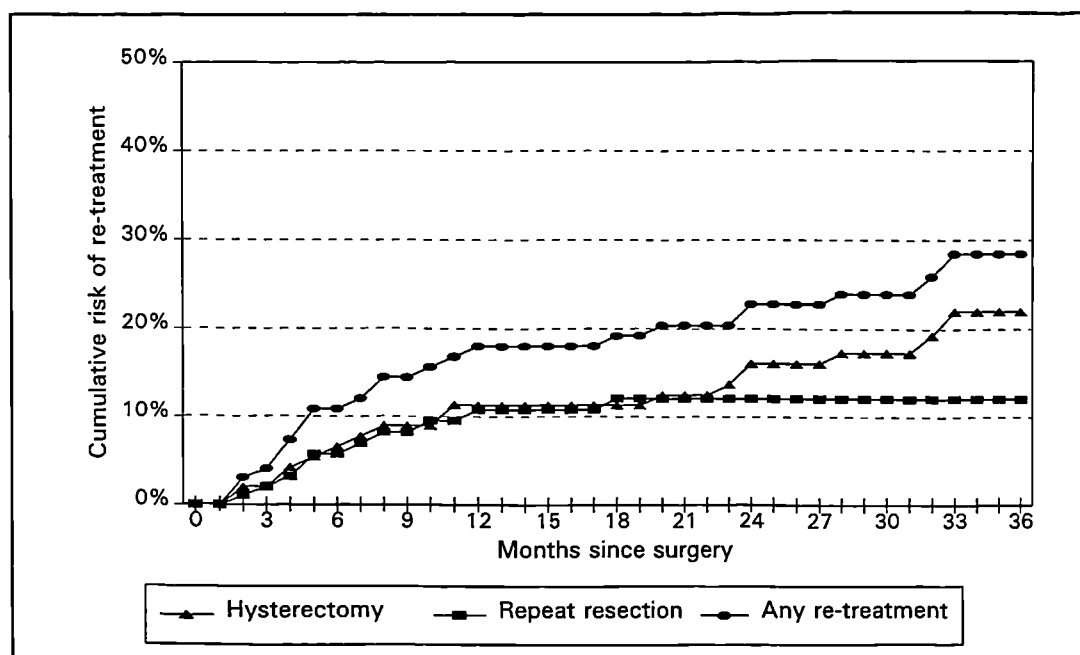
Figure 3.1   The cumulative probability of treatment failure for women randomised to TCRE based on longer-term follow-up.

would not, compared to 4 (6%) of patients randomised to AH (Chi squared, p = 0.002).

Resource use and costs. A number of women originally randomised to TCRE underwent a repeat resection and/or a hysterectomy due to failure of the initial procedure to ameliorate symptoms. A total of three women were known to have received re-treatment between four months follow-up and the current period of follow-up, despite not responding to the postal questionnaire. Including re-treatments by four months follow-up, a total of 10 patients received a repeat TCRE and 18 underwent a hysterectomy. Four women received both a repeat TCRE and a hysterectomy.

Figure 3.1 shows 'failure curves' for women randomised to TCRE. Three curves are shown relating to the cumulative probability of a repeat TCRE, of a hysterectomy and of any form of re-treatment, respectively. These estimates are based on life table analysis and the assumption that, if a woman did not

respond to the postal questionnaire and re-treatment details were not available through some other source, her data are censored at four months (the point of the last follow-up). On this basis, by two years after initial surgery, the cumulative probability of a repeat TCRE was 12%, that of a hysterectomy was 16% and that of any form of re-treatment was 23%. If it is assumed that details of any re-treatment would have been available through other sources for all women regardless of whether or not they had completed the questionnaire, these probabilities fall slightly to 11%, 14% and 20%, respectively.

Of the women responding to the questionnaire, 9 (11%) who had been randomised to TCRE were judged to have used hospital resources which were related to their menstrual problems after four months follow-up. This resource use is in addition to the re-treatment detailed above and mostly took the form of between one and four out-patient visits, but two women underwent an ultrasound scan and one women had a hysteroscopy. Five (6%) of the women randomised to AH who responded to the questionnaire reported other hospital resource use which was judged to be related to the treatment of their menstrual problems, which consisted of 1 or 2 out-patient visits, 2 ultrasound scans for one woman and a urinary dynamics test for another. Of the women randomised to TCRE and responding to the item in the questionnaire, 9 (11%) had been taking hormone replacement therapy for a mean duration of 11 months (SD = 8); 12 (17%) of the AH group had been using such therapy for a mean duration of 13 months (SD = 10).

Table 3.7 presents the results of the revised cost analysis which adds the cost of re-treatment and of other related resource use between four months and an average of 2.2 years follow-up to the short-term analysis. The table treats data relating to women not responding to the postal questionnaire, and whose re-treatment and other related resource use details were not available to the study through other routes, as missing. Compared to the results of the cost analysis at four months, the 'cost gap' between TCRE and AH has closed: as a percentage of the mean total cost of AH, the mean total cost of TCRE was 53% at four

Table 3.7   Revised analysis of the health service costs per patient in the trial based
               on a mean overall follow-up of 2.2 years.

	TCRE		AH	
	Mean (SD)	Median (Range)	Mean (SD)	Median (Range)
Initial surgery	513 (102)	483 (446-1256)	1123 (210)	1053 (876-2414)
Re-treatment costs to 4 months	85 (248)	0 (0-1029)	0	0
Re-treatment costs after 4 months	185 (396)	0 (0-1157)	0	0
Other related resource use after 4 months:				
- hospital costs	11 (40)	0 (0-284)	10 (43)	0 (0-238)
- HRT	3 (8)	0 (0-26)	5 (12)	0 (0-31)
Total costs*	790 (493)	523 (446-2148)	1110 (168)	1053 (876-2036)

*   Only relates to those patients for whom there is full follow-up (78 patients in the TCRE
       group and 70 in the AH group), so the summations of the mean and median columns do
       not equate exactly with the stated totals.

months; it has increased to 71% at a mean overall follow-up of 2.2 years.

However, total cost differences between the two groups remain statistically

significant (difference £320, 95% CI 198-442). Table 3.7 shows that the mean

total cost of AH at 2.2 years is actually slightly lower than that at four months.

This is due to the fact that there is full follow-up on only 70 of the 97 women

randomised to AH, so the two estimates of total cost are based on somewhat

different samples. As a result, there may be a risk of bias in the estimate of

longer-term cost, but the extent of this is likely to be small given the limited

resource use in the AH group between four months and 2 years.

Figure 3.2 shows the distribution of total costs by randomised group. The figure

emphasises the greater spread in total costs amongst women randomised to

TCRE, although, as Table 3.7 shows, the mean and median of the distribution

are lower for women in that group.
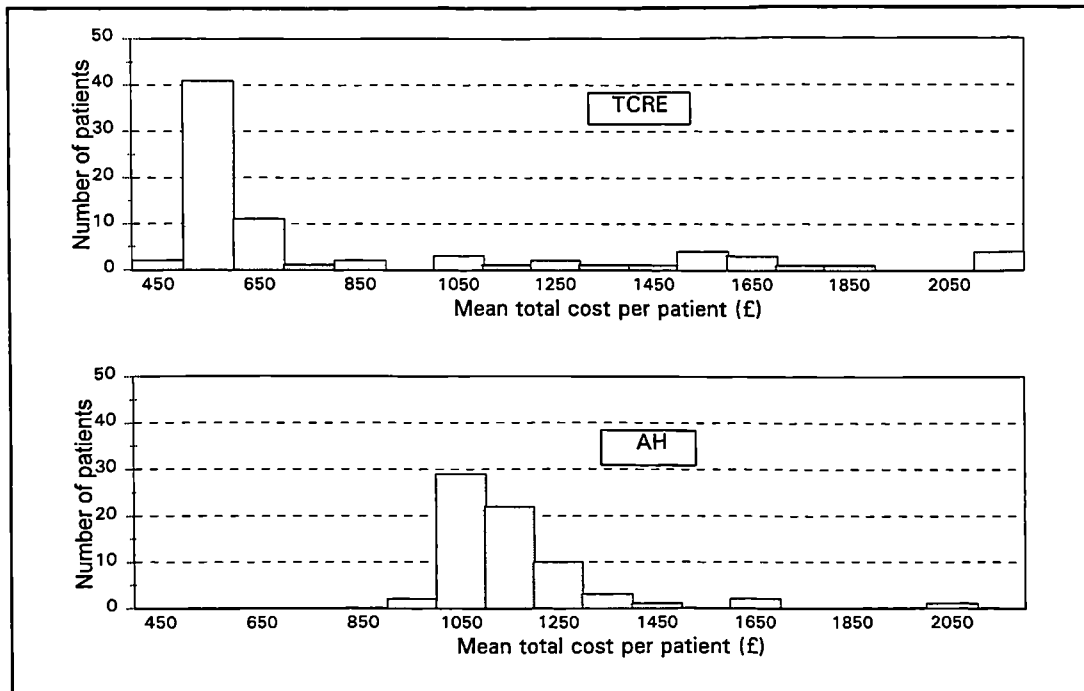
Figure 3.2   Distribution of mean total costs over 2.2 years. Includes only those
women for whom there is full follow-up data.

3.3.3 Cost-effectiveness analysis

Based on both four months and 2.2 years follow-up within the trial, AH appears
to be more expensive in terms of health service costs. The outcomes of the two
treatments during these two periods of follow-up, however, are equivocal: the
shorter-term outcomes tend to favour TCRE, but the longer-term impact on
menstrual symptoms and HRQL generally favours AH. If success is defined as a
woman indicating that she is 'very satisfied' or 'quite satisfied' with treatment,
AH can be considered statistically significantly more effective than TCRE, both
at four months (p = 0.04) and 2.2 years (0.002). Table 3.8 relates the additional
cost of AH to its additional effectiveness using CEA. At 4 months after surgery,
the difference in treatment success on the basis of women's satisfaction was
only 7% which results in an incremental cost of AH per additional success of
£7,557. After 2.2 years, the difference in success rates increased to 17% as
satisfaction with AH increased and that with TCRE declined. Therefore, as the
difference in costs falls as more women who were randomised to TCRE

Table 3.8   Cost-effectiveness relating costs to treatment success defined in terms of
              patient satisfaction

	Period of follow-up		
	4 months (All patients)	4 months (Full follow-up)	2.2 years
Total mean cost			
AH	£1123	£1095	£1110
TCRE	£594	£519	£790
Difference	£529	£576	£320
Success rates*			
AH	90%	96%	96%
TCRE	83%	84%	79%
Difference	7%	12%	17%
Incremental cost per additional success	£7557	£4800	£1882

*   Defined as women reporting that she is 'very satisfied' or 'quite satisfied' with the results of her
       treatment

required additional treatment, the incremental cost of AH per additional

treatment success declines to £1,882.

The apparent reduction in the total cost of AH at 2.2 years compared to four

months is a result of the fact that 24 (25%) women randomised to AH were lost

to longer-term follow-up, and most of these women consumed no resources

between four months and 2.2 years. Given the skewed nature of the cost

distribution shown in Figure 3.2, it only requires a small number of the higher

cost women at four months to be lost to follow-up for the mean cost to decline

at 2.2 years. To allow for the effect of loss to follow-up, Table 3.8 also reports

the costs and effectiveness of treatment at four months solely for those patients

followed up for the full 2.2 years. On this basis, the incremental cost of AH per

additional treatment success is £4,800.

3.4   Discussion

3.4.1 Resource costs

This study has found that, over a mean period of 2.2 years, the total mean cost of surgically treating menorrhagia using TCRE is statistically significantly lower than that using AH. At four months follow-up, the mean total cost of TCRE is 53% that of AH. This cost difference is largely explained by two components of health service resource use which are higher amongst AH patients: operative resources and hotel resources. As regards operative resources, the fixed cost (£24.89 versus £17.25), the variable cost per minute (£3.09 versus £2.63) and the mean operating time (62.9 minutes versus 51.2 minutes) are all higher for AH patients.

The most important difference in health service resource use between the two groups, however, is length of in-patient stay. The statistically significantly longer mean length of stay of AH patients in the trial (a mean of 6.37 nights versus 2.09) leads to the mean *difference* in ward costs accounting for 96% of the mean *difference* in total cost at four months.

The short-term cost analysis incorporates the cost implications of the finding in the randomised trial that, at four months follow-up, 11 (12%) patients who initially received TCRE were defined as having had treatment failure and required further surgery. For one additional patient AH was required because of an abnormal histological finding. Of these 12 patients, seven received repeat TCRE and five underwent hysterectomy.

The key longer-term issue in relation to resource costs is whether this re-treatment rate increases as follow-up continues. By 2.2 years after surgery, and adjusting for differential periods of follow-up, the cumulative probability of a woman undergoing a repeat TCRE was 12% and that of her having a hysterectomy was 16%. In terms of costs, this has the effect of closing the gap between TCRE and AH. Whereas the mean total cost of TCRE was 53% that of

AH at four months follow-up, that percentage had increased to 71% at a mean overall follow-up of 2.2 years.

A crucial issue as regards cost is whether this cost gap will narrow further - perhaps to the point that it closes entirely - as more time elapses. It is probably the case that re-treatment rates for women randomised to TCRE will not increase markedly over the next few years, as most women, for whom initial TCRE failed, have probably re-presented for further assessment and treatment. In the longer term, some other items of health service resource use may have an influence on the differential cost of the two treatments. Most women who undergo AH will not require cervical cytology every five years until the age of 64 years; and the resource cost of treating cervical and uterine cancer will also be avoided in a sub-group of these patients. However, hysterectomy has been associated with increased risk of premature ovarian failure [Siddle *et al*, 1987] and cardiovascular disease [Centerwall, 1981], which will have a cost impact. A large long-term follow-up study has been funded by the UK Department of Health to assess these risks. It is unlikely, however, that these serious clinical events will occur in a sufficient number of women to close the cost gap significantly.

3.4.2 Health outcomes

A range of outcome data was collected in the trial. Shortly after surgery, TCRE avoids much of the morbidity associated with AH, resulting in women returning to their normal activities more quickly. By four months follow-up, results are more equivocal. Although, overall, the majority of women in both treatment groups are treated successfully on the basis of satisfaction rates at four months (83% TCRE; 90% AH), this success rate is statistically significantly higher amongst AH patients, probably reflecting their greater improvement in menstrual symptoms and treatment failure in a proportion of women in the TCRE group.

As the period of follow-up lengthens in a trial comparing surgical treatments for menorrhagia, strictly clinical measures of health outcome become less relevant to an evaluation, and patients' perceived health status take on more importance.

84

Hence, the longer-term follow-up reported here concentrates on patients'
symptoms, their perceptions of their HRQL and satisfaction levels. On the basis
of the postal questionnaire, patients initially randomised to AH were doing as
well or better on each of these measures of outcome than patients randomised
to TCRE.

The longer-term follow-up study reported here has two weaknesses. The first is
the fact that a proportion of women has been lost to follow-up (17 (17%) and 24
(25%) of TCRE patients and AH patients, respectively), and no longer-term
symptom and HRQL data were available for these women. It is crucial to assess
the longer term outcomes of surgery for menorrhagia, and the only feasible way
of doing this is through postal questionnaires. However, this form of data
collection inevitably results in some patients being lost to follow-up. Indeed, the
79% overall response rate in this study might be considered quite high. The
possibility exists that non-responders differ from responders. However, for a
major bias to be introduced into the results, these differences would have to
apply asymmetrically between the two randomised groups. Although it might be
considered unlikely that, for example, non-responders randomised to AH are
largely dissatisfied with the results of surgery and non-responders randomised to
TCRE are generally satisfied, it is not possible unequivocally to exclude the
possibility of bias. A degree of caution is, therefore, necessary in interpreting
these results.

The second weakness of the long-term assessment of outcomes is the lack of
baseline data for the SF36. As the SF36 has been shown to be sensitive to
changes in the health of women who have undergone surgery for menorrhagia
[Jenkinson *et al*, 1994], it would have been useful for women in this trial to have
completed the instrument at randomisation and at various subsequent intervals.
However, when the trial was being planned, the SF36 had not been fully
developed, nor had it been validated in the UK. Furthermore, it was felt that
there was no sufficiently sensitive alternative general measure of HRQL available
at the time. Although the availability of the SF36 data at baseline would have

strengthened the study, the use of the instrument at a single point of follow-up, given that the trial is randomised, provides a valuable picture of HRQL, which no previously published trial in this area has given.

3.4.3 Cost-effectiveness

It is likely that the cost advantage of TCRE over AH will remain, whatever the period of follow-up. However, for purchasers and providers of health care in this area, the crucial issue to consider is the relative *cost-effectiveness* of the two procedures, which requires an assessment of both costs and outcomes. On current evidence, there is a clear trade-off between the two treatments. On the one hand, AH will resolve menstrual problems and is more likely to provide high levels of satisfaction with treatment. On the other hand, despite leaving a proportion of women dissatisfied, TCRE does seem to provide satisfactory treatment for most women and involves a much shorter period of convalescence than AH.

A CEA relates the differential cost of two interventions to a uni-dimensional measure of effectiveness. If women's satisfaction with treatment is seen as an adequate 'all-embracing' measure of effectiveness, then its use in CEA seems appropriate. On this basis, and assuming that the results presented here are typical of those that would prevail in routine clinical practice, the use of AH rather than TCRE to treat menorrhagia would, on average, have an incremental cost of £1882 for each extra treatment success on the basis of data collected at longer-term follow-up. The ultimate decision about whether this is a reasonable cost to pay, given the benefits effects generated, has to be left to health care purchasers.

3.4.4 Methodological issues

The trial-based economic analysis of AH versus TCRE reported in this chapter represents a starting point for the exploration of the methodological issues related to the economic evaluation of MAS in general and AH versus TCRE in particular. In economic analysis alongside clinical trials, uncertainty associated

with the results is usually expressed in terms of the variability in stochastic data, and the results presented here reflect this in the use of hypothesis tests and confidence intervals around mean cost differences. However, several other sources of uncertainty exist in economic evaluation [Briggs *et al*, 1994]. Two areas of uncertainty in particular are highlighted by this chapter, and these relate to the measurement of benefits and to generalisability.

It is clear from the results presented in this chapter that, whilst AH is apparently more costly than TCRE, at least on the basis of 2.2 years follow-up, it is unclear which treatment is the more effective given the trade-off discussed above in terms of short-term versus longer-term outcomes; a trade-off which is typical of many forms of MAS when compared with conventional surgery and which lies at the heart of the economic issues considered in this thesis. The CEA presented here is based on the premise that reported rates of satisfaction with treatment are an adequate all-embracing measure of effectiveness, reflecting how women trade-off the various characteristics of the two treatments. However, it is unlikely that this fairly crude way of defining a treatment success fully reflects women's attitudes to the alternative treatments. In order to develop a measure of benefit which more accurately reflects the preferences of women with menorrhagia, it is necessary to understand more about the nature of these preferences, and this is the focus of Chapter 4 of this thesis.

A further limitation of the use of satisfaction rates as a primary outcome measure within CEA is that such an outcome lacks the generic qualities required as a basis of informed resource allocation across disease areas and health care programmes. The SF36 data collected within the Bristol trial 2.2 years after surgery represents a means of expressing the outcomes of the two treatments on a generic scale. However, although work is being undertaken to translate SF36 data into a single index [Brazier *et al*, 1994], no way yet exists of expressing its results on the uni-dimensional scale necessary to assess cost-effectiveness. The use of cost-utility analysis, where benefits are measured in terms quality-adjusted life years (QALYs), offers a way of expressing the

outcomes of AH and TCRE in the form of a generic measure of benefit. Furthermore, in that patients' values can be used to construct QALYs, this form of benefit measure may have the further advantage of more adequately reflecting patients' treatment-related preferences than success rates based on patient satisfaction. Chapters 5 and 7 consider the use of cost-utility analysis in the evaluation of AH versus TCRE, focusing on its empirical implications and its methodological strengths and weaknesses.

The second methodological issue highlighted by the trial-based economic analysis reported in this chapter relates to generalisability. Although an increasing number of economic evaluations are being undertaken alongside RCTs [Adams *et al*, 1992], it has been recognised that the possible limits on the external validity of clinical trials discussed in Chapter 1 may have a significant impact on the economic variables collected within trials [Drummond and Stoddart, 1984; Eisenberg *et al*, 1989; Drummond and Davies, 1991; Drummond, 1995]. Firstly, clinical trials of MAS - and indeed other surgical procedures - tend to be undertaken in specialist units. The Bristol trial of AH versus TCRE was undertaken in a medical school by clinical enthusiasts and clinical practice, especially in relation to TCRE, may not be typical of routine clinical practice elsewhere in the NHS.

Secondly, MAS techniques tend to develop quickly and, no sooner has a trial been completed focusing on one version of an application, than another is beginning to diffuse. Women were recruited into the Bristol trial during 1990-1, and it is likely that clinical practice, in relation to both TCRE and AH will have changed since then. In particular, although these two surgical techniques are the most frequently used for the treatment of menorrhagia [RCOG Audit Unit, personal communication], other forms of MAS such as laser ablation are used in some centres. Furthermore, new approaches to TCRE and hysterectomy, such as the use of rollerball resection rather than the loop in TCRE and laparoscopic assistance in hysterectomy, are becoming more widely used.

The third impact that trial-based data might have upon the generalisability of an economic evaluation relates to the atypical clinical practice that tends to be generated by a clinical trial, wherever it is undertaken. In the Bristol trial, for example, the clinical investigators decided not to use drugs to prepare the endometrium prior to TCRE because this was not possible with AH and might bias the results. This decision was taken despite that fact that most centres use such drugs with TCRE [RCOG Audit Unit, personal communication].

The limitations that the use of data from RCTs impose upon the generalisability of an economic evaluation, together with the development of methods to address this problem, are the focus of Chapter 6 of this thesis.

3.5 Conclusions

The first route into the economic analysis of MAS is often the use of effectiveness and key resource use data from a clinical trial set up primarily to inform *clinical* policy. Given the high internal validity of data from these trials, this approach can begin to answer some of the questions that need to be addressed to inform resource allocation. However, trial-based analysis is rarely the final word in the economic evaluation of health care technologies and, typically, key uncertainties remain.

This is clearly the situation as regards the economic assessment of AH versus TCRE. The Bristol trial provides a useful way into the analysis of the two treatments, but several areas of uncertainty are highlighted, and these require additional data collection and analysis to be illuminated, which is the focus of subsequent chapters.

Appendix 3.1   Unit costs used to value resource use

Resource	Unit	Cost (£)
Tests		
Full blood count	Per test	1.59
'Group and save'	Per test	9.54
Urea and electrolytes	Per test	4.51
Pelvic ultrasound	Per test	14.58
Histology test	Per test	71.44
Blood transfusion		
Blood and regional handling cost	One unit of red cell (OAS) product	28.09
Hospital handling cost	One unit of red cell (OAS) product	2.65
Drugs		
Antibiotics		
Co-amoxiclav	Per 375mg tablet	0.48
Amoxycillin	Per 250mg capsule	0.20
Metronidazole	Per 400mg tablet	0.11
Amoxycillin (IV)	Per 500mg vial	1.06
Metronidazole (IV)	Per 100mg bottle	6.49
Gentamicin (IV)	Per 2ml vial	2.37
Frusemide	Per 40mg tablet	0.01
Ferrous suphate	Per 170mg tablet	0.01
Analgesics		
Papaveretum	Per 1ml amp	0.19
Paracetamol	Per 10 500mg tablets	0.06
Oestrogen	Per 625ug tablet	0.07
Theatre staff		
Surgeon (consultant)	Per minute	0.45
Anaesthetist (consultant)	Per minute	0.45
Anaesthetic nurse (Grade H)	Per minute	0.20
Instrument nurse (Grade G)	Per minute	0.18
Trolley nurse (Grade G)	Per minute	0.18
Circulating nurse (Grade G)	Per minute	0.18
Senior house officer	Per minute	0.19

Resource	Unit	Cost (£)
Theatre equipment		
'Non-routine'		
Camera	Per patient	2.36
Light source	Per patient	1.07
Telescope	Per patient	1.87
Resectoscope	Per patient	1.07
Theatre consumables		
Loops	Per patient	3.46
Irrigation tubing: in	Per patient	3.69
Catheter	Per patient	0.21
Gloves	Per patient	0.74
Glycine	Per 2 litre bag	2.48
Suturing materials	Per patient	8.43
Blades	Per patient	0.21
Dressings	Per patient	0.32
Supra pubic catheter	Per patient	5.83
'Foley' catheter	Per patient	2.64
'Filshie' clips	Per patient	17.69
Theatre anaesthesia		
TCRE	Per minute	0.27
AH	Per minute	0.35
Theatre overheads	Per minute	1.08
Hospital "hotel" services		
Ward	Per in-patient day	120.00
ITU	Per in-patient day	494.00
GP visit	Per visit	5.51

Chapter 4

Women's Views on Menorrhagia: Health-Related Quality of Life and Preferences

4.1 Introduction

The purpose of this chapter is to explore women's attitudes to menorrhagia and their preferences. Often in economic evaluations decisions are taken about methods for benefit measurement without a full understanding of patients' attitudes and preferences in the clinical area in question. The work reported in this chapter was undertaken to answer a range of research questions, not all of which were related to economic analysis. However, by assessing the characteristics of women's attitudes to heavy bleeding and alternative forms of management, the study serves as a bridge between the cost-effectiveness analysis in Chapter 3, which links costs to a measure of effect collected on patients within the context of a trial, and a more complete measurement of benefit which is consistent with women's health- and treatment-related preferences.

It is clear from Chapter 3 that the process and outcomes of surgical treatment for menorrhagia are inherently multi-dimensional. The choice of treatment involves trade-offs in terms of such things as complications and side effects, duration of convalescence, prophylaxis against specific cancers, probability of success, contraceptive effect and achievement of amenorrhoea and hypomenorrhoea. The advent of MAS procedures for menorrhagia has accentuated these trade-offs, presenting a stark choice to women who fail with drugs and require further treatment: undergo MAS and experience a swift return to usual activities but risk an unsuccessful treatment; or accept a hysterectomy which has a much longer convalescence, has various risks associated with it but which provides a once-and-for-all solution to menorrhagia and offers a prophylaxis against some gynaecological cancers.

These characteristics are typical of many applications of MAS in a range of clinical areas, and they pose some specific problems for economic evaluation. In particular, as emphasised by the trial-based analysis detailed in Chapter 3, it is unlikely that any single clinical or patient-based measure of outcome will adequately represent the net benefit of treatment from a patient's perspective. If the measurement of benefit is to be a central element of economic analysis of AH and TCRE, it is important to have a greater understanding of women's attitudes to the condition and of their treatment-related preferences.

Furthermore, although clinicians will have a crucial role in identifying the feasible treatments for a given women with menorrhagia - that is, in excluding treatments that are not medically advisable or not available within the centre - the existence of trade-offs in the process and outcomes of interventions would suggest an important role for women in identifying a preferred therapy. The concept of shared decision making has some important implications for economic analysis.

Little research has been undertaken on women's preferences in the area of menorrhagia. Warner [1994] identified 257 women with a range of menstrual problems and invited them, together with a control group of 105 women without

similar problems, to complete a lengthy questionnaire, part of which focused on preferences for particular treatments and for the effects of therapy. Of the 257 women, 39% had menorrhagia. It was found that women would prefer therapy that normalised, rather than eliminated, their periods (89%) and that also offered a reversible contraceptive effect (74%). Preferences regarding a 'one-off' operation compared to tablets were very similar in the group (47% versus 44%). Preferences were most strongly related to a woman's reproductive status: women without children or who had not been sterilised were less likely to prefer a treatment that affected their periods or fertility. The study concluded that 'treatment for menstrual complaints should be decided with reference to the full scope of the individual's menstrual problems and treatment aspirations' [Warner *et al*, 1994; p.109].

As part of a large study looking at the treatment of menorrhagia in general practice, Coulter *et al* [1994B] asked women about their preferences for treatment, and considered the extent to which these preferences influenced GPs' management decisions. Of the 488 women who completed a questionnaire, 36.5% indicated that they had a strong preference regarding treatment, and these patients were more likely to be older, to have received higher education, to be in social class I or II and to have previously consulted a GP with menstrual problems. Of these women, 14.8% indicated a preference for surgery, a preference which was more likely in women with severe menorrhagia and who had not received higher education. As a part of a separate questionnaire, GPs were asked to indicate their understanding of each woman's preferences, and this was correct in only 34.4% of those cases where a strong treatment preference was expressed.

With the exception of these two studies, very little has been published on women's preferences for, and for the characteristics of, treatment of menorrhagia. No studies have been identified that look specifically at women's attitudes to the trade-offs that present themselves in choosing between MAS and hysterectomy. Within economic evaluation, there is a strong tradition of

considering the value patients attach to the outcomes of care, and this is the focus of Chapters 5, 7 and 8 of this thesis. The aim of this chapter is to report on survey work which has elicited details of women's attitudes to menorrhagia and its treatment. The survey offers a descriptive overview of women's health-related quality of life (HRQL), attitudes to menorrhagia and alternative treatments, as well as their preferences and aspirations.

4.2   Methods

4.2.1 Survey sample

The survey was undertaken using a postal questionnaire. The population of interest was women who had recently been referred for the first time to hospital by their GP due to menorrhagia. The survey was undertaken in two centres: at St Michael's Hospital in Bristol and at the Princess Margaret Hospital in Swindon; and women were recruited during two overlapping time periods. To identify a sample, all new referral letters sent to the hospitals' gynaecology departments by GPs were scrutinised by a medical secretary to see if sufficient information existed to include the women in the study. The inclusion criterion was that a woman was to be a new and non-urgent referral for heavy bleeding. Each woman identified in this manner over a specific period was sent a questionnaire. On the basis of the information provided in the questionnaire, women were excluded from the analysis if they had serious concomitant illness; if they had previously undergone gynaecological surgery; if, by the time they completed the questionnaire, they had already had their out-patient appointment; or if their symptoms suggested that they did not have uncomplicated menorrhagia (eg. inter-menstrual bleeding).

4.2.2 Components of the questionnaire

The questionnaire was designed to elicit information in a number of areas. Prior to the main survey beginning in Bristol, a pilot study was undertaken with the purpose of testing how women comprehended the exercise and to identify any element that was easily misunderstood. In the pilot study, questionnaires were

sent to 10 women who had been referred to St Michael's hospital, eight of whom replied. Small changes were made to the questionnaire on the basis of their responses.

In Section 1 of the questionnaire, women were asked to provide their date of birth and their post code. In Section 2, they were asked a series of questions about their menstrual symptoms, including the duration of the problem and the presence of blood clots. In Section 3, the respondent was asked to complete the EuroQol instrument, a non-disease specific means of describing and valuing HRQL [EuroQol Group, 1990; Brooks, 1996]. The first part of the EuroQol asks respondents to categorise their health state on five dimensions (mobility, self care, usual activities, pain and anxiety/depression), where each dimension has three possible levels of response. In the second part of the instrument, respondents are asked to value their health status using a vertical rating scale presented like a thermometer, which is labelled with 'best imaginable health state' at the top and 'worst imaginable health state' at the bottom. Women in the survey were asked to complete both parts of the EuroQol from two perspectives: their health today and their health on the day during which their menstrual bleeding is heaviest.

Section 4 was made up of two parts. In the first part, a series of 10 characteristics of treatment was listed and framed in the first person from the woman's perspective. Examples of these were 'I want to stay in hospital for as short a period as possible', 'I want treatment that will put a stop to my periods for good', and 'I don't want to have to worry about contraception after treatment'. Women were asked to rate each characteristic of treatment on a four point categorical scale defined as 'very important', 'of some importance', 'of little importance' and 'not important'. They were then asked to list, and to rank, the three most important characteristics from their point of view.

In the second part of Section 4 women were presented with two unnamed scenarios describing the process and outcomes of TCRE and AH, respectively.

The scenarios were based on a review of the relevant clinical literature and on clinical opinion. They used a mixture of condition- and treatment-specific items, as well as more general ones, and they covered the whole period between the operation and one year follow-up. In the case of TCRE, the description included details of the risk of treatment failure. Women were asked to indicate, on the basis of the scenarios, which treatment option they would choose to have, and they were given the opportunity to respond that they did not know or that they would not like either, preferring to put up with their symptoms. They were then asked to value the two scenarios on the EuroQol rating scale.

In Section 5 women were asked to indicate which information sources they had been exposed to in relation to treatments for menorrhagia, and they were asked whether they felt well-informed by these sources. Women were then asked whether they had any strong positive or negative preferences about treatments.

4.2.3 Analysis

The samples from the two centres are analysed separately and together for Sections 2 and 3 of the questionnaire, which focus on descriptive data for the samples. The analyses of Section 4 and part of Section 5 are also undertaken centre-by-centre and pooled. Due to the fact that it concentrates on only a sub-group of women (ie. those with strong positive or negative treatment preferences), the analysis of the remainder of Section 5 is based only on pooled data.

4.3   Results

4.3.1 Survey population

In Bristol women were sent questionnaires between January 1994 and October 1995; in Swindon recruitment took place between August 1994 and March 1995. In Bristol, a total of 175 women were identified from GP referral letters and were sent a questionnaire. Of these, 115 (66%) women returned a completed questionnaire. Two of these women were excluded: one because her

questionnaire responses indicated that she did not actually have menorrhagia; the other was because she had already undergone a TCRE. Therefore, a total of 113 women were included in the Bristol sample.

In the Swindon survey, 202 questionnaires were sent out and 119 (59%) were returned completed. Of these, 11 women were excluded because they had already had an out-patient appointment (3), they had a concomitant illness (6), they had undergone previous gynaecological surgery (1) and they had inter-menstrual bleeding (1). Hence 108 women were entered into the study in Swindon.

4.3.2 Characteristics of women in the survey

Table 4.1 details the characteristics of women in the survey. In addition to women's ages, the table shows the severity of their menorrhagia in terms of a range of questions including duration, days with heavy flow and the passing of blood clots. In both centres, 50% of the median number of days of a woman's period consisted of heavy flow. Overall, 88% and 91% of women passed blood clots and experienced flooding episodes, respectively.

An important point as regards the analysis of the questionnaire is the comparison between the two centres of the women in the survey Table 4.1 shows that there are no statistically significant differences in most of the characteristics of the two groups. The only exception to this is the duration of women's menstrual problems, which was significantly shorter in the Swindon sample (p=0.02). For this reason, the bulk of the remainder of the analysis presents the results of the two centres separately as well as pooled.

4.3.3 General health status

The severity of women's menorrhagia evident in Table 4.1 is mirrored in their responses to the EuroQol questionnaire, which are shown in Table 4.2. The table emphasises the major impact of women's 'heaviest' days on their health

Table 4.1   Characteristics of the samples of women included in the survey in
             Bristol and Swindon

Characteristic	Bristol (n = 113)	Swindon (n = 108)	P Value	Pooled (n = 221)
Mean (SE) age (years)	41.38 (0.73)	40.47 (0.82)	0.41*	40.94 (0.55)
Median (range) duration of menorrhagia (months)	24 (3-360)	12 (1-420)	0.02[t]	18 (1-420)
Median (range) days per month bleeding	8 (2-31)	8 (3-31)	0.93[t]	8 (2-31)
Median (range) days per month with heavy flow	4 (1-21)	4 (1-25)	0.67[t]	4 (1-25)
Number (%) of women passing blood clots	94 (86)	93 (90)	0.40[ø]	187 (88)
Number (%) of women experiencing flooding	105 (95)	89 (87)	0.09[ø]	194 (91)
Number (%) of pads on heaviest day:			0.54[ǂ]	
1-4	1 (1)	1 (1)		2 (1)
5-9	36 (33)	38 (37)		74 (35)
10-14	51 (47)	45 (44)		96 (46)
15 or more	21 (19)	18 (18)		39 (18)
Median (range) days of work lost during previous year due to menstrual problems	0 (0-50)	0 (0-36)	0.62[t]	0 (0-50)

*	T test
t	Wilcoxon rank-sum test
ø	Fisher's exact test
ǂ	Mann Whitney U test

status, by showing the large number of women who moved from having no
impairment (ie. level 1 on a given dimension of the instrument) on the day the
questionnaire was completed to having an impairment (ie. levels 2 or 3) on their
'heaviest day'.

Figure 4.1 brings this point out further using pooled data from the two centres.
The figure shows, for each dimension of the EuroQol, the percentage of women
with an impairment on the day the questionnaire was completed and on their
heaviest day, respectively. Overall, 43% of women had an impairment to their

Table 4.2   EuroQol classifications and valuations of women in the survey in Bristol and Swindon

	Bristol (n = 113)		Swindon (n = 108)		Pooled (n = 221)	
	Today	'Heaviest'	Today	'Heaviest	Today	'Heaviest'
EuroQol groups (n(%))						
Group 1: Mobility						
No problems in walking about	86 (86)	23 (23)	94 (93)	29 (29)	180 (90)	52 (26)
Some problems in walking about	14 (14)	73 (72)	5 (5)	67 (68)	19 (9)	140 (70)
Confined to bed	0 (0)	5 (5)	2 (2)	3 (3)	2 (1)	8 (4)
Group 2: Self care						
No problems with self care	97 (96)	77 (84)	97 (98)	84 (88)	194 (97)	161 (86)
Some problems with self care	4 (4)	15 (16)	2 (2)	11 (12)	6 (3)	26 (14)
Unable to wash or dress	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Group 3: Usual activities						
No problems with usual activities	83 (82)	14 (14)	92 (92)	15 (15)	175 (87)	29 (15)
Some problems with usual activities	14 (14)	66 (65)	6 (6)	70 (71)	20 (10)	136 (68)
Unable to perform usual activities	4 (4)	22 (22)	2 (2)	13 (13)	6 (3)	35 (18)
Group 4: Pain						
No pain or discomfort	72 (71)	4 (4)	72 (73)	6 (6)	144 (72)	10 (5)
Some pain or discomfort	28 (27)	48 (45)	25 (26)	54 (53)	53 (27)	102 (49)
Extreme pain or discomfort	2 (2)	54 (51)	1 (1)	42 (41)	3 (2)	96 (46)
Group 5: Emotional						
Not anxious or depressed	73 (72)	12 (12)	68 (71)	13 (13)	141 (71)	25 (12)
Moderately anxious or depressed	28 (27)	56 (54)	27 (28)	54 (53)	55 (28)	110 (53)
Extremely anxious or depressed	1 (1)	36 (35)	1 (1)	35 (34)	2 (1)	71 (34)
EuroQol visual analogue scores (mean (SE))	78.59 (1.91)	40.79 (2.19)	82.79 (1.46)	41.91 (2.46)	80.59 (1.22)	41.32 (1.63)

'Heaviest' means the day during which menstrual bleeding is heaviest
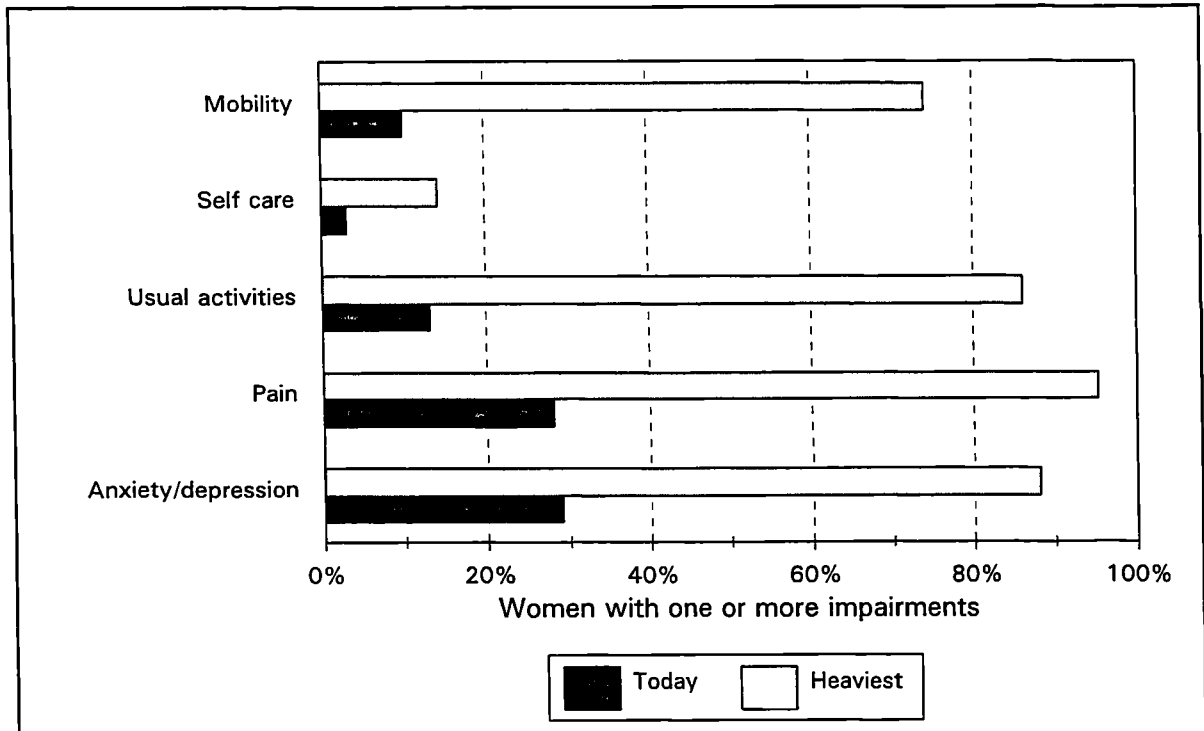
Figure 4.1   Proportion of women with an impairment on EuroQol dimensions on the day they completed the questionnaire and on their heaviest day. Data are pooled from the two centres.

health status, on one or more dimensions, on the day they completed the questionnaire, compared to 99% on the heaviest day of their period.

The major down-turn in women's health status during the day when their blood loss is heaviest is also emphasised in the EuroQol rating scale results, which are also shown in Table 4.2. The mean value women attached to their health state on their heaviest day is only 51% that of the value they associated with their health state on the day the questionnaire was completed (52% in Bristol, 51% in Swindon).

4.3.4 Treatment characteristics

Table 4.3 shows how women rated the various characteristics of surgical treatment on a scale running from 'very important' to 'not important'. On the basis of pooled data, the three characteristics most frequently rated as 'very important' were getting back to usual activities as soon as possible, experiencing the least pain and discomfort and spending as

Table 4.3   Levels of importance women attached to various characteristics of surgical treatments for menorrhagia.   Figures are numbers (%)

Characteristics of treatment	Bristol (n=113)				Swindon (n=108)				Pooled (n=221)			
	Very imp	Of some imp	Of little imp	Not imp	Very imp	Of some imp	Of little imp	Not imp	Very imp	Of some imp	Of little imp	Not imp
Hospital stay as short as possible	54 (50)	31 (28)	12 (11)	12 (11)	43 (42)	38 (37)	10 (10)	12 (12)	97 (46)	69 (33)	22 (10)	24 (11)
Not removing womb	22 (21)	18 (17)	19 (18)	44 (43)	30 (31)	21 (21)	19 (19)	28 (29)	52 (26)	39 (19)	38 (19)	72 (36)
Removing womb	30 (31)	19 (19)	23 (23)	26 (27)	21 (23)	24 (27)	16 (18)	29 (32)	51 (27)	43 (23)	39 (21)	55 (29)
Not leaving scar	22 (20)	24 (22)	22 (20)	41 (38)	17 (16)	28 (27)	27 (26)	32 (31)	39 (18)	52 (24)	49 (23)	73 (34)
Least pain & discomfort	52 (47)	33 (30)	16 (15)	9 (8)	46 (44)	39 (37)	13 (12)	7 (7)	98 (46)	72 (33)	29 (13)	16 (7)
Stop periods for good	49 (45)	18 (17)	20 (19)	21 (19)	41 (40)	24 (24)	13 (13)	24 (24)	90 (43)	42 (20)	33 (16)	45 (21)
Reduce periods	23 (23)	19 (19)	16 (16)	41 (41)	27 (28)	29 (31)	13 (14)	26 (27)	50 (26)	48 (25)	29 (15)	67 (35).
No worry about contraception	24 (24)	7 (7)	12 (12)	58 (57)	16 (16)	4 (4)	9 (9)	68 (70)	40 (20)	11 (6)	21 (11)	126 (64)
Resume sex life on soon as possible	29 (26)	54 (49)	8 (7)	19 (17)	26 (25)	42 (41)	15 (15)	20 (19)	55 (26)	96 (45)	23 (11)	39 (18)
Back to usual activities as soon as possible	62 (56)	38 (34)	4 (4)	7 (6)	60 (58)	35 (34)	6 (6)	2 (2)	122 (57)	73 (34)	10 (5)	9 (4)

short a time as possible in hospital. A large proportion of women in Bristol (45%) also rated the stopping of periods for good as 'very important'.

Figure 4.2 shows the mean ranking on the importance scale, where 'very important' counts as the highest rank (1) and 'not important' counts as the lowest rank (4). The figure shows that, with the exception of the characteristic of reducing periods which women in Swindon ranked much higher than those in Bristol due perhaps to the fact that women in Swindon are referred earlier, the mean ranks were very similar between the centres. Reflecting the large number of women rating them as 'very important', the highest mean ranks were for getting back to usual activities as soon as possible, experiencing the least pain and discomfort and having a short hospital stay.

Women were also asked to indicate which three characteristics listed in the questionnaire were most important to them, and to rank these. Table 4.4 details their responses. The characteristic most frequently rated as the most important by women in both centres was the stopping of periods for good (27% and 29% in Bristol and Swindon, respectively). In contrast, the characteristic with the next highest overall proportion of women rating it as the most important was not removing the womb (18% overall; 20% and 17% in Bristol and Swindon, respectively). Getting back to usual activities quickly was considered the most important characteristic of treatment by 23% of women in Swindon, but by only 9% in Bristol.

If a woman's top three characteristics are scored 3 (most important), 2 (second most important), 1 (third most important) or 0 if it is not rated in the top three, the mean score per characteristic can be calculated, and these are also detailed in Table 4.4. Overall, the highest mean scores were for stopping periods for good, an early return to usual activities and the least pain and discomfort. However, the characteristic of not removing the womb also scored quite highly in both centres.
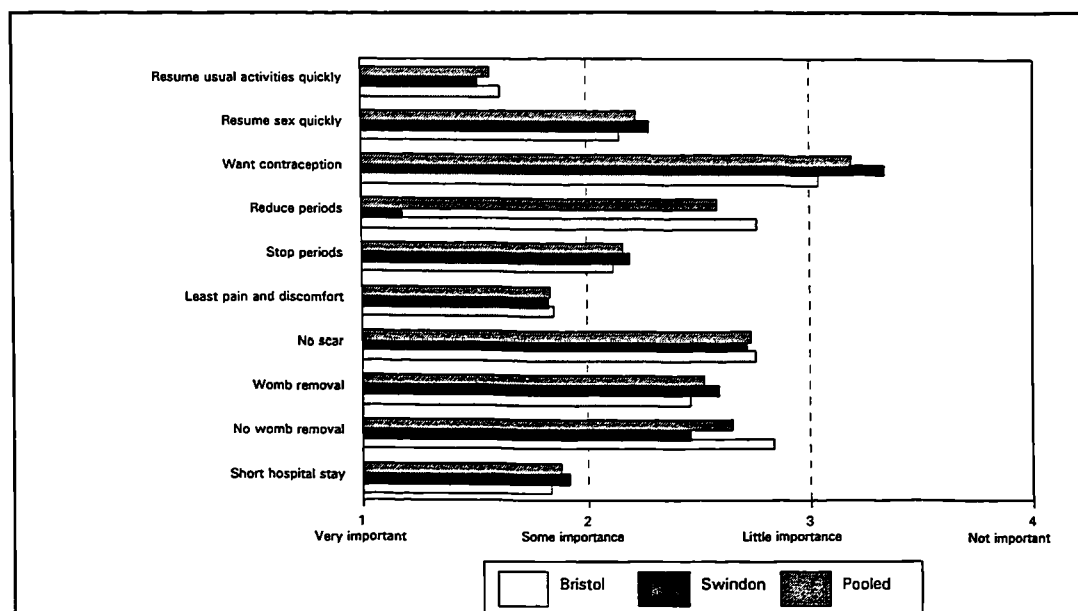
Figure 4.2   Mean ranks women attached to the characteristics of surgical treatment for menorrhagia.

In order to encourage women to think about the trade-offs existing between treatments in terms of these characteristics, they were asked to choose between, and to value, two treatment scenarios, one describing AH and the other TCRE. Table 4.5 shows that very similar proportions of women in the two centres preferred each of the two treatments; overall 43% preferred AH and 41% preferred TCRE. A total of 13% and 19% in Bristol and Swindon, respectively, either would accept neither treatment or felt unable to choose. The approximately equal division of women in the survey in terms of preferences for AH and TCRE was also reflected in the values they attached to the two scenarios using the EuroQol rating scale. Table 4.5 shows the very similar mean scores for AH and for TCRE in both centres.

4.3.5 Information and treatment preferences

Figure 4.3 shows the proportions of women in the two samples receiving information from various sources about treatments for menorrhagia. The main source is women's GPs, but a large proportion of women had received information from friends (37% overall) and from magazines (30%). Despite the fact that 70% of women overall had received information from their GPs, only

Table 4.4   Details of the three most important characteristics and mean scores of surgical treatments for menorrhagia. Figures are numbers (%) unless otherwise stated

Characteristic	Bristol (n = 113) Level of importance			Mean Score*	Swindon (n = 108) Level of importance			Mean Score*	Pooled (n = 221) Level of importance			Mean Score*
	1st	2nd	3rd		1st	2nd	3rd		1st	2nd	3rd	
Hospital stay as short as possible	10 (9)	10 (9)	8 (8)	0.55	7 (7)	14 (14)	13 (13)	0.63	17(8)	24 (12)	21 (10)	0.59
Not removing womb	21 (20)	6 (6)	3 (3)	0.74	17 (17)	7 (7)	5 (5)	0.67	38 (18)	13 (6)	8 (4)	0.71
Removing womb	16 (15)	4 (4)	1 (1)	0.54	6 (6)	6 (6)	8 (8)	0.39	22 (11)	10 (5)	9 (4)	0.47
Not leaving scar	3 (3)	4 (4)	6 (6)	0.22	0 (0)	2 (2)	2 (2)	0.06	3 (1)	6 (3)	8 (4)	0.14
Least pain & discomfort	9 (8)	12 (11)	19 (18)	0.66	8 (8)	18 (18)	9 (9)	0.70	17 (8)	30 (14)	28 (14)	0.68
Stop periods for good	29 (27)	20 (19)	8 (8)	1.23	29 (29)	7 (7)	9 (9)	1.12	58 (28)	27 (13)	17 (8)	1.18
Reduce periods	5 (5)	10 (9)	4 (4)	0.37	7 (7)	12 (12)	4 (4)	0.48	12 (6)	22 (11)	8 (4)	0.42
No worry about contraception	1 (1)	3 (3)	7 (7)	0.13	1 (1)	5 (5)	2 (2)	0.15	2 (1)	8 (4)	9 (4)	0.14
Resume sex life as soon as possible	3 (3)	21 (19)	22 (21)	0.66	2 (2)	13 (13)	20 (20)	0.52	5 (2)	34 (16)	42 (20)	0.59
Back to usual activities as soon as possible	10 (9)	18 (17)	28 (26)	0.89	23 (23)	15 (15)	28 (28)	1.27	33 (16)	33 (16)	56 (27)	1.07

*   If a characteristic was judged most important then a score of 3 is given, 2nd most important then a score of 2 and third most important a score of 1. Otherwise a score of zero is given.

Table 4.5   Women's preferences concerning the two treatment options
described in unlabelled scenarios in the questionnaire

	Bristol (n = 113)	Swindon (n = 108)	Pooled (n = 221)
Number (%) preferring:			
AH	50 (46)	40 (39)	90 (43)
TCRE	44 (41)	42 (42)	86 (41)
Would accept neither	4 (4)	5 (5)	9 (4)
Unable to choose	10 (9)	14 (14)	24 (11)
Mean (SE) rating scale value for*			
AH	61.77 (3.41)	56.50 (3.56)	59.34 (2.46)
TCRE	57.25 (2.96)	56.31 (3.02)	56.82 (2.11)

*   EuroQol rating scale - 0 ('worst imaginable health state') to 100 ('best imaginable health
state')

44% of women considered themselves well-informed about menorrhagia and its
treatment (51% in Bristol and 37% in Swindon).

When asked to think about specific treatments for their menorrhagia, 46% of
women in Bristol and 39% in Swindon (43% overall) indicated that they had a
strong positive preference for a particular treatment. Very similar proportions -
47% and 37% in Bristol and Swindon, respectively - reported strong negative
preferences for treatments. Table 4.6 shows where these preferences lay, for
the group overall, by listing the actual treatments women noted on the
questionnaire. The first part of the table shows the specific treatments listed by
women. In the second part, treatments have been grouped into broad classes;
for example, the various medical treatments have been grouped into drug
therapy.

Table 4.6 clearly shows that women are heterogenous in their treatment
preferences, with roughly equal proportions having a strong positive and negative
preferences, respectively, for hysterectomy. A similar case is true with drug
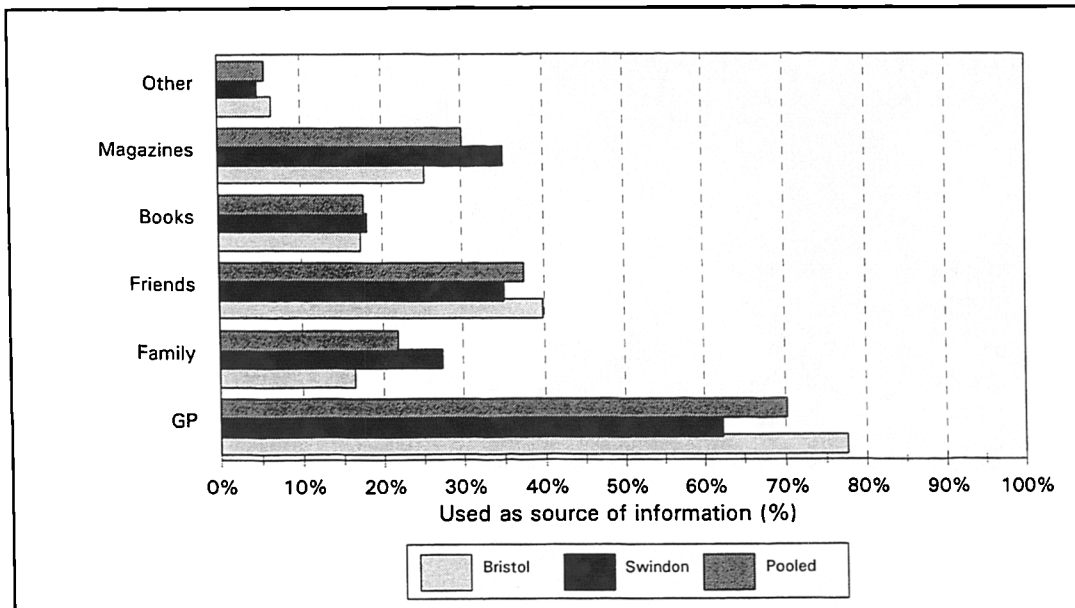
Figure 4.3   Sources of information on menorrhagia to which women had access.

therapy, although the proportion with negative preferences is much larger than that with positive ones: 11% have a strong preference for a particular drug or for medical treatment in general, whilst 22% have a strong negative preference. Although the major issue in surgical treatment for menorrhagia is the choice between MAS and hysterectomy, relatively few women have strong (positive or negative) preferences about the former.

4.4   Discussion

A broad choice exists between MAS and open surgery in many clinical areas, and these treatments invariably have different risks, types of process and outcomes over which patients are likely to have preferences which will influence how they benefit from treatment. This has important implications for the measure of benefit used in the economic comparison of MAS and open surgery. In addition, if patients' preferences are to be given a greater role in determining their optimal treatment, ways of incorporating this form of management into economic analysis are required.

Table 4.6   Treatment cited by women for which they had a strong positive or
             negative preference. Data only relate to those women who indicated a
             strong preference and are pooled across centres

Treatment	Number (%)*
Treatments cited by women for which they have a *positive* preference	
Hysterectomy	31 (54)
Option 1 in questionnaire (AH)	8 (14)
HRT	4 (7)
Option 2 in questionnaire (TCRE)	3 (5)
Laser treatment	3 (5)
D&C	2 (4)
Other	6 (11)
Treatments cited by women for which they had a *negative* preference	
Hysterectomy	32 (49)
Tablets	6 (9)
Laser treatment	4 (6)
Option 1 in questionnaire (AH)	3 (5)
Option 2 in questionnaire (TCRE)	3 (5)
D&C	3 (5)
HRT	3 (5)
The pill	3 (5)
Surgery	2 (3)
Other	6 (9)
Treatment groups for which women had a *positive* preference	
Hysterectomy	39 (68)
Minimal access surgery	17 (12)
Drug therapy	6 (11)
D&C	2 (4)
Other	3 (5)
Treatment groups for which women had a *negative* preference	
Hysterectomy	35 (55)
Drug therapy	14 (22)
Minimal access surgery	6 (9)
D&C	3 (5)
Other	6 (9)

*   Percentages based on total number of treatments detailed (ie. some women detailed more
     than one)

Despite the importance of preferences in the management of menorrhagia, little

is known about the strength and direction of women's preferences, particularly in

relation to surgical treatments. This chapter describes a survey of women who had recently been referred to hospital with heavy periods, providing valuable information on their HRQL and their preferences for the characteristics of treatment, and for the treatments themselves.

It is clear from the survey that heavy periods have a serious impact on women's HRQL, and this confirms the results of other studies [Coulter *et al*, 1994A]. On the basis of women's responses to the EuroQol instrument, the most frequent health state into which women classified themselves in relation to the heaviest day of their period was 21222. That is, the typical woman on her heaviest day had moderate impairment on all dimensions of health status except self-care. A recent study by the University of York, funded by the Department of Health, asked 3395 members of the public to value a number of the EuroQol health states using the time-trade-off valuation technique [Williams, 1995][1]. Using the data from these interviews and further modelling techniques, the study was able to allocate values - on a zero (death) to 1 (perfect health) scale - to all EuroQol states. The mean value for the health state 21222 on the basis of the public's values was 0.62, which means that, on average, the public would be willing to trade-off nearly 40% of their remaining life-years to avoid a permanent impairment to health status to the extent that this survey indicates is associated with menorrhagia on the heaviest day of a woman's period.

In choosing between AH and TCRE, a number of risks and benefits need to be considered. The questionnaire asked women to consider a range of treatment characteristics, and their responses indicate that some were considered particularly important, including a speedy return to usual activities, the least possible pain and discomfort following surgery and stopping periods for good. However, neither treatment has all the characteristics that women feel are important. When women were asked to consider the three most important characteristics of surgical treatment to them, the highest scores in the groups overall were associated with stopping periods for good (best achieved with

---

[1] The time-trade-off valuation technique is discussed in more detail in Chapter 5

hysterectomy) and with a speedy return to usual activities (best achieved with minimal access surgery). This conflict of objectives was evident in some women at the individual level also: 11% of women indicated that both being in hospital for as short a period as possible and having a treatment that ended their periods for good were 'very important' to them; 11% of women rated a treatment that removed their womb and one that caused the least pain and discomfort during convalescence as 'very important'. This emphasises the point that the benefit measure used in the economic evaluation of AH and TCRE should reflect how women trade-off the characteristics of these two treatments.

If women are to play a greater role in identifying their optimal treatment, trade-offs are again all important. When women were asked to think about these trade-offs in the questionnaire, by way of two unlabelled scenarios describing AH and TCRE, 15% of women overall either felt unable to choose or indicated that they would rather put up with their symptoms than have either. It is likely that this rate would alter if women were provided with additional information, but it shows that a large proportion were unprepared to accept the trade-offs inherent in the choice between AH and TCRE. Of those who were able to state a preference for one of the treatments described, very similar proportions preferred hysterectomy and TCRE, suggesting that women have quite different attitudes to the trade-offs.

Clearly the issue of information is crucial to the preferences women express about therapy. Although 77% of women said that they had received information about treatments from their GP, only 44% felt they were well-informed. If women are to make a greater contribution to decisions about treatment in this area, the sources of information they currently have access to need to be supplemented by up-to-date, balanced, accurate and accessible sources.

Given the apparent inadequacy of existing information sources, it is perhaps not surprising that less than half the women in the survey had strong positive (43% overall) or negative (42%) preferences for specific treatments. These results are

similar to Coulter *et al's* [1994B] study which found that 46% of women with menorrhagia consulting GPs had no positive treatment preference. Of those women in the current survey who did express preferences, it was clear that views were most often formed about hysterectomy, with similarly large proportions of women expressing strong positive preferences for the removal of their uterus as negative ones. In contrast to hysterectomy, preferences concerning MAS were not well formed, which again is likely to reflect inadequate information. In comparison with Coulter *et al's* [1994B] survey in general practice which found that 17% of women would prefer drug therapy, very few women in this survey reported positive preferences about drug treatment, probably reflecting the fact that, as hospital referrals, many of them would have tried drugs already which had failed to ameliorate their symptoms adequately.

4.5   Conclusions

This chapter emphasises the importance of some descriptive assessment of patients' treatment-related attitudes and preferences regarding MAS prior to detailed benefit measurement for economic evaluation. The chapter has shown several important things about women's attitudes to menorrhagia: that it has a major detrimental impact on HRQL; that women have clear preferences about the characteristics of TCRE and AH in terms of process and outcomes; that there is heterogeneity between women in these preferences; and that many women are able to make trade-offs between the characteristics of treatments. These findings have important implications for the benefit measure used in the economic evaluation of AH and TCRE. The first of these is the need to develop a measure of benefit which reflects the impact of the two treatments on the characteristics which women think important. The second implication is that, if patients' values are considered important for resource allocation in this area, the benefit measure used in the economic evaluation should also reflect how women trade-off the various characteristics of treatment. Chapter 5 considers the extent to which cost-utility analysis based on the standard QALY meets these requirements.

Chapter 5

A Cost-Utility Analysis of AH Versus TCRE Using the Standard QALY Model

5.1 Introduction

The purpose of this chapter is to explore the role of cost-utility analysis (CUA) in the economic evaluation of MAS procedures. Chapter 3 assessed the relative cost-effectiveness of AH and TCRE using the methods of cost-effectiveness analysis (CEA) and treatment success, defined in terms of women's satisfaction with the results of surgery, as the measure of effectiveness. In the Bristol trial, AH was found to be significantly more costly than TCRE, but more effective, at 2.2 years. Therefore, to assess the relative cost-effectiveness of the two interventions, it is necessary for commissioners to judge whether the greater chance of a treatment success with AH is worth paying for in terms of its incremental cost.

The use of CEA has a number of limitations, both in general and specifically in the economic evaluation of MAS. This chapter uses CUA, and the standard QALY as a measure of benefit, in an attempt to overcome some of these limitations and, therefore, to provide some firmer evidence on the relative cost-effectiveness of the two procedures. A secondary aspect of the chapter is to develop a decision analytical model to provide a framework for the CUA, to facilitate the movement away from the comparison of two groups within a RCT. The model also provides a link to Chapter 6 which explores the generalisability of the estimates of cost and benefit presented here. Given the uncertainty that is associated with estimates of cost and benefit in all economic evaluations, a further element of the chapter deals with how to deal systematically with uncertainty.

Section 5.2 of the chapter discusses the limitations of simple CEA and Section 5.3 reviews the methods of CUA. Section 5.4 details the methods used in the CUA of AH versus TCRE, and Section 5.5 presents the results. Section 5.6 provides a discussion of the results and their implications for clinical and purchasing policy, and of the methodological issues associated with CUA in this area; and Section 5.7 offers some conclusions.

5.2   The limitations of cost-effectiveness analysis

As a form of economic evaluation, CEA has some important limitations. The first of these is that the measure of effectiveness incorporated into the analysis must be uni-dimensional. Drummond *et al* [1987] state that two conditions must hold for a CEA to be appropriate:

> '(a) that there is one, unambiguous, objective of the intervention(s) and therefore a clear dimension along which effectiveness can be assessed; or (b) that there are many objectives, but that the alternative interventions are thought to achieve those to the same extent.' (p.74)

In the case of MAS, it is not clear that these conditions exist, and particularly in the case of the surgical treatment for menorrhagia where there is no *single* objective of treatment. Clearly the amelioration of excessive blood loss is important; but the diffusion of TCRE has taken place with the clear understanding that AH is 100% effective in this regard. As highlighted in Chapter 3, the comparison between TCRE and AH is characterised by clear trade-offs: TCRE results in less post-operative morbidity but is less effective at improving symptoms, and frequently requires women to have further surgery.

The CEA reported in Chapter 3 uses satisfaction rates as its key measure of effectiveness. In principle, it is possible that a patient's satisfaction with surgery reflects their perception of its effectiveness in terms of improving symptoms, whilst allowing for the 'process disbenefits' of treatment. In practice, it is important to question whether dichotomizing this measure into 'satisfied with treatment' (successful treatment) and 'not satisfied with treatment' (unsuccessful treatment) adequately embraces women's preferences regarding the trade-offs between the technologies. Furthermore, the measurement of patient satisfaction in health services research in general has been criticised on various grounds including patients' ability to make technical judgements, the factors that influence their responses to questions, the reliability and validity of instruments to measure satisfaction and the practical use of satisfaction results [Fitzpatrick, 1993].

A second limitation of CEA relates to its usefulness in health care resource allocation. For within-programme resource allocation, CEA has a clear role to play. For example, for a decision maker whose objective is to maximise benefits from a fixed budget in the area of asthma care, information on the incremental cost per additional episode-free day of new treatments may be of value as episode free days may be considered a good composite measure of effectiveness [Sculpher and Buxton, 1993]. For between-programme resource allocation, however, condition-specific measures of effectiveness have limited usefulness. For example, the episode-free day does not represent useful information for a

decision maker considering whether to allocate additional funding to asthma care or to surgical oncology, as the measure has little relevance to the latter area.

Satisfaction rates do have generic characteristics, in that it is possible to ask patients or service users across various programmes whether or not they are satisfied with care. However, satisfaction will mean quite different things to different people in different clinical contexts: for example, satisfaction with gynaecological surgery for a benign condition cannot be assumed to be of equal *value* to satisfaction with treatment for a life-threatening illness.

The cost-effectiveness ratio relating to the economic comparison of AH and TCRE, estimated in Chapter 3, therefore has limitations. Firstly, it is not clear whether, as a simple measure of effectiveness, satisfaction can fully represent women's preferences regarding the inherent trade-offs between the technologies. Secondly, the relevance of the cost per successful treatment for health service resource allocation is likely to be limited to narrow questions concerning funding treatments for menorrhagia, rather than broader between-programme and between-specialty issues.

5.3   Cost-utility analysis

The limitations of CEA in many contexts have encouraged the development of fuller forms of economic evaluation. Cost-benefit analysis (CBA) is grounded firmly in the principles of welfare economics [Mishan, 1971A], and has been used widely in economic appraisals relating to transport and the environment [Jones-Lee, 1976; Pearce *et al*, 1989]. In the field of health care, however, the use of CBA has been hampered by the need to value health benefits in monetary terms [Pauly, 1995]. In the 1960s and 1970s, controversial benefit valuation methods were adopted in health, based on the human capital method [Becker, 1964], but the limitations of this approach [Mishan, 1971B] resulted in few genuine CBAs being published [Backhouse *et al*, 1992]. In recent years, CBA has experienced something of a renaissance, with the use of stated preference

(willingness to pay) methods to value health benefits [O'Brien and Viramontes, 1994; O'Brien *et al*, 1995; Donaldson *et al*, 1995; Chestnut *et al*, 1996]. However, these methods have not been used widely in full economic evaluations and, although promising, require further research [Johannesson, 1993; Arrow *et al*, 1993].

Given the difficulties in applying CBA in the field of health care, CUA has developed as a means of more fully reflecting the outcomes of health care in the evaluation calculus. Although described in general texts as a separate form of economic evaluation [Drummond *et al*, 1987; Luce and Elixhauser, 1990; Robinson, 1993], CUA was originally conceived as a particular example of CEA [Weinstein and Stason, 1977].

CUA is characterised by the use of a generic measure of benefit which embodies the impact of a technology on both health-related quality of life (HRQL) and life expectancy, and the trade-offs between the various dimensions of HRQL and between HRQL and life expectancy [Williams, 1985]. This dual impact and trade-off characteristic has traditionally been expressed in terms of the quality-adjusted life year (QALY) [Loomes and McKenzie, 1989]. As a generic measure of benefit, the QALY can, in principle, be used to compare the cost-effectiveness of technologies across health care programmes, in terms of their incremental cost per additional QALY [Maynard, 1991]. This form of economic evaluation has been used to assess a range of health care technologies [Gerard, 1992].

The standard approach to CUA is illustrated in Figure 5.1 in the form of a QALY profile. The curves show the impact of an intervention on duration of life along the horizontal axis. Over time, patients move between health states which are associated with varying levels of HRQL, and a patient's duration in a given health state is quality-weighted according to the relevant level of HRQL. The multi-dimensional nature of HRQL [Patrick and Erickson, 1993] is dealt with by the various dimensions being valued on a single utility, preference or valuation scale,
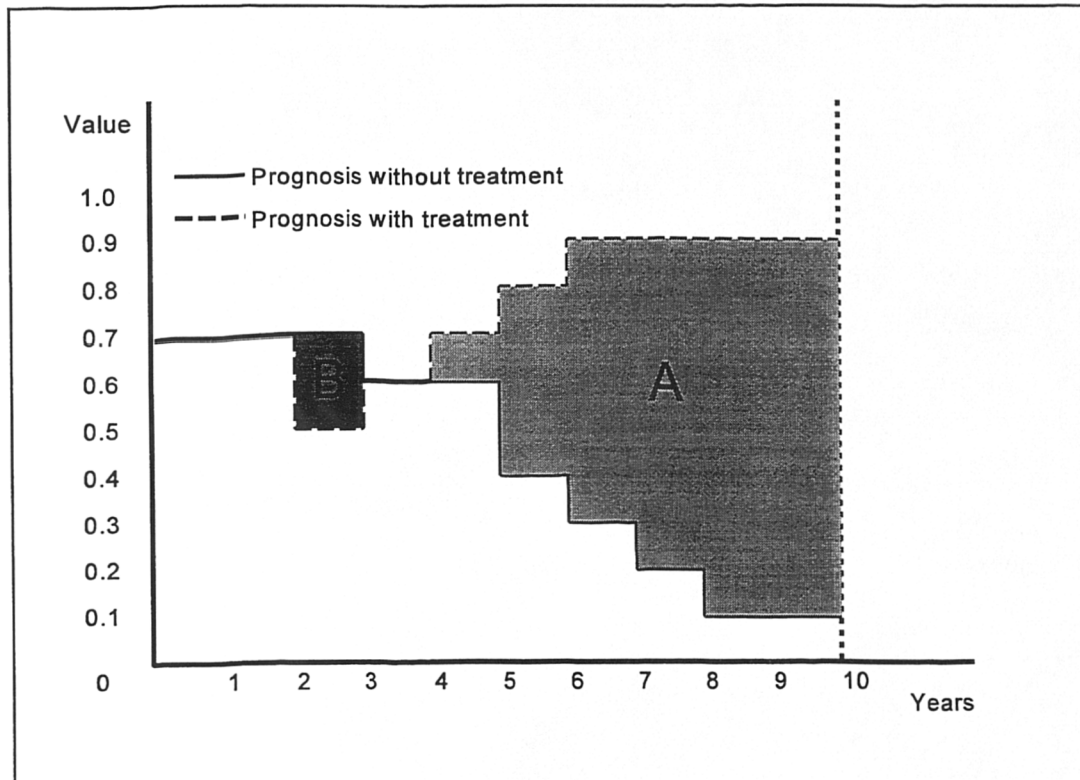
Figure 5.1   **The QALY profile showing prognosis with a hypothetical treatment versus no treatment. The effect of treatment on patients' duration of life is shown on the horizontal axis, and the vertical axis shows the quality weightings based on a 0 to 1 valuation scale. The areas between the two profiles show the difference in QALYs as a result of treatment. The area marked B shows the initial reduction in QALYs as a result of treatment, perhaps due to adverse events. The area marked A shows the larger increase in QALYs as a result of treatment.**

between 0 (equivalent to death) and 1 (equivalent to good health), and has been taken as having cardinal measurement properties. Not only does valuation overcome the problem of comparing alternative dimensions of HRQL, in principle it also provides a means by which the preferences of key groups, such as patients or the general public, can be incorporated into the measure of the benefit of health care interventions. In Figure 5.1, the areas under the QALY profiles represent the QALYs associated with the interventions; the difference in areas between the profiles is the additional QALYs generated by the more effective technology.

One of the earliest examples of a CUA of alternative health care interventions was Stason and Weinstein's evaluation of therapy for hypertension [Stason and Weinstein, 1977]. Seen as a simple extension to CEA, their evaluation used assumptions for the health state valuations (quality weights). Since then, a range of techniques has been used to attach values to health states [Torrance, 1986]. Although there has recently been a consideration of the assumptions underlying the QALY [Loomes and McKenzie, 1989], and a debate, at a theoretical level, about alternative ways of measuring the benefits of health care to reflect society's (or some other group's) preferences [Gafni, 1989; Gafni *et al*, 1993; Johannesson *et al*, 1993; Culyer and Wagstaff, 1993], the measure of benefit invariably used in CUA is the standard QALY. [Chapter 7 reviews this literature more fully.]

The standard QALY model shown in Figure 5.1 can also be expressed as in Equation 5.1 below:

$$QALYs = H(Q).T \qquad (5.1)$$

where H(Q) is the value function and T is life years [Johannesson *et al*, 1993].

The economic evaluation of TCRE in comparison with AH can be usefully extended by undertaking a CUA of the interventions. The two weaknesses of CEA in this clinical context can, in principle, be addressed by expressing the benefits of the treatments in terms of QALYs. The multi-dimensionality of, and the inherent trade-offs between, the outcomes of the two forms of surgery, can be overcome with CUA, by the valuation of HRQL on a single scale and the synthesis of these data with those relating to patients' duration in relevant health states. Moreover, by evaluating the technologies in terms of QALYs, their relative value for money can be compared with that of other technologies within and outside gynaecology.
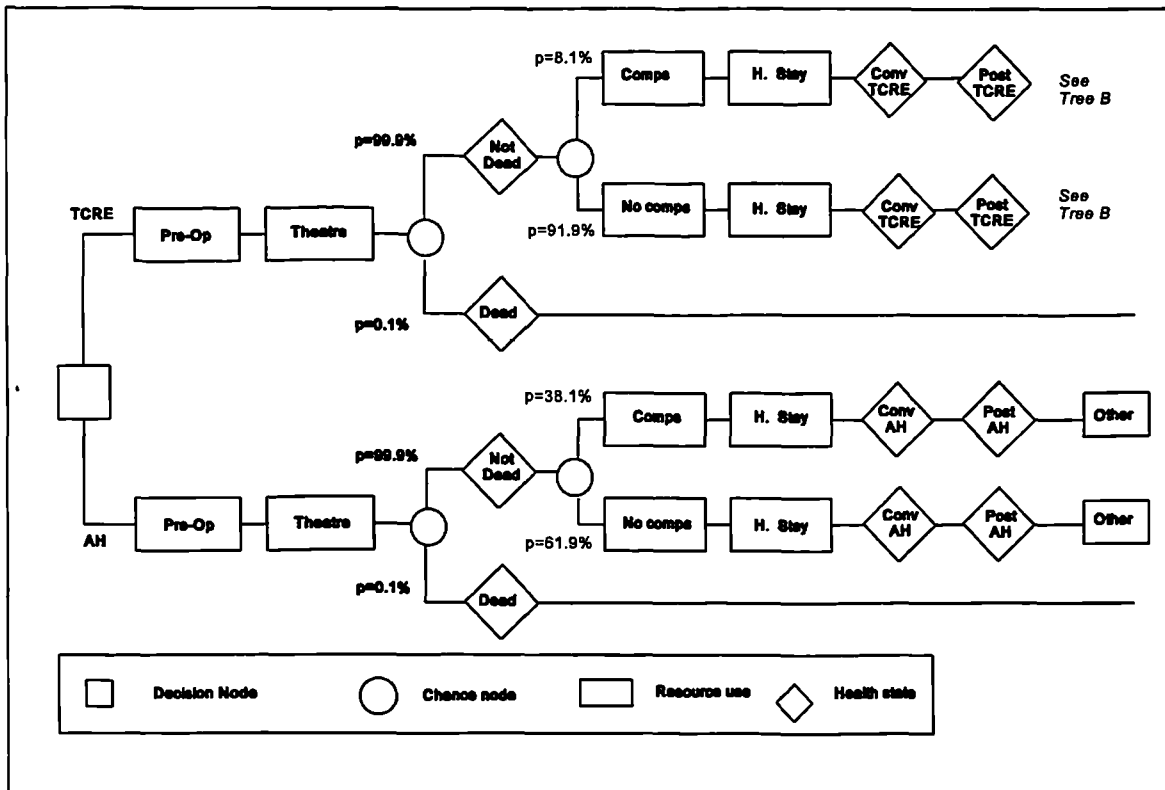
Figure 5.2 The first part of the decision tree model representing the initial surgical intervention and convalescence (Decision Tree A). Abbreviations: pre-op = pre-operative; comps = complications; h.stay = hospital stay; conv = convalescence.

5.4 Methods

The CUA described here is made up of six elements: the decision analytical model; health state description; health state valuation; QALY estimation; resource costs; and dealing with uncertainty. Each element of the analysis is described in detail below.

5.4.1 The decision analytical model

Model structure. The main source of data for the CUA is the Bristol RCT comparing AH and TCRE described in Chapter 3. However, to move to the more detailed level of analysis that CUA represents, some additional data are required. To provide a framework within which to synthesise data from all sources, a decision tree model has been developed, which is illustrated in two parts in
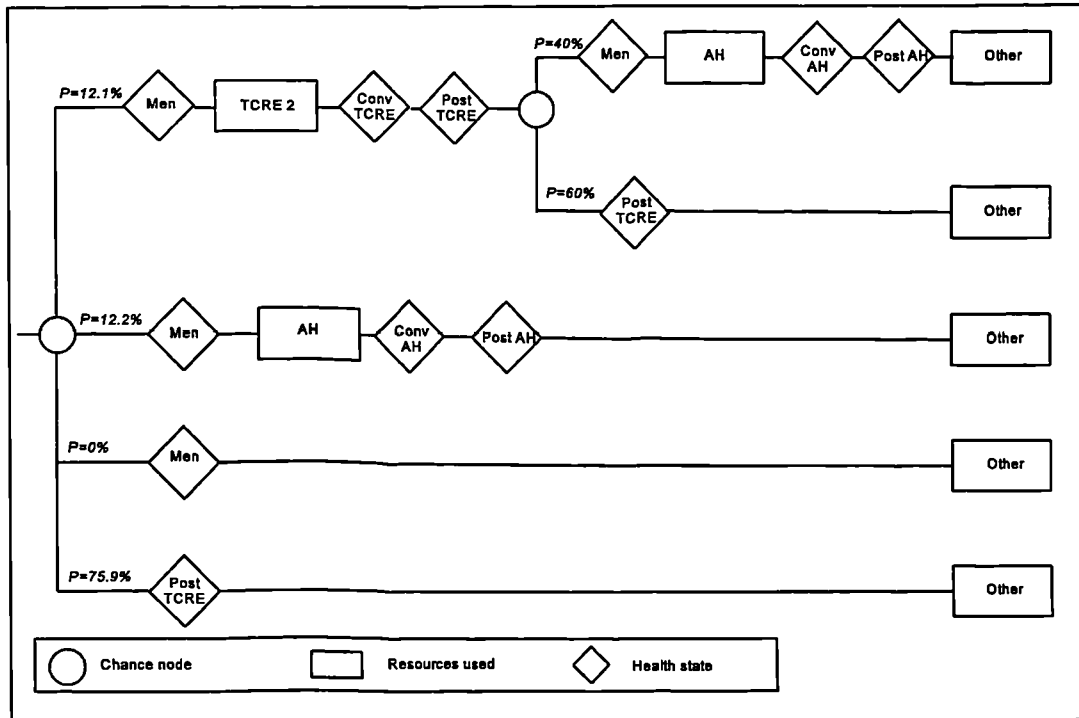
Figure 5.3 **The second part of the decision tree model representing the subsequent prognosis after initial surgical intervention (Decision Tree B). Abbreviations: men=menorrhagia; conv=convalescence.**

Figures 5.2 and 5.3. Figure 5.2 shows Decision Tree A, which represents the period of the initial surgical intervention and subsequent convalescence; and Decision Tree B in Figure 5.3 shows the period after initial surgery for women initially undergoing TCRE.

Following the usual convention, decision nodes and chance nodes are shown as boxes and circles, respectively. In addition, rectangles within branches signify resource use and diamonds represent health states. In Decision Tree A, for example, a patient initially undergoing TCRE will consume pre-operative resources and theatre resources; there is a risk of operative mortality and of complications which will result in resource use; all patients consume ward-related resources during their hospital stay and experience a convalescence; once they have recovered from the operation, women will move into the Post-AH or Post-TCRE health state, depending on which operation they have undergone; over time, women may consume other related resources such as hormone

120

replacement therapy. In the base-case analysis, the model considers the two-year period after surgery, which is broadly equivalent to the average period of follow-up of women in the Bristol RCT.

Decision Tree B relates solely to women initially undergoing TCRE. This shows that women can 'fail' on treatment (ie. their menorrhagia returns). Treatment failure may result in re-treatment, a repeat TCRE or an AH. A repeat TCRE could be followed by further failure and an AH. The risk of operative mortality and of complications with subsequent re-treatments for women initially undergoing TCRE is allowed for in the model although this is not shown explicitly in Decision Tree B.

Taken together, the two sub-trees provide a number of possible pathways through which a woman can pass over two years. These pathways differ according to the risk and number of treatment failures, and, consequently, health status, and the number of additional treatments. For each pathway, estimates of cost and benefit are generated.

Model probabilities. The probability data used in the CUA are shown in Figures 5.2 and 5.3 and relate to the branches coming out of each chance node. These data are taken largely from the Bristol RCT. Although no procedure-related deaths occurred in the Bristol trial, this may have been because the trial sample was too small to detect such a rare event. Other studies have indicated that the mortality risk associated with AH is approximately 0.1% [Dicker *et al*, 1982]. As yet, insufficient data have been collected on TCREs to estimate its mortality risk adequately. Therefore, in the base-case analysis, it is assumed that the mortality risk per procedure is 0.1% for both TCRE and AH. The implications of this assumption are tested using sensitivity analysis.

The probability of complication is based on the proportion of women in the Bristol RCT who experienced any of the operative or post-operative complications detailed in Table 3.2 in Chapter 3. The longer-term follow-up data

from the Bristol trial have permitted the calculation of treatment failure
probabilities allowing for differential follow-up using life table methods (see
Chapter 3). The failure probabilities used here are based on failure rates until
two years. The decision tree allows for the possibility of a woman failing initial
TCRE but not receiving additional surgery, because a woman may rather put up
with her symptoms than experience the disbenefits of further surgery. It was
not possible, prospectively, to ascertain which women firmly fell into this group
in the Bristol trial. Therefore, in the base-case analysis, it is assumed that 0% of
women experience this sort of failure. The implications of this assumption are
tested using sensitivity analysis.

5.4.2 Health state descriptions

In order to estimate the relative benefits of TCRE and AH in terms of QALYs, it is
necessary to value the outcomes of the two technologies on the standard 0 to 1
valuation scale; that is, to estimate quality weights. To achieve this, those
health states relevant to menorrhagia and its treatment need to be described and
then valued by a sample of individuals from an appropriate population.

Identifying the relevant health states. The standard QALY model splits
the outcomes of treatment into discrete health states. As regards the
comparison of TCRE and AH, this involves identifying the key health states that
constitute the possible prognoses of these two types of surgery. For those
health states that may recur, in a broadly similar form, over time (eg.
menorrhagia and convalescence after TCRE), the standard QALY model has a
single description and value, however the state is sequenced - whatever the
duration of the state.

As a result of the inevitable variation between patients in the process and
outcome of surgery, there are numerous possible health states associated with
the treatment of menorrhagia, and it is unlikely to be feasible to describe and to
value each of them within a CUA. One means of simplification is to use a
clinical trial to ask patients, in effect, to describe their *own* health at various

points in time using a standardised descriptive classification of health status like the EuroQol instrument [Brooks, 1996] which is linked to a tariff of externally-generated values. This has the potential advantage of capturing information on patients at a number of time points which may coincide with key clinical events. Although this CUA of AH versus TCRE is based largely on data from a RCT, the latest and validated version of the EuroQol was not available at the outset of the trial, and no other valuation system was considered appropriate. So the use of a standard descriptive and valuation system was not feasible in this analysis. Furthermore, there are factors which might lead to an alternative approach to identifying health states for a CUA. For example, in certain clinical contexts, the descriptive system offered by generic valuation systems may lack sensitivity to differences between, and changes within, patients in underlying health status [Cook and Richardson, 1993]. A number of CUAs have, therefore, selected study-specific health states as a way of representing the outcomes of interventions [Mohide *et al*, 1988; De Haes *et al*, 1991; Hall *et al*, 1992; Sculpher *et al*, 1996B], and this was the approach adopted here.

In the context of the comparison of AH and TCRE, it has been necessary to identify key health states related to the surgical management of menorrhagia that can be described and subsequently valued. These health states should relate to the key phases of the typical patient's prognosis following surgery, but should be manageable in number: the more health states, the greater the burden of the valuation task and the more complex the modelling.

For the purposes of this CUA, five health states have been identified, as detailed below.

(a)   *Menorrhagia ('Men' in Figures 5.2 and 5.3).* For CUA of acute conditions, patients' health status prior to treatment would not need to be established as a distinct health state because it represents a common baseline for each comparator. However, for women receiving any treatment other than hysterectomy, menorrhagia can be viewed as a

chronic condition with the potential to represent a burden on the HRQL for some years. In particular, because there is a risk of treatment failure following TCRE, women may again experience the condition some months or years after initial surgery. Hence, it is necessary for menorrhagia to represent a distinct health state within this CUA. It has been assumed that, if women fail on TCRE, the health state they return to is identical to the one they experienced prior to surgery.

(b)   *Convalescence following TCRE ('Conv TCRE' in Figures 5.2 and 5.3).* A major difference between TCRE and AH is in convalescence following surgery. This difference is manifested not only in the duration of convalescence, but also in the HRQL associated with the phase. It is, therefore, necessary for convalescence following TCRE to be a separate health state within the CUA. Given the possible need for re-treatment following a TCRE, this period of convalescence might be experienced twice or more by women.

(c)   *Convalescence following AH ('Conv AH' in Figures 5.2 and 5.3).* The rationale for including convalescence following TCRE as a distinct health state also applies to convalescence following AH. Of course, a woman can experience this health state only once in her prognosis.

(d)   *Pre-menopausal following recovery from successful TCRE ('Post-TCRE' in Figures 5.2 and 5.3).* Once a woman has recovered from the convalescence following a TCRE, she may enter one of many health states. These will differ according to the amount of bleeding and menstrual pain she experiences and the impact on her functional status. Within a CUA, it is necessary to simplify these health states and here it is assumed that a woman's prognosis after TCRE can either be unsuccessful (in which case she reverts to the state of menorrhagia (see (a) above)), or successful. This health state, therefore, describes the typical woman's

HRQL after TCRE, when there has been an improvement in symptoms.
The health state focuses only on the period prior to the menopause.

(e)   *Pre-menopausal following recovery from AH ('Post-AH' in Figures 5.2 and*
       5.3). As for TCRE, there is a large number of health states a woman can
       experience. To keep the valuation exercise and CUA as manageable as
       possible, this health state describes the typical HRQL of a pre-menopausal
       woman after undergoing AH, and having passed through the
       convalescence phase.

Describing health states. Various approaches have been adopted for
describing health states within CUA [Llewellyn-Thomas *et al*, 1984; Froberg and
Kane, 1989A; Furlong *et al*, 1990; Gerard *et al*, 1993]. Decisions have to be
taken about whether the health states are described in point form phrases or
narrative paragraphs; whether they will be covered in the first, second or third
person, or in summary form; how much information will be provided; whether to
adopt a holistic design, where a descriptive scenario represents a combination of
many attributes, or a decomposed design, which allows specific attributes within
health states to be analysed separately; how to frame the descriptive scenario
(eg. whether to detail the probability of dying or the probability of surviving); and
whether to use disease labels within the scenario.

Some empirical work has shown that these decisions influence how individuals
value health state scenarios. Llewellyn-Thomas *et al* [1984] found values varied
considerably according to whether scenarios were written in the first person
singular and were full in detail, or in point-form with only the most severe health
problems detailed. McNeil *et al* [1982] found that values associated with
different treatments for lung cancer were influenced by whether the treatments
were mentioned, and by whether probabilities were couched in terms of living or
dying.

Other studies have found that values are not sensitive to how the health state scenarios are described. O'Connor *et al* [1987] found that the way information was presented to cancer patients on the side effects of a hypothetical drug did not influence their values. Gerard *et al* [1993] found that the labelling and framing methods of scenarios relating to breast cancer had no significant effect on the values provided by a convenience sample of women.

Although there is uncertainty in the literature both about the effects of alternative ways of describing health states on subsequent valuations, and about the implications of any such effects for the appropriateness of different descriptive methods, some authors have produced guidance for developing scenarios based on empirical work. Furlong *et al* [1990] suggest the following guidelines, which are discussed below in relation to the valuation exercise undertaken in the CUA of AH versus TCRE.

(a)   *The descriptions should be comprehensible for the intended respondents.*
       The choice of women with menorrhagia as the sample of valuers is likely
       to satisfy this point (see below).

(b)   *All important aspects of the health state should be explicitly included.*
       The process of selecting the information for descriptive scenarios in this
       study was made up of the following elements.

       (i)   A set of basic general attributes of HRQL relevant to menorrhagia
              and its treatment was selected including pain, social function,
              vitality, emotion and sexual function. In addition, some condition-
              specific information was included. These attributes and condition-
              specific details were identified from a review of the published
              literature on the HRQL implications of menorrhagia, and from
              access to databases relating to its treatment from three major
              prospective studies: the Bristol RCT comparing AH and TCRE
              [Dwyer *et al*, 1993] (see Chapter 3); the Oxford Treatment for

126

Menorrhagia Study [Peto *et al*, 1993; Coulter *et al*, 1994A and
1994B; Jenkinson *et al*, 1994]; and the North West Thames
Hysterectomy Study [Clarke *et al*, 1995]. In addition, information
was identified from videos of two focus groups organised by the
King's Fund Centre for Health Services Development, which
provided a valuable source of data on the implications of
menorrhagia and its treatment for women's HRQL. A set of draft
descriptive scenarios was developed using these sources of
information.

(ii)   The draft scenarios were presented to a group of gynaecologists
         and health service researchers with extensive experience in the
         area of menorrhagia. One individual had, in the course of her
         research, interviewed 80 women with menorrhagia, or whom had
         been treated for menorrhagia, to assess their perceptions of the
         condition and its treatment. On the basis of the comments of this
         group, a set of revised scenarios was developed.

(iii)   A pilot study was undertaken on a convenience sample of 20
         women. One objective of the pilot study was to assess women's
         reactions to the revised scenarios: whether they understood them
         and found the language appropriate. On the whole, the women in
         the pilot study commented favourably on the scenarios, but some
         small changes were made, resulting in the descriptive scenarios
         used in the valuation exercise.

5.4.3 Health state valuation

A key characteristic of CUA is that the health states through which a typical
patient passes before, during and after some form of health care intervention, are
valued. In this context, valuation means that each health state is 'quality- or
preference-weighted' on a 0 to 1 scale. In order for values to be used to quality-

adjust information on life expectancy to arrive at QALYs, the 0 on the scale accords with death and the 1 with good health.

Valuation instrument. A range of alternative valuation instruments has been used in CUA [Torrance *et al*, 1986; Patrick and Erickson, 1993]. The four most frequently used, however, are category rating, magnitude estimation, standard gamble and time trade-off [Gerard, 1992]. The choice between these methods is far from straightforward. In principle, magnitude estimation, where respondents compare a health state with some standard and report the extent to which it is 'better' or 'worse', produces a ratio scale, but this has not been adequately tested [Mulley, 1989]. Category rating, which involves asking respondents to complete a visual analogue scale where each step is taken as being an equal interval, has the advantage of being easy to understand on the part of the respondent. However, there are doubts about its ability to generate the genuine interval scale required for CUA [Nord, 1991], and concerns about respondents distributing their values across the full scale [Mulley, 1989].

Some economists argue that the standard gamble valuation technique should be used in CUA, as it has the soundest theoretical basis [Gafni, 1994]. With its origins in von Neumann and Morgenstern's [1944] work on utility theory, underlying the standard gamble is a set of axioms based on a normative theory of decision making under uncertainty. As well as a strong theoretical underpinning, it is argued that values generated by the standard gamble embody respondents' attitude to risk because the values are elicited in a risky context [Gafni, 1994]. Although the standard gamble is considered by many to be the gold standard for eliciting health state values for CUA, it has disadvantages. These include the fact that the normative theory of behaviour which underlies it has been shown to be an inadequate explanation of how individuals actually make decisions under uncertainty [Schoemaker, 1982]. Furthermore, respondents often find it difficult to use. The limitations of the standard gamble are discussed further in Chapter 7.

The time trade-off (TTO) valuation technique was developed by Torrance [1972]
as a means of avoiding some of the difficulties respondents have with the
standard gamble, while retaining its forced choice element. Although it can be
criticised because of the absence of a strong theoretical underpinning, the TTO
has been used widely in studies in Canada [Torrance, 1976; Mohide *et al*, 1988]
and the UK [Buxton *et al*, 1987; Daly *et al*, 1993; Sculpher et al, 1996B].
Evidence exists to suggest that the TTO produces very similar values to the
standard gamble [Krabbe *et al*, 1996].

Although the two forced choice-based instruments, the standard gamble and
TTO, have been shown to have acceptable intra rater reliability and
reproducibility [Froberg and Kane, 1989B], it is very difficult to validate these
techniques, as no true gold standard means of valuation exists. The criterion
validity of some techniques has been explored. Torrance [1987] showed that
the criterion validity of the TTO against the standard gamble was adequate; most
valuation methods have been shown to generate values consistent with expected
direction in types and severity of illness [Patrick and Erickson, 1993].

On the practical level, there seems to be some evidence pointing towards the
superiority of the TTO. Torrance [1976] concluded:

> '... this study points to the time trade-off method as the best of the three
> (standard gamble, TTO and category rating) tested for use on the general
> public in the measurement of social preferences for health states.'
> (p.135).

As a preliminary to the largest health state valuation exercise undertaken in the
UK, the standard gamble and TTO were compared within-respondent on the
basis of 335 interviews with a sample of the general public [Williams *et al*, 1995;
Dolan *et al*, 1996]. Five criteria were used to compare the two forms of
valuation: completeness, logical consistency, concurrent validity, discriminant
validity and test-retest reliability. Against these criteria, the study found that the

standard gamble and the TTO had little to choose between them, but the TTO was selected for the large-scale survey because it resulted in more complete data and more consistent valuations at the individual level.

These considerations, both theoretical and practical, have prompted the use of the TTO in the current study.

   Sample of respondents. Values were elicited from a sub-sample of the women referred to St Michael's Hospital in Bristol with uncomplicated menorrhagia who took part in the survey detailed in Chapter 4. Of the women referred to the hospital and who completed the questionnaire, a target sample of 60 women for interview was established. Chapter 4 provides full details of the sample from which the women interviewed was drawn. In brief, potentially eligible women were identified from GP referral letters received by the hospital between January and October 1994. These women were sent a letter explaining the study and a questionnaire to complete. The letter also asked if they were willing to be interviewed by a trained female interviewer prior to their visit to the out-patient clinic.

If women responded positively to the invitation to be interviewed, they were contacted by telephone to arrange a convenient time and day. Women were told they would be interviewed in their homes unless they preferred to come to St Michael's Hospital. They were excluded from valuation exercise if, on the basis of available data including the information provided in the questionnaire, they had significant concomitant illness; they lived too great a distance from St Michael's Hospital to make an interview practicable; or interview prior to their hospital appointment was not feasible.

Interviews were undertaken by two trained female researchers. The interview schedule consisted of three elements.
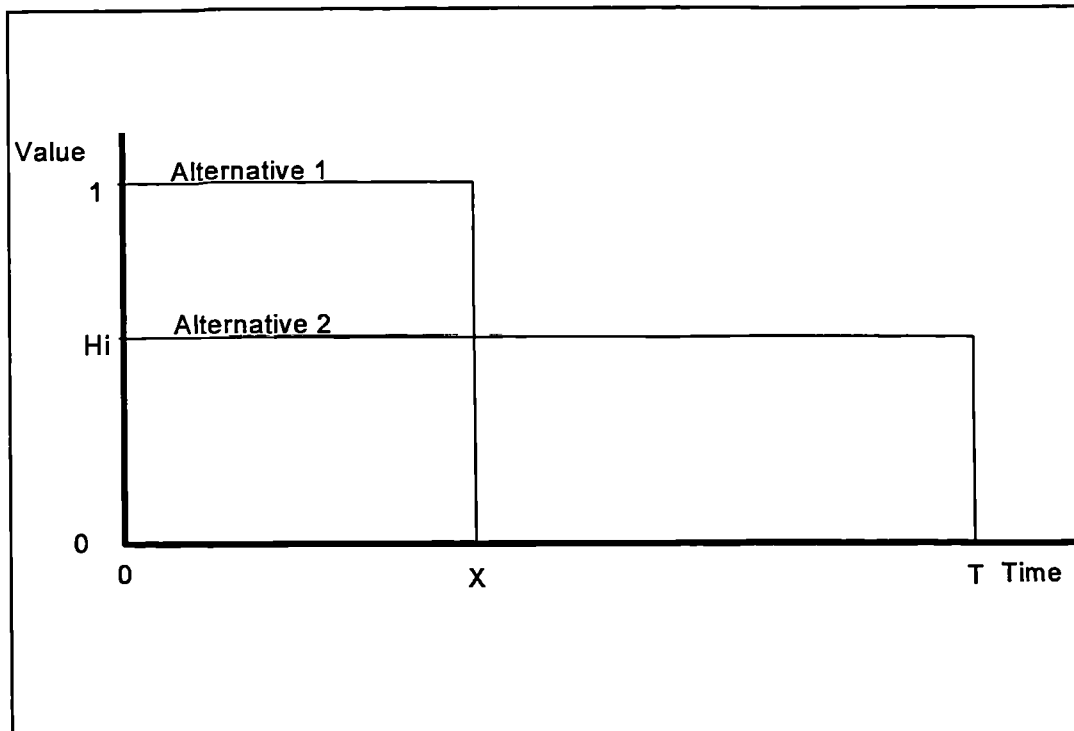
Figure 5.4   The time trade-off instrument used to value chronic health states
              considered better than death. The intermediate health state i, which is
              likely to be considered better than death, is valued relative to good
              health and death. Respondents are asked to consider a chronic health
              state i and to imagine spending the rest of their life (Time T) in that state
              (Alternative 2). They are asked to compare that situation with one
              where they spend a shorter period (X) in good health (Alternative 1).
              The time period X is varied until the respondent is indifferent between
              Alternative 1 and Alternative 2. The value of health state i is then
              calculated as x/t. [Source: Torrance, 1986].

(a)   *Introduction.* During this section of the interview the researcher

introduced herself and the nature of the exercise. She also asked the

respondent for permission to tape the interview. Finally, a short

additional questionnaire was given to the woman to complete, asking a

series of socio-demographic questions.

(b)   *Valuation of chronic health states.* The next stage of the interview

consisted of respondents ranking and valuing the chronic health states.

Written scenarios were presented to respondents on cards describing

states (a), (d) and (e) detailed in Section 5.4.2 above: menorrhagia; pre-

menopausal following recovery from successful TCRE; and pre-
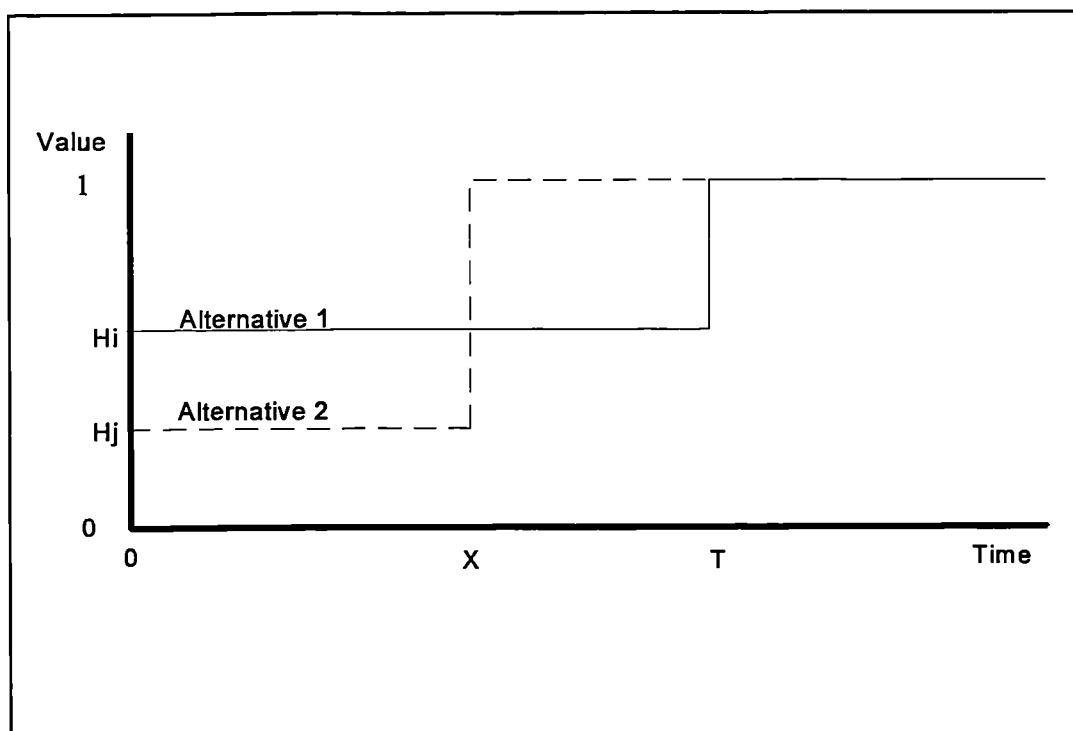
Figure 5.5   The time trade-off instrument used to value temporary health states.
The valuation of temporary health states is made up of two stages. In
the first stage, an intermediate temporary health state i is valued relative
to the best state (good health) and the worst temporary health state j.
The respondent is offered a choice between health state i lasting for the
duration of the temporary states (T) followed by good health (Alternative
1), and health state j for a shorter duration (X) followed by good health
(Alternative 2). The time period X is varied until the respondent is
indifferent between the two alternatives, and the value of the temporary
state i is calculated as 1-x/t, if the value of state j is assumed to be 0.
The second stage of the process involves re-scaling the value of health
state i onto the standard 0 to 1 scale. This involves valuing the worse
temporary health state j as a short duration chronic state, where the
duration is the same as the temporary states, as shown in Figure 5.4,
and calculating the value of state i using the following formula: $Hi = 1-(1-Hj)(X/T)$. [Source: Torrance, 1986].

menopausal following recovery from AH. In addition to these three
descriptive scenarios, a fourth card reading 'your health state today' was
shown to women. The respondents were then asked to rank the health
states described in the scenarios.

Following ranking, the TTO technique was carefully explained to
respondents. The three health states described in the scenarios were

considered sufficiently long-term in duration (about 10 years, on average) to value using the standard TTO technique for chronic states illustrated in Figure 5.4. To avoid anchoring bias, the 'converging ping-pong' approach to the TTO was adopted, whereby the period in the particular state being valued is systematically altered over the interval of the respondent's life expectancy on a high-low basis, converging inwards, until indifference is established [Mohide *et al*, 1989].

(c)   *Valuation of temporary health states.* The two other health states listed in Section 5.4.2 above - convalescence following TCRE (b) and convalescence following AH (c) - were valued as temporary health states, using Torrance's [1986] two-stage technique, as illustrated in Figure 5.5. Respondents were asked to rank the two temporary states. The preferred health state was then valued, relative to the worst, for the duration of the temporary health states, which was assumed to be 10 weeks.

The second stage of the valuation technique for temporary health states is required to translate the value of the temporary state onto the standard 0 to 1 scale. To achieve this, the least preferred of the two temporary states is valued as a short duration chronic state, where the duration is not the respondent's life expectancy as for the chronic states, but the duration of the temporary health states (10 weeks). It was felt that, faced with the choice between the worst temporary health state for 10 weeks followed by death and a shorter period in good health, the short life expectancy would dominate the details of HRQL in a woman's response. Hence, following Cook *et al* [1994], the worst temporary health state was valued as a short duration chronic state twice with two different durations: 10 weeks (Approach I) and 10 years (Approach II). Therefore, two alternative sets of values have been calculated for the two temporary health states. It was planned that, if there were no clear differences between the two approaches, Approach I would be used because of its theoretical advantages.

Table 5.1   Duration of convalescence by type of surgery, and times until treatment for
women failing on initial TCRE. Figures are means (standard errors)*

Duration	TCRE	AH
Duration of convalescence (weeks)		
With surgical complications (0.64)	4.71 (0.68)	10.65
Without surgical complications (0.70)	2.32 (0.15)	11.63
Time until re-treatment (months)†		
First repeat TCRE	7.4 (1.54)	-
Hysterectomy following one repeat TCRE	14.9 (5.81)	-
Hysterectomy without prior repeat TCRE	10.4 (2.83)	-

*   All data taken from Bristol RCT (see Chapter 3). Mean estimates are used in the base-case
analysis, information on standard errors is used in the sensitivity analysis.

†   Months after first TCRE.

5.4.4 QALY estimation

As illustrated in Figure 5.1, within the standard model, QALYs are estimated by
multiplying the health state value by the length of time a patient spends in that state,
and summing across health states over the time span of the evaluation. Therefore, for
the possible pathways through which a woman can pass over two years shown in
Figures 5.2 and 5.3, assumptions are made about which health states a woman
experiences and for how long.

The duration of time in a given health state is, as far as possible, taken from the
Bristol RCT. The period of convalescence is assumed to be the mean time
women in the Bristol trial, who were in employment, reported that it took them
to return to work following surgery, allowing for whether they experienced any
complication. For women initially undergoing TCRE, there is a risk of treatment
failure and the need for additional surgery. The trial provides data on the risk of
failure and the time period until re-treatment(s), but it was not feasible to collect
data on *when* a woman's health status deteriorated to the extent that she
sought further surgery. In the base-case analysis of the model, it is assumed
that, if a woman undergoes re-treatment, she experienced half the period

between recovery from initial surgery and re-treatment in the 'pre-menopausal following recovery from successful TCRE' state, and half that period in the 'menorrhagia' state.

For all women undergoing a hysterectomy (whether initially or as a re-treatment following initial TCRE), they are assumed to remain in the health state 'pre-menopausal following recovery from AH' until two years. In the base-case analysis, it is assumed that, as in the Bristol trial, women's average age is 41 years. Table 5.1 details the duration of convalescence for the two groups according to whether there were complications with initial surgery, and times until re-treatment for women failing initial TCRE.

It is standard practice in economic evaluation to discount future costs from health care interventions to reflect individuals' and society's time preference [Drummond *et al*, 1987; Weinstein and Stason, 1977]. Until recent years, it was widely accepted that the benefits generated by health care programmes in future years should also be discounted, usually at the same rate as costs [Keeler and Cretin, 1983]. Recently, however, there has been a debate about whether it is appropriate to use a positive discount rate for benefits [Parsonage and Neuberger, 1992; Cairns, 1992]. Although Chapter 7 includes some consideration of the validity of the standard approach to discounting benefits in economic evaluation, this thesis does not consider the methods of discounting in detail. Department of Health guidelines on discounting health benefits expressed in natural units show some inconsistency: joint guidelines with the Association of the British Pharmaceutical Industry recommend that two approaches be adopted: a 6% discount rate and a 0% rate [Association of the British Pharmaceutical Industry/Department of Health, 1994]; but guidelines for public policy appraisal in the field of health suggest a rate of 1.5% to 2% for these benefits [Department of Health, 1995]. Given the uncertainty that currently exists in this area of method, the following approach is taken here. In the base-case analysis, future QALYs are discounted at 6% per annum. However, the sensitivity analysis considers alternative values for the discount rate including a 0% rate.

5.4.5 Resource costs

The health service resource use and unit costs relating to the surgical procedures and used in the model are at a June 1994 price base, and are based on those estimated from the Bristol RCT and described in Chapter 3. Procedure costs are divided into those which, on average, are incurred when a patient has surgical complications and those which, on average, are incurred when surgery is straightforward. This distinction in procedure costs reflects the difference in length of hospital stay and time in theatre following complicated procedures, as well as the cost of complications in terms of such things as additional disposables, diagnostic tests and drugs. Table 5.2 details the lengths of hospital stay and times in theatre for TCRE and AH, according to whether there were complications associated with the procedure or not. Table 5.3 presents the procedure costs used in the model, again distinguishing between procedures with and without complications. It is assumed that, for the very small proportion of women who die during surgery, the only costs incurred are the pre-operative and theatre costs.

For women initially undergoing TCRE, there is a risk of additional subsequent surgery for their menorrhagia. For re-treatments, the expected cost of a procedure includes the cost of complications weighted by the probability that complications will occur.

The Bristol RCT did not collect data on the proportion of women undergoing cervical screening, so the cost of this area of resource use was not included in the trial-based economic evaluation described in Chapter 3. However, cervical screening generates a NHS resource cost that will differ between women having a TCRE and those undergoing AH. It has been assumed, therefore, that women who retain their uterus are offered screening every five years, and that 74% of women undergo screening [Brown and Sculpher, 1993]. The unit cost of

Table 5.2   Length of hospital stay and time in theatre according to whether women had complications. Figures means (standard errors)*

Resource use	TCRE	AH
Length of hospital stay (days)		
With surgical complications	3.25 (0.70)	6.64 (0.19)
Without surgical complications	1.99 (0.02)	6.19 (0.12)
Time in theatre (minutes)		
With surgical complications	63.75 (3.09)	63.16 (2.23)
Without surgical complications	50.12 (0.97)	62.75 (1.24)

*   All data are taken from the Bristol RCT (see Chapter 3).

cytology, including laboratory costs, is taken as £8 (Dr Flanally, personal communication).

The model also incorporates the cost of women using hormone replacement therapy (HRT). Based on data collected in the Bristol trial at 2.2 years after initial surgery, 10% of women who undergo TCRE and who do not have a hysterectomy are taken as using HRT compared to 17% of women who have a hysterectomy. Whether a woman initially undergoes TCRE or AH, there is a chance she will use other health service resources because of menstrual problems, as shown by the Bristol RCT. Hence an 'other' cost category is added to the two-year costs of therapy based on the Bristol results. All costs occurring after initial surgery are discounted at an annual rate of 6% [HM Treasury, 1991].

Due to the fact that building a model as a framework for economic evaluation requires simplifying assumptions that a RCT-based analysis may not, it would be expected that the model detailed here would generate cost estimates that are slightly different to those presented in Chapter 3, which were taken directly from the Bristol trial. The reasons for this include the fact that the model is using two year treatment failure probabilities for the cost of TCRE, whereas the RCT-based analysis estimated costs based on all failures. Furthermore, HRT rates in the trial

Table 5.3   Procedure costs used in the model (£)

Cost element	TCRE	AH
With complications		
Pre-operative	17	17
Theatre	267	292
Ward	389	795
Complications	68	84
Post-operative	3	7
General practice	4	12
Other	11	10
Total	759	1217
Without complications		
Pre-operative	17	17
Theatre	231	291
Ward	238	741
Post-operative	3	7
General practice	4	12
Other	11	10
Total	504	1078

were analysed on an intention to treat basis; the rate of HRT use incorporated into the model is calculated separately for women who had a hysterectomy, regardless of their randomised group.

5.4.6 Data synthesis and decision rules

The purpose of CUA is to assist purchasers in the decision as to whether a particular intervention or programme is more cost-effective than its comparator, and hence whether it represents a good use of health service resources. The RCT-based economic evaluation of AH and TCRE detailed in Chapter 3 showed that AH is statistically significantly more costly than TCRE, and the model-based analysis presented here will be consistent with this. The main focus of this CUA, therefore, is to assess whether AH is worth purchasing: is its incremental cost justified in terms of any additional benefits?

In order to address this question, CUA synthesises cost and benefit estimates into a ratio which represents the incremental cost per additional QALY of the more costly, but more effective, intervention under comparison. In principle, purchasers can then use this information in comparison with similar ratios generated by economic evaluations of other interventions. This comparison can help to decide whether the additional resources required can be taken from other treatments which have higher incremental cost per QALY ratios, or from an expanding overall budget. The concept of the 'QALY league table' has developed as a possible means of assisting in this decision making process, ranking a range of interventions in ascending order of their incremental cost per QALY [Williams, 1985; Maynard, 1991]. Some authors have criticised a formalised use of league tables on the grounds of the current weakness and variability of the methods used in CUA [Drummond *et al*, 1993B; Mason *et al*, 1993]; others have doubted the value to specific decision makers of league tables which consist of studies from various locations and contexts [Gerard and Mooney, 1993]. However, if CUA is to be of use to health care purchasers, decision rules need to be established involving incremental cost per QALY thresholds below which an intervention would be considered cost-effective.

No such thresholds have emerged in the UK, either by central diktat or consensus. Laupacis *et al* [1992] made tentative suggestions about such thresholds for the Canadian health care system. They argued that various grades of recommendation exist. For example, a Grade B recommendation would suggest that there is 'strong evidence for adoption and appropriate utilisation' if a new technology is more effective and costly than an existing comparator and each additional QALY costs less than (Can) $20,000; a Grade C recommendation would suggest that 'moderate evidence for adoption and appropriate utilisation' exists if a new technology is more costly and more effective that an established one with each additional QALY generated costing no more than (Can) $100,000.

Even if these thresholds are broadly appropriate to Canada, they are unlikely to be sufficiently conservative for the UK, which spends, per head, about between 60% and 70% the amount spent in Canada [OECD, 1996]. Using an exchange rate of £1 to (Can)$2 and adjusting for the lower health care expenditure in the UK by weighting the Canadian values by 65% would imply a lower threshold of £6,500 and an upper threshold of £33,000. These two values have no grounding in policy or empiricism. However, within any economic evaluation presenting results in the form of an incremental cost to effect ratio, it is essential to use some form of threshold in order to handle adequately the uncertainty in the analysis. To assist in the presentation of results and the formulation of broad policy conclusions from the CUA presented here, it is assumed that these two illustrative incremental cost per QALY thresholds are broadly acceptable lower and upper bounds to define cost-effectiveness.

5.4.7 Dealing with uncertainty

Inherent in all economic evaluations is uncertainty about its results and conclusions. Briggs *et al* [1994] identified four types of uncertainty in economic evaluation which relate to data inputs, extrapolation, generalisability and analytical methods.

Data inputs. The trial-based economic evaluation in Chapter 3 considered the implications of variation in stochastic data using standard statistical methods. This form of uncertainty also exists in the modelling undertaken here, in relation to estimates of resource use, probabilities of clinical events and health state valuations. In most economic evaluations, the unit costs of items of resource use are deterministic; that is, they come in the form of point estimates, usually from particular health care facilities, with no sampling variation around them. The major source of uncertainty with unit costs is measurement error as a result of the differences in costing methods between different health care facilities and between different countries. The key uncertain unit cost estimate in relation to the differential cost of AH and TCRE is that of a hospital in-patient day.

Using a series of sensitivity analyses, the implications of variation in data inputs in the model is the particular area of uncertainty considered in this chapter. The objective is to consider how robust the conclusions of the base-case analysis are to variations in the values of particular parameters. To present the results of the sensitivity analyses, a cost-effectiveness (CE) plane is used as shown in Figure 5.6. For a given comparison of two health care interventions, the CE plane plots their cost difference against the difference in their effects (in this case QALYs).

Hence the economic comparison can be summarised in terms of four quadrants. If the plane were used to illustrate the economic comparison of AH and TCRE, the expected location for the comparison would be Quadrant I (AH more costly and more effective) or Quadrant IV (AH more costly and less effective). Although the RCT-data do not support it, at least until two years follow-up, the economic comparison of AH and TCRE could be located in Quadrant III (AH less costly and less effective) or Quadrant II (AH less costly and more effective). If the comparison were located in Quadrant II, the policy conclusion would be clear: AH would be more cost-effective because it dominates TCRE, being less costly and more effective. A similar policy conclusion in favour of TCRE would be valid if the comparison were located in Quadrant IV.

In Quadrants I and III, the concept of a threshold incremental cost per QALY ratio becomes crucial for a policy decision, as purchasers have to decide whether incremental costs are worth incurring in order to generate the additional benefits. The dotted line in Figure 5.6 illustrates a maximum threshold ratio. If the AH-TCRE comparison is located in Quadrant I and the incremental ratio is less than the threshold (ie. in the area marked b), AH would be considered more cost-effective than TCRE. Conversely, if the ratio is greater than the threshold (ie. in
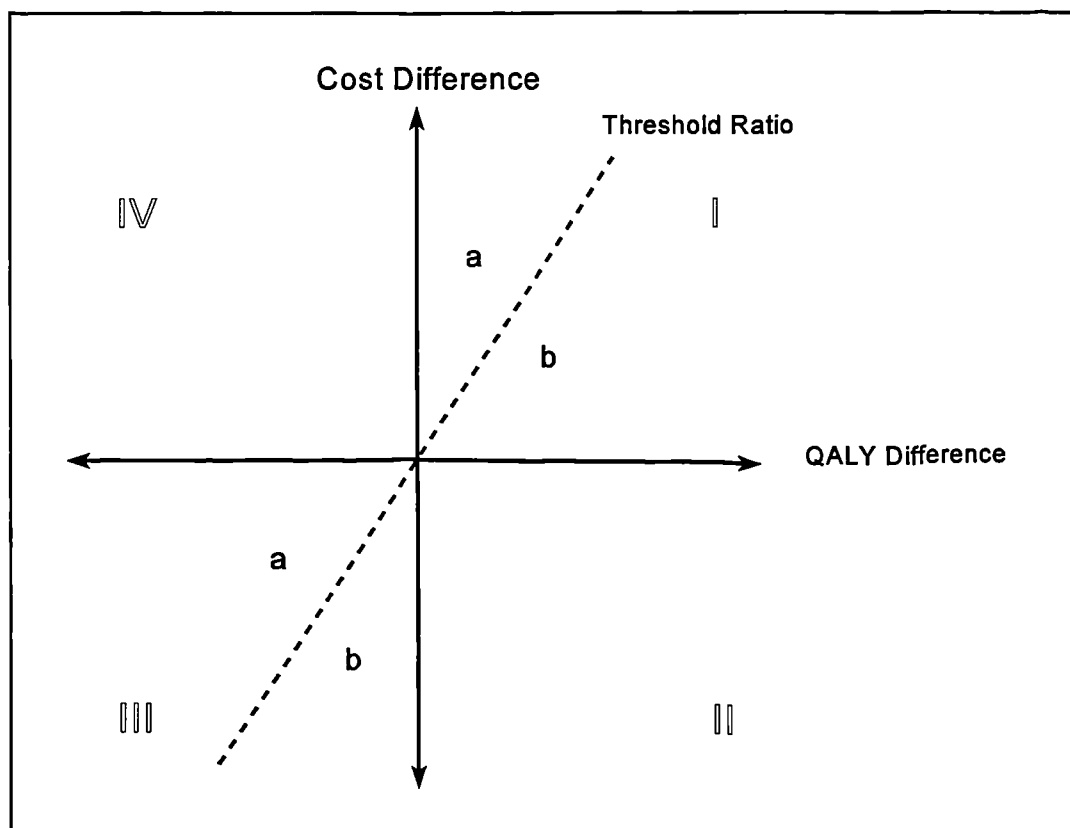
Figure 5.6   The cost-effectiveness plane used to present the results of a cost-utility analysis. (Source: Briggs, 1995).

the area marked a), TCRE would be considered more cost-effective.

Using base-case data, a policy conclusion from the CUA is suggested using the illustrative lower and upper bound threshold ratios described above. The aim of the sensitivity analysis is to assess how robust that conclusion is to alternative values of parameters. In other words, if the values of particular data inputs change, will the economic comparison of AH and TCRE switch quadrant, or alter from (to) area a to (from) area b in Quadrants I or III?

Extrapolation. Uncertainty associated with extrapolation is concerned with the process of trying to make the results of the analysis more comprehensive by moving away from the primary data source. In the context of this analysis, a major limitation imposed by the Bristol RCT as the main source of data, is its short period of follow-up. Ideally, a detailed CUA of these two

interventions would be based on a time horizon of women's entire lifetimes. Such long-term data will not be available for many years. However, given that purchasers require some indication of the likely longer term cost-effectiveness of the two procedures, two simple extrapolations are added to the analysis presented here.

(a)   *Until the menopause.* The first extrapolation models the costs and benefits of treatment until the menopause based on the following assumptions: that women will continue until the menopause in the health state in which they are located at two years; that there are no differences between the treatments in terms of consumption of health service resources over that period; and that the average age at the menopause is 51 years [Luoto *et al*, 1994].

(b)   *Until death.* The second extrapolation models costs and benefits until death using the following assumptions: after the menopause women move into a good health state for the remainder of their life valued at 1.0; that there are no differences between the treatments in terms of consumption of health service resources over that period; and that the average age at death is 80 years.

These simple extrapolations allow a consideration of the potential cost-effectiveness of the two therapies over the longer term.

   Generalisability. Uncertainty relating to generalisability is concerned with the extent to which the results of an evaluation when applied to a particular context (eg. to a specific population, hospital, set of clinicians) hold true when the focus of the evaluation is altered to another context. In relation to the current analysis, uncertainty of this type is generated by the fact that only one RCT is used to provide parameter estimates; that a RCT is the main source of data, and trials may generate estimates of resource use and benefits that are unrepresentative of routine practice; that the focus is on one type of MAS

treatment and hysterectomy, when several others are also used in practice; and that unit costs are related largely to one centre. Chapter 6 focuses specifically on an analysis of the generalisability of the CUA presented here.

Analytical method. Uncertainty relating to analytical method springs from the fact that many of the methods used in economic analysis are subject to controversy about their appropriateness. The sensitivity analysis in this chapter considers the implications of using a 0% discount rate on benefits. However, perhaps the main source of analytical uncertainty in the current study is the methods used to estimate benefits. These are explored, and alternatives are presented, as a particular focus of Chapter 7.

5.5   Results

5.5.1 Health State valuation

The final health state scenarios, related to menorrhagia and its treatment, used in the valuation exercise are shown in Box 5.1.

As part of the valuation exercise, 175 women were identified from referral letters, to whom letters were sent explaining the study, as well as a questionnaire and an interview consent form to complete. A total of 115 women returned completed forms, of whom 89 agreed to be interviewed. Out of these 89 women, the target sample of 60 was achieved. In reaching this number, 29 exclusions were made, details of which are provided in Table 5.4.

Table 5.5 presents the socio-demographic and clinical characteristics of the sample of women interviewed, and details of the severity of their menorrhagia. In order to value the health states, assumptions about women's life expectancy were required. One (2%) woman was in the 20 to 29 years age group (assumed life expectancy 60 years); 26 (43%) women were in the 30 to 39 years age group (assumed life expectancy 50 years); 29 (48%) women were in the 40 to

Box 5.1   Health state scenarios used in valuation exercise. All scenarios were presented unlabelled to women.

Health State A (Menorrhagia - chronic state)

- She experiences heavy periods requiring 10 or more pads during the worst day of her period.
- She experiences painful periods.
- She worries about flooding in public. She avoids wearing light-coloured clothing and makes sure she is never far from a lavatory.
- Because of these worries, she is limited in her social activities such as meeting friends and sporting activities.
- She often has difficulties in performing her usual daily activities, especially her work.
- She generally feels tired and lacking in vitality.
- She often feels moody, irritable and depressed.
- Her menstrual problems prevent her from enjoying her sex life.

Health State B (Convalescence following TCRE - temporary state)

- *Recently, she had surgery for her heavy menstrual bleeding and painful periods.*
- It is likely that her heavy bleeding will have disappeared.
- She is aware that there is a risk that the operation may not solve her menstrual problems.
- She has returned home from hospital and is feeling some discomfort.
- She feels tired for some of the day.
- Her discomfort and tiredness mean she is not prepared to go back to work, although she has resumed her activities around the house and most of her social activities.
- She occasionally feels moody, irritable or depressed.
- She is currently not able to have a sex life.

Health State C (Convalescence following AH - temporary state)

- Recently, she had surgery for her heavy menstrual bleeding and painful periods.
- Her menstrual problems have disappeared and she will no longer have periods.
- She has returned home from hospital but sometimes needs to take pain killers.
- She is limited physically as she cannot drive, lift objects or walk very far. She finds it difficult to bend.
- She feels tired for much of the day.
- Because of these problems, she is not prepared to go back to work or to resume her usual activities fully around the house or her social activities.
- She occasionally feels moody, irritable or depressed.
- She is currently not able to have a sex life.

Health State D (Premenopausal following recovery from successful TCRE - chronic state)

- She had surgery for her menstrual bleeding and painful periods three months ago.
- Her operation has not left her with a scar.
- She still has periods but they are much lighter since her operation.
- She still has some pain with her periods.
- She still has her womb, although it is very unlikely that she would become pregnant.
- She is not limited in her social activities.
- She is able to perform her usual daily activities such as work.
- She occasionally feels moody, irritable or depressed.
- Because of the improvement in her menstrual symptoms, she is more able to enjoy her sex life.

Health state E (Premenopausal following recovery from AH - chronic state)

- She had surgery for her menstrual bleeding and painful periods three months ago.
- Her operation has left her with a faint scar on her abdomen.
- She no longer has periods or pain.
- She no longer has a womb so she is unable to bear children.
- She is not limited in her social activities.
- She is able to perform her usual daily activities, such as work.
- She occasionally feels moody, irritable or depressed.
- Because of the improvement in her menstrual symptoms, she is more able to enjoy her sex life.

145

Table 5.4   Details of reasons for exclusions from the Bristol valuation exercise

Number	%	Reason for exclusion
19	66	Date of out-patient appointment too close
2	7	Decided to use private health care
1	3	No telephone, so not possible to arrange appointment
1	3	Lived too far away
1	3	Convenient interview-time could not be identified
1	3	Unable to read
1	3	Too ill to be interviewed
1	3	Not experiencing menorrhagia
1	3	Had undergone a previous endometrial resection
1	3	Interview abandoned due to unsuitable interview conditions

49 years age group (assumed life expectancy 40 years); and 4 (7%) women were in the 50 to 59 years age group (assumed life expectancy 30 years).

Interviews took place between 1st March and 4th November 1994. All women were asked if they were happy for the interview to be taped, and all but 9 (15%) said they were. The mean (SE) duration of interviews was 34.6 (1.69) minutes.

Table 5.5   Socio-demographic and clinical details of the sample of women interviewed in the valuation exercise

Characteristic		
Socio-demographic variables		
Mean (SE) age (years)	41.09	(0.77)
Number (%) who have experienced serious illness:		
- Themselves	18	(33)
- In their family	33	(64)
- In caring for others	14	(30)
Number (%) currently smoking	19	(32)
Number (%) who have worked in health or social services	22	(37)
Number (%) in employment	33	(55)
Number (%) leaving school at minimum leaving age	39	(65)
Number (%) with degree or equivalent professional qualification	10	(17)
Clinical variables		
Median (range) duration of menorrhagia (months)	24	(3-360)
Median (range) days per month bleeding	8	(4-20)
Median (range) days per month with heavy flow	4	(2-14)
Number (%) passing clots	51	(86)
Number (%) with flooding episodes	58	(97)
Maximum number of pads/tampons on heaviest day of period (numbers (%)):		
- 1-9	18	(31)
- More than 9	40	(69)
Median (range) days lost from work due to menstrual problems over last last year for those in work (n = 37)	2	(0-48)

Table 5.6   Values given by women to chronic and temporary health state scenarios
              and to their own health state (n = 60)

Health state	Mean (SE)	Median (Range)
Chronic states		
Menorrhagia	0.50 (0.04)	0.55 (0-0.95)
Pre-menopausal following recovery from successful TCRE	0.73 (0.04)	0.90 (0-1)
Pre-menopausal following recovery from AH	0.86 (0.03)	0.95 (0.05-1)
Own health state	0.65 (0.04)	0.75 (0-1)
Temporary health states*		
Convalescence following TCRE		
Approach I	0.76 (0.04)	0.85 (0-1)
Approach II	0.75 (0.04)	0.85 (0-1)
Convalescence following AH		
Approach I	0.74 (0.05)	0.95 (0-1)
Approach II	0.79 (0.04)	0.95 (0-1)

*   Approach I is where the temporary health state ranked second out the two is valued as a
      short duration chronic health state lasting 10 weeks. For Approach II the state lasts 10
      years.

Table 5.6 shows the values women provided for the three chronic health state

scenarios, their own health state valued as a chronic state and the two

temporary health state scenarios. The results show that the ordering for the

chronic health states is the same on the basis of mean and median values, and

generally as expected. The chronic health state scenario valued lowest by

women was menorrhagia which, in terms of mean values, women were prepared

to trade 50% of their future life expectancy to avoid. The mean and median

values women attached to their own health state were higher than those for the

described state of menorrhagia, probably because many women would not, at

the time of the interview, have been experiencing their period. The chronic

Table 5.7   **Base-case results: expected costs and QALYs over a two year period.**
All costs and benefits are discounted

	TCRE	AH
Expected costs		
Initial Surgery		
Theatre	£234	£292
In-patient	£250	£761
Complications	£5	£32
Other*	£34	£45
Later costs		
Re-treatment costs	£264	£0
Cervical cytology	£2	£0
Hormone replacement therapy	£5	£9
Total expected cost	£794	£1,139
Expected QALYs	1.363	1.593
Difference in costs	£345	
Difference in QALYs	0.23	
Incremental *cost per additional* QALY	£1,500	

*   Includes pre-operative, post-operative and general practice costs in first four months after
initial surgery and other related costs over a two year period.

health state scenario describing health after an AH, but prior to the menopause,
was valued most highly.

As regards the values women provided for the temporary health states, Table
5.6 shows the results on the basis of the two approaches to standardising
values on the conventional 0 to 1 scale. Approach I gives the values generated
for a given temporary health state valued against the worst temporary health
state, where the worst state has been valued as a short duration chronic state
lasting as long as the temporary state (10 weeks). Approach II relates to the
values calculated when the worst temporary health state is valued as a short
duration chronic state lasting 10 years. The table indicates that the mean and
median values calculated from the women's responses are similar, whatever the

method used for standardising the values. Furthermore, the values are broadly similar for both health states. Hence the values based on the valuation of the worst temporary health state as a short duration chronic state lasting 10 weeks is employed to estimate QALYs.

5.5.2 Estimates of costs and benefits

Table 5.7 presents the base-case results of the CUA. The present value of expected costs over two years is £794 for women initially undergoing TCRE, compared to £1,139 for women having an AH. Although AH costs, on average, £345 more per patient than TCRE over two years, the base-case results of the model indicate that it also generates an additional 0.23 of a QALY. Hence, each additional QALY generated by AH has an incremental cost of £1,500. This ratio lies below the lower illustrative threshold ratio of £6,500 introduced above, which would imply that the incremental cost of AH is worth incurring for the additional benefit generated over a two year period. In other words, if the illustrative threshold ratios used are generally accepted on the part of purchasers, the base-case results suggest AH is more cost-effective than TCRE.

5.5.3 Dealing with uncertainty

Data inputs. But how robust is this conclusion to the uncertainty that surrounds the data inputs in the model? Table 5.8 presents the results of a series of one-way sensitivity analyses focusing on the uncertainty in the analysis related to data inputs. The table shows, for each uncertain parameter, the base-case value and the alternative (higher and lower) values used, together with the incremental cost per QALY ratios for each of these values. The table shows that plausible variation in each parameter individually is not sufficient to alter the base-case finding that, over a period of two years after initial surgery, AH is both more costly and more effective than TCRE.

To give a better sense of the variables to which the results are most sensitive, Figure 5.7 plots each of the alternative differential cost and benefit estimates from the one-way sensitivity analyses onto the cost-effectiveness plane. The

Table 5.8   Results of a series of one way sensitivity analyses to assess the implications of uncertainty in data inputs. Unless shown, all lower parameter values are lower 95% confidence intervals and all higher parameter values are upper 95% confidence intervals

Parameter	Parameter values			Incremental Cost per QALY		
	Base-case	Lower	Higher	Base-case	Lower	Higher
Health state values						
1. Menorrhagia (I)*	0.50	-	0.65	£1500	-	£1585
2. Menorrhagia (II)	0.50	0.42	0.58	£1500	£1443	£1538
3. Post AH	0.86	0.80	0.92	£1500	£2449	£1070
4. Post TCRE	0.73	0.65	0.81	£1500	£ 971	£3190
5. Convalescence after TCRE	0.76	0.68	0.84	£1500	£1460	£1519
6. Convalescence after AH	0.74	0.64	0.84	£1500	£1624	£1375
Complications						
7. Probability with AH	38.14%	28.47%	47.81%	£1500	£1441	£1536
8. Probability with TCRE	8.08%	2.71%	13.45%	£1500	£1554	£1424
Failure probabilities in women initially having TCRE						
9. Repeat TCRE	12.06%	4.69%	19.43%	£1500	£1879	£1123
10. Repeat AH given repeat TCRE†	40.00%	20.00%	60.00%	£1500	£1601	£1376
11. AH only	12.17%	4.83%	19.50%	£1500	£1867	£1116
12. Failure with no re-treatment∅	0.00%	-	16.00%	£1500	-	£1299

*   Use of the mean value women attach to their own current health state

†   Use of plausible range rather than 95% confidence intervals.

∅   Relates to possibility that some women undergoing TCRE experience no benefit but do not seek further surgical treatment. Calculated from Bristol RCT as percentage who are 'not very satisfied' or 'very dissatisfied' after TCRE but who have no re-treatment.

**   In the base-case it is assumed that women's treatment fails half way (0.5) between initial surgery and re-treatment. In the sensitivity analysis, this point is varied between failure immediately (0.00) and just before re-treatment (1.00).

‡   Based on Chartered Institute of Public Finance and Accountancy (CIPFA) [1990].

151

Table 5.8 Results of a series of one-way sensitivity analyses to assess the implications of uncertainty in data inputs. Unless shown, all lower parameter values are lower 95% confidence intervals and all higher parameter values are upper 95% confidence intervals (continued)

Parameter	Parameter values			Incremental Cost per QALY		
	Base-case	Lower	Higher	Base-case	Lower	Higher
Re-treatment times (months)						
13. Until repeat TCRE	7.40	4.38	10.42	£1500	£1503	£1475
14. Until AH given repeat TCRE	7.50	0.00	15.87	£1500	£1516	£1461
15. Until AH only	10.40	4.85	15.95	£1500	£1542	£1442
16. Point where treatment fails between initial surgery and re-treatment**	0.50	0.00	1.00	£1500	£1359	£1642
Duration of convalescence (weeks)						
17. AH without complications	11.63	10.26	13.00	£1500	£1476	£1501
18. AH with complications	10.65	9.40	11.90	£1500	£1482	£1496
19. TCRE without complications	2.32	2.03	2.61	£1500	£1486	£1491
20. TCRE with complications	4.71	3.38	6.04	£1500	£1487	£1490

*   Use of the mean value women attach to their own current health state

†   Use of plausible range rather than 95% confidence intervals.

ø   Relates to possibility that some women undergoing TCRE experience no benefit but do not seek further surgical treatment. Calculated from Bristol RCT as percentage who are 'not very satisfied' or 'very dissatisfied' after TCRE but who have no re-treatment.

** In the base-case it is assumed that women's treatment fails half way (0.5) between initial surgery and re-treatment. In the sensitivity analysis, this point is varied between failure immediately (0.00) and just before re-treatment (1.00).

‡   Based on Chartered Institute of Public Finance and Accountancy (CIPFA) [1990].

152

Table 5.8 Results of a series of one-way sensitivity analyses to assess the implications of uncertainty in data inputs. Unless shown, all lower parameter values are lower 95% confidence intervals and all higher parameter values are upper 95% confidence intervals (continued)

Parameter	Parameter values			Incremental Cost per QALY		
	Base-case	Lower	Higher	Base-case	Lower	Higher
Hormone replacement therapy (HRT)						
21. Probability of using HRT after AH	17.00%	8.32%	25.68%	£1500	£1471	£1506
22. Probability of using HRT after TCRE	10.00%	2.59%	17.41%	£1500	£1504	£1474
Cervical cytology						
23. Proportion of women presenting†	74.00%	64.00%	84.00%	£1500	£1490	£1487
Length of hospital stay (days)						
24. AH without complications	6.19	5.95	6.43	£1500	£1425	£1552
25. AH with complications	6.64	6.27	7.01	£1500	£1427	£1551
26. TCRE without complications	1.99	1.95	2.03	£1500	£1509	£1468
27. TCRE with complications	3.25	1.88	4.62	£1500	£1553	£1425

* Use of the mean value women attach to their own current health state

† Use of plausible range rather than 95% confidence intervals.

ø Relates to possibility that some women undergoing TCRE experience no benefit but do not seek further surgical treatment. Calculated from Bristol RCT as percentage who are 'not very satisfied' or 'very dissatisfied' after TCRE but who have no re-treatment.

** In the base-case it is assumed that women's treatment fails half way (0.5) between initial surgery and re-treatment. In the sensitivity analysis, this point is varied between failure immediately (0.00) and just before re-treatment (1.00).

‡ Based on Chartered Institute of Public Finance and Accountancy (CIPFA) [1990].

Table 5.8 Results of a series of one-way sensitivity analyses to assess the implications of uncertainty in data inputs. Unless shown, all lower parameter values are lower 95% confidence intervals and all higher parameter values are upper 95% confidence intervals (continued)

Parameter	Parameter values			Incremental Cost per QALY		
	Base-case	Lower	Higher	Base-case	Lower	Higher
Time in theatre (minutes)						
28. AH without complications	62.75	60.32	65.18	£1500	£1471	£1506
29. AH with complications	63.16	58.79	67.53	£1500	£1470	£1508
30. TCRE without complications	50.12	48.22	52.02	£1500	£1511	£1467
31. TCRE with complications	63.75	57.69	69.81	£1500	£1495	£1483
Mortality						
32. Risk of death with AH†	0.1%	-	0.2%	£1500	-	£1496
33. Risk of death with TCRE†	0.1%	0.05%	-	£1500	£1492	-
Unit costs						
34. Cost of in-patient day‡	£120	£53	£249	£1500	£619	£3174

* Use of the mean value women attach to their own current health state
† Use of plausible range rather than 95% confidence intervals.
ø Relates to possibility that some women undergoing TCRE experience no benefit but do not seek further surgical treatment. Calculated from Bristol RCT as percentage who are 'not very satisfied' or 'very dissatisfied' after TCRE but who have no re-treatment.
** In the base-case it is assumed that women's treatment fails half way (0.5) between initial surgery and re-treatment. In the sensitivity analysis, this point is varied between failure immediately (0.00) and just before re-treatment (1.00).
‡ Based on Chartered Institute of Public Finance and Accountancy (CIPFA) [1990].

Figure 5.7 Graphical representation of the one-way sensitivity analyses undertaken on data inputs using Quadrant I of the cost-effectiveness plane. The numbers shown relate to the specific variables listed in Table 5.8.

155

origin runs to the base-case result, the gradient of this line being the base-case incremental cost per QALY ratio (£1,500). Hence points below this line represent a lower incremental cost per QALY estimate for AH than in the base-case, and points above the line represent higher estimates.

The ward cost per in-patient day is subject to considerable uncertainty, due partly to genuine cost differences between hospitals, but also to the inconsistency in hospitals' costing methods. Chapter 6 considers the impact on the conclusions of the study of variation in key unit costs between hospitals in more detail. However, Table 5.8 and Figure 5.7 emphasise how sensitive the incremental cost per QALY ratio is to ward cost. For indicative purposes, when a particularly large range for this unit cost is used (£52 to £249) based on hospital cost returns [Chartered Institute of Public Finance and Accountancy (CIPFA), 1990], the incremental cost per additional QALY of AH varies between £619 and £3174.

The one-way sensitivity analyses also show that the incremental cost per QALY of AH is particularly sensitive to the health state values of the post-convalescence /pre-menopausal states (Post-AH and Post-TCRE in Figures 5.2 and 5.3). Table 5.6 shows that the mean and median values of these health states are different, in favour of AH. Furthermore, this difference is statistically significant (Wilcoxon rank sum test, $p = 0.008$), which explains why AH remains more effective than TCRE in the one-way sensitivity analyses. However, the incremental ratio is sensitive to the *difference* between these values: the mean difference of 0.13 used in the base-case has 95% confidence intervals ranging from 0.03 to 0.23.

Figure 5.8 looks at the sensitivity of the incremental cost per QALY to variation in the difference in these two health state values, by keeping the Post-AH value in the model fixed at 0.86, and varying the Post-TCRE value according to the 95% confidence intervals. The figure shows that, although the incremental ratio is sensitive to this variation, the smallest difference between these values (0.03)

156

Incremental cost per additional QALY

Value of the Post-TCRE health state
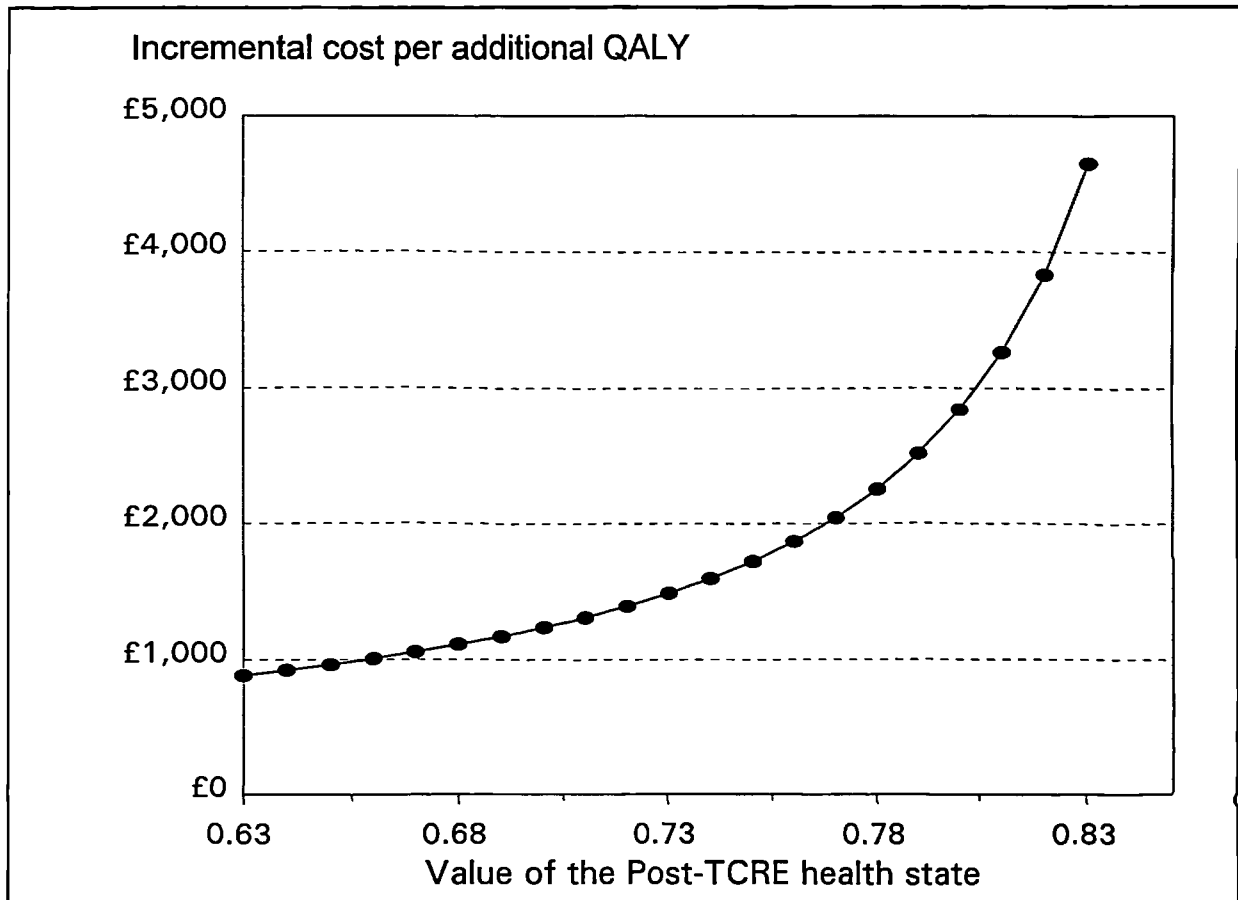
Figure 5.8   A graphical representation of the variation in the incremental cost of AH per additional QALY as the value of the Post-TCRE health state varies. The value of the Post-AH health state remains at its base-case value (0.86).

is sufficient to keep the ratio below the lower illustrative threshold ratio of £6,500.

An implicit assumption of one-way sensitivity analysis is that variability in a parameter is independent of variability in one or more other parameters. This is unlikely to be the case in practice, so, in the absence of stochastic data for all variables, one way to explore the robustness of the conclusions of an analysis to co-variance in parameters is an analysis of extremes [Briggs *et al*, 1994]. This form of sensitivity analysis compares the base-case incremental cost per QALY

with the ratio generated under two alternative cases: one where all the parameters in the model are simultaneously altered to the extremes of their plausible range in a way that favours AH (the 'optimistic for AH' scenario); and the second where all the parameters are simultaneously altered to the extremes of their plausible range in a way that favours TCRE (the 'pessimistic for AH' scenario). Using the 'optimistic for AH' scenario, AH would remain more effective, but also less costly; that is, AH would dominate TCRE. Using the 'pessimistic for AH' scenario, the incremental cost of AH per additional QALY would be over £255,000, which is significantly higher than the upper illustrative threshold ratio and very unlikely to be considered cost-effective. Even if the rather wide plausible range for ward cost per in-patient day were narrowed somewhat so that the upper value were £180, the 'pessimistic for AH' scenario would still generate an incremental cost for AH per additional QALY of nearly £190,000.

Extrapolation. In order to provide an indication of how robust the conclusions of the base-case analysis are to taking a longer time horizon for costs and benefits, Table 5.9 shows the extrapolated costs and QALYs until the menopause and until death. Although, as described above, these extrapolations have been based on simple assumptions, they do indicate that AH looks increasingly cost-effective over a longer time horizon. This is because the Post-AH health state is valued higher than the Post TCRE health state and is allowed to have an effect for longer in these extrapolations. Given that women undergoing both procedures are assumed to experience equally valued health states after the menopause, it is not surprising that the incremental cost per QALY ratio alters little for the extrapolation until death. The small difference between the menopause and death only reflects the fact that women having TCRE can expect to undergo, on average, more surgical procedures and, therefore, experience a greater overall risk of operative mortality.

The implications for the conclusions of the analysis of uncertainty relating to generalisability is considered in detail in Chapter 6. The main source of

uncertainty in analytical method - the methods used to estimate the benefits of the two procedures - is explored in detail in Chapter 7. Another source of uncertainty in analytical method is the rate used to discount benefits in the model. If this is reduced from the 6% used in the base-case analysis (the same as for costs) to 0%, the estimates of expected benefit increase sharply, particularly when extrapolated to the menopause and to death. However, the incremental cost per QALY ratios change very little: £1463 in the two year analysis, £349 in the extrapolation to the menopause and £346 in the extrapolation to death. This modest change in ratio is because the temporal distribution of benefits is very similar for the two interventions.

5.6   Discussion

5.6.1 Cost and benefit estimates

Given the uncertainty surrounding aspects of the CEA presented in Chapter 3, this chapter builds on that analysis, presenting the results of a CUA using the standard QALY model. The purpose of the CUA is to estimate the relative cost-effectiveness of AH and TCRE when benefits are expressed in terms of QALYs. Although the main source of data for the CUA is the Bristol RCT, the analysis develops a decision analytic framework to provide increased flexibility to explore uncertainty in the cost-effectiveness estimates, both in this and subsequent chapters. The data from the Bristol RCT have been augmented by a specific health state valuation study, which elicits values for health states associated with menorrhagia and its treatment from a sample of women with the condition.

The results of the analysis of costs reported here confirm that, despite a 23% re-treatment rate over two years in women initially undergoing TCRE, AH remains more costly, from a health service perspective, over that period. However, given the failure rates with TCRE observed in the Bristol RCT and the health state values elicited in the valuation study, AH is also more effective in terms of QALYs.

Table 5.9   Results of the extrapolation exercise. All costs and benefits are discounted

	TCRE		AH
Extrapolation until menopause			
Expected costs (£)	816		1162
Expected QALYs	5.179		5.958
Differential cost (£)		346	
Differential QALYs		0.779	
Incremental cost per additional QALY (£)	-		444
Extrapolation until death			
Expected costs (£)	816		1162
Expected QALYs	14.413		15.195
Differential cost (£)		346	
Differential QALYs		0.782	
Incremental cost per additonal QALY (£)	-		442

5.6.2 Cost-effectiveness

The base-case analysis indicates that each additional QALY generated by AH has an incremental cost of £1,500. These results would suggest that the decision facing health care purchasers currently purchasing AHs for women with menorrhagia is whether to continue with AHs, or to switch to purchasing TCREs. By devoting the additional resources to AH, purchasers may be forgoing greater benefits elsewhere that could be realised if less costly TCREs were purchased. Similarly, those currently purchasing TCREs for their population with menorrhagia will need to decide whether to use any increase in revenues to facilitate a switch to AH, and/or to make purchasing changes elsewhere to release resources to permit a change to AH.

Those QALY league tables that have been published certainly suggest that the health service is funding interventions and programmes with incremental cost per QALY ratios in excess of the base-case result here. For example, the league table presented by Maynard [1991] from a review of the literature suggested that kidney transplantation, breast cancer screening and heart transplantation each has incremental ratios appreciably higher than £1,500. It is clear that rigid adherence to league tables would be to ignore their methodological limitations [Drummond *et al*, 1993B; Mason *et al*, 1993; Mooney and Gerard, 1993]. However, if purchasers accept the findings of CUAs, they will need to decide on a threshold cost per QALY above or below which a serious reappraisal of purchasing policy will be triggered. The illustrative incremental cost per QALY thresholds used here are based on some tentative proposals made in Canada, which have been adjusted for the UK [Laupacis *et al*, 1992]. If purchasers think these threshold ratios are acceptable, then, on the basis of base-case data, AH would be considered more cost-effective than TCRE.

It is interesting to note that Laupacis *et al* make a distinction between cost per QALY thresholds relating to the acceptance or rejection of *new* technologies, and those relating to *existing* programmes. They suggest that an existing technology that is more effective and expensive than a new comparator might have a higher threshold incremental cost per QALY than a new technology which is more costly and effective than an existing comparator. There would seem to be a case for imposing a stronger 'burden of proof' on new technologies. However, in relation to the treatment of menorrhagia, TCRE has now diffused widely in the UK and centres will differ according the whether it is the new or the existing intervention for the condition.

5.6.3 Uncertainty

All economic evaluations are subject to sources of uncertainty. This chapter has looked in detail at the robustness of the base-case conclusion to uncertainty in data inputs into the model. The sensitivity analyses show that the incremental ratio is particularly sensitive to variation in the unit cost of a day on a ward and to the health state values of the Post-AH and Post-TCRE health states. Varying these parameters individually across a plausible range, however, does not reverse the base-case

conclusion that AH is more cost-effective than TCRE, if the lower illustrative incremental ratio (£6,500) is considered acceptable. However, an analysis of extremes, using a particularly pessimistic scenario regarding data inputs from the perspective of AH, indicates that the combined uncertainty of all data inputs could be sufficient to alter the conclusion that AH is the more cost-effective.

An important area of uncertainty regarding the relative cost-effectiveness of AH and TCRE is the costs and benefits that will accrue in the future. The base-case results presented here are based on the firm evidential basis of a RCT, and a reasonable level of confidence can be attached to these results over the period of follow-up in the Bristol trial of two years. However, women's prognoses over subsequent years may alter the economic balance between the interventions significantly. For example, AH offers a prophylactic effect against some gynaecological cancers, but it may be associated with premature ovarian failure and early menopausal symptoms [Siddle *et al*, 1987]. This emphasises the importance of continued follow-up of women in the RCTs comparing AH and TCRE, and the particular value of the large long-term cohort study currently underway in the UK. However, purchasers need to make decisions about the relative cost-effectiveness of the two interventions prior to the longer-term data becoming available. To assist in this process, two simple extrapolations are presented here: one until the menopause and one until death. These extrapolations show that AH is likely to remain the more cost-effective option if judged against the illustratice threshold ratios. However, this conclusion must be a cautious one, and the model presented here should be updated when longer-term follow-up data become available.

A third area of uncertainty concerns the methods used within the CUA. The choice of how to value the health states used within the model is likely to be crucial to the final results. The sensitivity analysis shows that the benefit (and hence the incremental cost per QALY) estimates are sensitive to sampling variation in some of the health state values (for example, see points 3 and 4 in Figure 5.7). For this study, a valuation strategy of presenting a sample of women experiencing menorrhagia with descriptive scenarios for key health states and eliciting values

using the TTO method was selected. Alternative approaches - for example using generic descriptions from a valuation system such as the EuroQol or adopting the standard gamble valuation instrument - may have generated different health state values and, consequently, estimates of incremental cost per QALY. However, no 'gold standard' approach exists to the choice and valuation of health states, making the validation of the methods used here difficult.

The additional data source introduced into the economic analysis in this chapter relates to the health state values. The relative magnitude of the health state values is broadly as expected. However, the fact that the mean value women provided for the untreated menorrhagia scenario (0.50) was markedly less than that for their own current health state (0.65) was surprising. As discussed in Section 5.5.1, this difference was probably due to the fact that only a proportion of women would have been having their period at the time of the interview, and that women were valuing the health state of menorrhagia imagining it as a state of 'continual menstruation' (ie. a period lasting over their life expectancy), rather than a state where the worst effects would be experienced for approximately one week out of four, but some effects of which would extend over the full month. There are problems in eliciting values for health states relating to essentially chronic conditions which affect HRQL on a daily basis, but the worst effects of which are episodic. Some uncertainty, therefore, exists in the value to attach to the menorrhagia health state within the CUA, in addition to the sampling variation. In Table 5.8 and Figure 5.7 (point number 1), the sensitivity analysis substitutes the mean value elicited from women for their own current health state for the mean value women attached to the described state of menorrhagia, and this has only a modest effect on the incremental ratio.

Table 5.8 and Figure 5.7 show that the key values are those for the Post-AH and Post- TCRE health states. Although these values are statistically significantly different, the absolute difference between them is subject to uncertainty. When looked at in isolation, the conclusions of the base-case analysis are robust to this source of uncertainty (Figure 5.8). However, the uncertainty in the difference

between these values makes a major contribution to the absence of robustness in the analysis of extremes.

5.6.4 Methodological issues

The focus of this chapter has been an empirical contribution to understanding the relative cost-effectiveness of AH and TCRE, in particular estimating benefits in terms of a generic measure reflecting patients' preferences. In addition, some 'incremental' contributions to how uncertainty in the results of economic analysis is handled and presented have been made. For example, the 'scatter-plot' in Figure 5.7 used to illustrate the one-way sensitivity analyses provide a valuable visual indication of the importance of individual variation in parameters.

In addition, the chapter does raise some important methodological issues. Firstly, a fundamental source of uncertainty regarding analytical method is the validity of the standard QALY model used in the analysis. A major motivation for undertaking a CUA of AH and TCRE is the need to incorporate women's preferences about the different outcomes of treatment into the economic evaluation framework. The standard QALY model here seeks to achieve this by using women's preferences to value the individual health states in the model. However, the assumptions necessary to link the QALY to individual preferences are strong. Chapter 7 considers this source of uncertainty, and alternative measures of benefit for CUA, in more detail.

The second methodological issue raised by the chapter relates to generalisability. Although the CUA presented here broadens the evaluation by expressing benefits in generic terms which may more adequately reflect patients' preferences, the analysis still relies largely on resource and effect data from the Bristol RCT, calling into doubt the generalisability of the estimates of cost and benefit. The decision analytic model developed in the chapter provides a framework for incorporating data inputs from other sources and, therefore, offers a bridge to the analysis of generalisability detailed in Chapter 6.

5.7   Conclusions

The base-case CUA results presented here suggest that, for an intervention which is more effective and more costly than its comparator, if a threshold incremental cost per QALY of £6,500 is acceptable to purchasers, then AH is more cost-effective than TCRE. However, on the basis of existing data, there are important uncertainties associated with this conclusion. This chapter has shown that the conclusion is robust to variation in individual data inputs, but not to extreme co-variation in these inputs. Furthermore, methodological issues are raised in relation to benefit estimation and generalisability. The next three chapters of the thesis consider these methodological issues further.

Chapter 6

The Generalisability of the Costs and Benefits of AH Versus TCRE

6.1 Introduction

The use of MAS is likely to be characterised by variation between centres in the process and outcomes of care. Related to this is the fact that MAS applications are developing quickly over time. Therefore, in order for health service decision makers to use the data presented in Chapter 5 to assist in the resource allocation process, they need to know whether the conclusions of the analysis are consistent with those that would be expected generally in routine practice.

This chapter considers the issues of method that are raised in relation to the generalisability of economic evaluations. In particular, the chapter focuses on the reasons why economic studies might lack generalisability, and the available approaches to assess levels of generalisability within an evaluation. Using the general tool of

sensitivity analysis, the chapter provides an assessment of the robustness of the base-case conclusions in Chapter 5 to alternative data inputs drawn from sources that reflect the variation in the costs, process and outcomes of health care delivery in this clinical area.

The chapter is structured in the following way. Section 6.2 considers the methodological issues related to generalisability in economic evaluation. Sections 6.3 to 6.8 detail five separate analyses of the generalisability of the base-case CUA, which re-estimate the costs and benefits of TCRE and AH by using alternative sources for the data inputs used in the model. Section 6.9 provides a discussion of the results, and Section 6.10 offers some conclusions.

6.2   Generalisability in economic evaluation

Generalisability (or external validity) is concerned with the extent to which the conclusions of an analysis, as they apply to a specific population, location or context, hold true in relation to a different population, location or context [Briggs *et al*, 1994]. Much has been written *about issues of generalisability in clinical evaluation [Bailey,* 1994; Davis, 1994]; particularly about its trade-off with internal validity when choosing a study design [Schwartz and Lellouch, 1967]. However, with the exception of a small number of studies looking at international generalisability [Drummond *et al*, 1992; Leese *et al*, 1992], there has been relatively little consideration, at a methodological or empirical level, of the external validity of economic evaluations. This is the case despite the fact that generalisability is probably more difficult to achieve in economic than clinical evaluation because, in addition to data inputs relating to clinical effectiveness, economic analysis incorporates other categories of data which tend to be particularly influenced by location and context.

The need to consider the concept of generalisability springs from the variation that exists in clinical practice that cannot easily be handled using standard statistical methods within a specific study. The variation likely to be exhibited in relation to MAS technologies relates, in particular, to the effect of the learning curve on resource and

non-resource consequences of treatment, the differences between centres in the detailed process of treatment and changes in practice over time. The impact of variation on an economic evaluation can be considered using the four basic categories of data input: effectiveness, resource use data, unit costs and health state valuation, each of which is discussed below.

6.2.1 Effectiveness data

The effectiveness data used in economic evaluations may be taken from various sources, and the type of data source will have a major impact on the generalisability of results. Data sources can sensibly be grouped under randomised controlled trial (RCTs) and observational studies.

Randomised controlled trials. It is widely considered that the gold standard of clinical evaluation is the RCT [Pocock, 1983], and this is a key design for evaluation funded as part of the NHS programme of health technology assessment [Advisory Group on Health Technology Assessment, 1992]. These studies are also now seen as an important source of effectiveness data for economic evaluation [Drummond and Davies, 1991; Drummond, 1995]. However, as discussed in Chapter 2, it has long been recognised that such studies may exhibit a lack of external validity. This realisation led Schwartz and Lellouch [1967] to distinguish between explanatory and pragmatic trials. Explanatory trials seek to test specific hypotheses in 'ideal' clinical conditions. By keeping 'extraneous variation' in the process of care to a minimum, it is hoped that the treatment differences identified in these studies reflect true differences in *efficacy*. Pragmatic trials on the other hand explicitly recognise that the *effectiveness* (as distinct from efficacy) of interventions will partly reflect a range of factors associated with the process of care which interact with the specific interventions under consideration. These factors include the skill of the clinician, the willingness of the patient to comply with treatment and the availability of other therapeutic and diagnostic technologies.

In reality, it is unlikely that any RCT is a perfect example of either an explanatory or a pragmatic trial. The result of this is that RCTs invariably have some limitation to their

generalisability. These limitations have been described elsewhere [Simon *et al*, 1995], and include the following.

(a)   RCTs tend to focus on atypical populations. This can occur for a range of reasons including the fact that they are often undertaken in tertiary referral centres where 'unusual' cases are sent, because trials may use careful screening procedures to achieve as homogenous a study sample as possible and because the need for informed consent may result in the inclusion of patients who have atypical attitudes to health care delivery.

(b)   The centres taking part in RCTs may also be atypical. As RCTs are often undertaken within specialist centres, the care a patient receives is likely to differ from the routine in more ways than just the technologies being evaluated. For example, it is likely that clinical staff will be more skilled and experienced than in non-specialist centres.

(c)   Even outside specialist centres, the process of care in RCTs may be quite atypical. For example, monitoring may be necessary for safety purposes which in themselves alter the management a patient receives; the use of questionnaires and interviews to measure the impact of a technology on patients' HRQL may, in itself, affect outcome through a 'Hawthorne effect'. Indeed, the whole process of standardising care may impact on outcomes and is quite different to the routine, where significant variation in medical practice is evident [Cleary *et al*, 1991].

(d)   When a patient enters a RCT, the nature of the doctor-patient relationship may alter. For example, in a trial, the need for long-term follow-up can result in the hospital doctor retaining an interest in a patient for a longer period than would routinely be the case. This may have a range of consequences, such as a tendency for greater use of diagnostic and therapeutic interventions.

(e)   Some trials seek to blind patients and doctors to the treatment allocation. However, in routine practice, knowing which intervention is being used may influence patients' and doctors' attitudes to and compliance with treatment, thus impacting on overall effectiveness.

Observational studies. In the absence of trial data, observational studies are often used to provide data for economic evaluation. Indeed, in analyses based on models, a mixture of RCT and observational data may be used. Although the observational study may exhibit biases because of the strong possibility of confounding variables influencing results, they can have the advantage of greater generalisability, as they are less likely to involve major interference with routine practice. Furthermore, because of the absence of a formal trial infrastructure, observational studies tend to be less costly to undertake and are, therefore, more likely to cover the wide variation in patient characteristics and clinical practice. However, as with all forms of clinical evaluation, the results can only be generalised to the patient sub-groups and clinical contexts covered, and no single study is likely to be able to cover all variations within these.

6.2.2 Resource use data

As well as generating effectiveness data for economic evaluations of health care technologies, RCTs are increasingly used to provide resource use data for those studies. As a result, this category of data input may also be characterised by limitations in generalisability. The atypical nature of trial patients and of the process of care in RCTs directly impacts on the resources patients consume. The most obvious way that this can manifest itself is with the use of protocol-determined investigations and hospital visits, which need to be allowed for in an economic evaluation seeking to reach conclusions relevant to routine practice.

For non-trial-based evaluations or modelling studies involving the synthesis of data from several sources, resource use data can be based on observational studies, clinical opinion or *ad hoc* surveys. It remains the case, however, that these data can usually be characterised as 'the best that can be acquired in the circumstances', and are very unlikely to reflect routine practice. Large administrative databases, usually based on claims data, can sometimes provide a useful source of resource use data, but they are not usually available in publicly-funded health care systems such as the NHS, they often fail to provide the fine detail of resource consumption that may be needed and they may not even allow resource consumption to be distinguished from overall costs.

6.2.3 Unit costs

The process of valuing resource use may also limit the generalisability of an economic evaluation. In principle, an item of resource should be valued according to its opportunity cost - the benefit forgone in using it in one way rather than in the next best alternative. Given the speed with which health care technologies develop and the major variation that exists in the configuration and organisation of health care facilities, the true opportunity cost of a resource will vary with time and place. At the extreme, therefore, the unit cost of a resource can be seen as unique to the time and facility at which it is used. This places a major limitation on the generalisability of economic evaluation.

In practice, a more pragmatic approach to costing has been taken in economic evaluation, with the acceptance that market prices, or surrogates for them, represent an adequate means of valuing resource use in health care. In the recent past, the quality of unit cost data available in the NHS was poor. If a hospital was identified which could provide some data, the costing methods used were invariably limited; and the quest for data from a range of hospitals, to explore the robustness of cost estimates to variations in unit costs, was handicapped by the lack of standard costing methods. For example, although hospitals have usually been able to provide an estimate of the average cost of an in-patient day, it has not been clear to what extent the significant variation between hospitals in these costs was due to genuine underlying cost differences (eg. the market price of land or labour), to differences in the mix of patients treated or to inconsistent costing methods. This variation is illustrated by the fact that, in 1990, the average stated cost per in-patient day for cardiology by district ranged between £41 and £539 [CIPFA, 1990].

In recent years, the quality of unit cost data available in the NHS has improved, at least in some centres. In large part this is due to the need to set prices within the reformed NHS of separate purchasers and providers. This process has been assisted by Department of Health guidelines on costing methods [NHS Management Executive, 1993], and by the development of financial information systems tailored to these guidelines. It is now possible to identify a group of hospitals which can provide unit

cost data using similar costing methods, so that differences between them are likely to reflect differences in underlying costs.

6.2.4 Health state valuation data

The process of valuing the outcomes of medical technologies in order to construct generic measures of benefit such as QALYs has been more concerned with methodological issues such as the choice of valuation instrument, than the generalisability of the data. Generalisability in this category of data input is related to the choice of whose values are to be used. If the values of the general public are to count, generalisability would require an adequately sized and stratified sample. The work recently undertaken by the University of York to elicit valuation data from a sample of over 3300 members of the public in the UK represents by far the most important move towards a generalisable set of health state values relating to the UK population as a whole [Williams, 1995]. If patients' values are considered important, it would again be necessary to collect data from an extensive sample reflecting diversity of opinion, preferences and clinical characteristics.

6.3 Assessing generalisability within economic evaluation

The variation that exists in the resource and non-resource consequences of health care gives rise to two analytical concepts, generalisability and extrapolation, and it is useful to distinguish the two. Extrapolation is concerned with taking the results of a study undertaken in a specific context, where context can be defined in terms of such things as location and point in time, and attempting to translate the results to another *specific* context. Generalisability, on the other hand is concerned with taking the results of a study undertaken in a specific context and seeking to assess the extent to which the results hold true in clinical practice *as a whole*; that is, across a range of different contexts. Extrapolation analysis will be of interest to particular decision makers as long as it is their contexts that the results are being extrapolated to - that is, as long as the analysis relates to their mix of patients, reflects the process and resource use in their clinical practice and incorporates their unit costs. Analyses of generalisability, however,

will be of interest to decision makers more generally by exploring the extent to which the conclusions of an analysis are sensitive to variation.

A number of ideas have been put forward on how to increase the external validity of *clinical* evaluations. These include the use of 'naturalistic' or 'real life' RCTs which aim to superimpose randomisation on routine practice, without the research exerting any extraneous effect on process or outcomes [Simons *et al*, 1995]. The greater use of observational studies, which might be strengthened to allow for the various sources of bias they often exhibit, has also been suggested [Sechrest and Hannah, 1990]. To the extent that economic evaluations are undertaken alongside these clinical studies, they may benefit from the success of these methods in increasing generalisability. However, as noted above, no single study is likely to be able adequately to cover or reflect the large variation that exists in patient characteristics and clinical contexts. Whatever the source of the effectiveness data, caution is required in generalising from a single evaluation to a specific context [Rubins, 1994]. Furthermore, even if an economic analysis is undertaken alongside a 'real world' RCT which provides generalisable data on effectiveness and resource consumption, there is no guarantee that the unit cost data used to value the resource use measured in the trial will exhibit a high level of generalisability.

There have been attempts to generalise economic assessments across national boundaries, by undertaking separate studies in different countries [Drummond *et al*, 1992; Leese *et al*, 1992], and this has been feasible largely because they were based on models or observational studies. However, it is clearly quite impractical to undertake separate economic evaluations for each location and context that might emerge in practice *within* a country. In the review of published economic evaluations of MAS detailed in Chapter 2, only 2/16 studies provided any significant assessment of generalisability

Therefore, an important role in the assessment of the level generalisability in an economic evaluation will be played by sensitivity analysis. This may take the form of the generation of scenarios concerning how one or more parameters in an evaluation might differ in routine practice from that in the base-case analysis. These scenarios are

often not based firmly on actual data, but can relate to plausible contexts or situations. For example, in the review in Chapter 2, Mays' [1991] economic evaluation of extracorporeal shockwave lithotripsy undertook a base-case analysis and then recalculated its results making different assumptions about the installation cost and utilisation of the equipment. Sculpher [1993] undertook an exploratory economic evaluation of the diode laser in ophthalmology, basing the analysis on a number of alternative scenarios about the hospitals considering the purchase of a laser; for example, whether they already had another type of laser and, if so, if it had any useful life remaining.

A potentially more rigorous approach to the use of sensitivity analysis to assess generalisability is the development of a base-case analysis and then the incorporation of data from alternative sources that are considered likely to reflect different aspects of routine practice. The objective would be to assess the *robustness* of the base-case results: the extent to which the alternative parameter estimates alter the conclusions of the base-case analysis. This is perhaps most frequently undertaken with unit costs in a study alongside a RCT: base-case unit costs are taken from the trial centre(s), but sensitivity analysis incorporates unit costs from other non-trial centres [Sculpher *et al*, 1994]. However, this process can work with all components of an economic evaluation. In the MAS review in Chapter 2, only one study [England *et al*, 1987] provided this level of detail in assessing generalisability.

There is a key role for the decision analytic model in this more rigorous approach to assessing generalisability. As an evaluative tool, the model provides a framework within which to explore the implications of alternative data sources. This implies that, even for RCT-based economic evaluations, the model can be used within which to organise data. Furthermore, the approach can explicitly recognise the trade-off between internal and external validity in selecting study methods. This could be achieved by the base-case analysis incorporating data with high levels of internal validity, and sensitivity analysis using other data sources as alternative parameter estimates. These alternative sources would be expected to have greater external validity, but at the expense of lower internal validity when used to compare interventions; for example large-scale surveys or other observational studies undertaken in routine practice.

One alternative source of data are other RCTs undertaken in the same clinical area. Although these other experimental studies might also lack generalisability, incorporating data from several RCTs rather than one will facilitate some assessment of the generalisability of the base-case. The use of several RCTs to answer a single research question has recently acquired a great deal of interest in the form of meta-analysis [Eysenck, 1994]. Meta-analysis is a formal means of synthesising data from a number of studies to provide a statistically more powerful estimate of effect size than that offered by a single study. This method has been used widely in clinical evaluation, including the menorrhagia field [Coulter *et al*, 1995]. Indeed, meta-analysis is increasingly being used to provide estimates of effectiveness for economic evaluations. For example, O'Brien *et al* [1994] used a meta-analysis of the efficacy of enoxaparin and warfarin as prophylaxis against deep vein thrombosis in a cost-effectiveness analysis of the two drugs.

One of the problems with meta-analysis, however, is the heterogeneity of RCTs, and the danger that combining results will inappropriately mask the underlying differences *between* trials [Thompson, 1994]. Statistical tests are available to assess the extent of heterogeneity in a meta-analysis but, as Eddy [1990] has said:

> 'The test for heterogeneity is basically a significance test, with the null
> hypothesis being that there is homogeneity. In fact, there is virtually never
> homogeneity between studies. Differences in subjects, settings, provider skills,
> techniques, and other factors make it very questionable to support the
> assumption that every study is estimating the same population parameter.'
> (p175)

The problem of heterogeneity is likely to be one factor explaining the observed differences between the results of meta-analyses and subsequent large RCTs [Borzak and Ridker, 1995].

Meta-analysis is likely to play an increasingly important role in providing estimates of effectiveness (and perhaps resource use also) for the *base-case* analyses of economic evaluations. However, in order to assess the generalisability of these studies, it will be

necessary to incorporate the results of each trial individually, to consider the robustness of the base-case conclusions to variability in data inputs which will partly reflect heterogeneity in trials.

To illustrate how these methods might be used, the remainder of the chapter considers the assessment of generalisability in economic evaluation in the context of the comparison of AH and TCRE. Six specific analyses are presented, each of which seeks to assess whether the base-case conclusions described in Chapter 5 are robust to alternative parameter estimates based on data from sources reflecting different aspects of routine practice.

6.4   Analysis of Generalisability I: alternative trial results

6.4.1 Purpose

The CUA presented in Chapter 5 is based on effectiveness and resource use data taken from the Bristol RCT described in Chapter 3. The fact that the study was undertaken in a specialist centre, that it took place relatively early in the diffusion of TCRE (January 1990 to May 1991) and that it was a RCT may limit its external validity and hence that of the economic evaluation. The first analysis of generalisability, therefore, considers the results of the other two published RCTs of TCRE and AH and incorporates their key results into the cost-utility model to assess the robustness of the base-case results.

6.4.2 Methods

The two published RCTs used for alternative parameter estimates are Gannon *et al* [1991] and Pinion *et al* [1994]. The two RCTs have been used to provide alternative estimates of a given parameter if the following apply:

(a)   the results of the base-case analysis were shown to be sensitive to variation in the parameter in the sensitivity analysis of data inputs in Chapter 5, or such sensitivity is thought likely; and

(b)   the alternative trial data sources have published estimates for that parameter.

In the base-case analysis, the availability of patient-based data from the Bristol RCT allowed some parameters to be estimated separately for patients who experienced complications, and for those who did not. This level of detail is not available for the two alternative trials, so the overall mean values of those parameters are used in the sensitivity analysis.

The proportions of women randomised to TCRE who required repeat TCRE and/or hysterectomy are based on a two-year follow-up within the Bristol RCT in the base-case analysis. The follow-up period in the two alternative trials is, however, only one year. In order that the sensitivity analysis is based on the same time horizon as the base-case (ie. two years), the one year failure rates from the Gannon *et al* and Pinion *et al* trials are extrapolated to two years. This is done using life table analysis, where the cumulative one-year failure rates from the two alternative trials are substituted for the one-year rates in the Bristol trial. The monthly hazard rates from month 13 to month 24 from the Bristol trial are then assumed to apply to those women from the alternative trials. The probability of a woman having an AH by two years given a repeat TCRE is assumed to retain its base-case value (ie. 40%). The hysterectomies undertaken on women randomised to TCRE are all assumed to be by the abdominal route.

The alternative parameter estimates are incorporated into the cost-utility model jointly; that is, as a multi-way sensitivity (or scenario) analysis to get a single estimate of the two-year expected costs and QALYs for the two treatments. The parameter estimates provided by the alternative trials have not been pooled to provide a single set of alternative parameters. As discussed in Section 6.3 above, the reason for this is that a meta-analysis of this type would, in effect, hide the variability between the Pinion *et al* and the Gannon *et al* trials, the importance of which this analysis is seeking to explore.

6.4.3 Results

The characteristics of the two trials in terms of their samples and of the process of care are detailed in Table 6.1, which also compares them with the Bristol RCT [Dwyer *et al*, 1993]. As the table shows, the two alternative trials are broadly similar in design to the Bristol trial. However, there are differences between them which, by contributing alternative parameter estimates, will provide one useful assessment of the

generalisability of the base-case analysis. These differences include that fact that the Gannon *et al* trial was undertaken in a district general hospital rather than a medical school; that the Pinion *et al* study was undertaken in Scotland rather than England and somewhat later than the other two trials; and that both the Gannon *et al* and the Pinion *et al* studies used drug therapies to prepare the uteri of women randomised to TCRE prior to surgery. An important difference between Pinion *et al* and the other two studies was that women not randomised to AH could undergo either laser ablation or TCRE. For this analysis, only the TCRE results are used.

Although the ages of the women in the three trials are similar, it is not easy to assess the relative severity of the condition in the three groups. The Gannon *et al* trial provides little detail of pre-operative severity; the information published on the other two trial cohorts would suggest that the extent of menorrhagia was worse in women in the Pinion *et al* study, but that more women experienced dysmenorrhoea in the Bristol trial.

Table 6.1 Characteristics of the two RCTs used for alternative parameter estimates compared with the Bristol RCT used in the base-case analysis

Characteristics	Alternative Parameter estimates		Base-case
	Gannon et al [1990]	Pinion et al [1994]	Dwyer et al [1993]
Inclusion criteria	Women awaiting AH for menorrhagia	Women aged under 50 years with menorrhagia who would otherwise have hysterectomy	Women aged under 52 years with menorrhagia which could not be controlled by conservative means and were candidates for hysterectomy
Exclusion criteria	Women known to have: leiomyomata, endometrial or cervical neoplasia; ovarian pathology; pelvic inflammatory disease; endometriosis	1. Weight of 100kg or more 2. Uterine size of 10 weeks or more 3. Abnormal histology	1. Uterine size greater than 12 weeks 2. If additional symptoms or other pathology made AH the preferred treatment
Treatments evaluated	TCRE versus AH	TCRE (loop and rollerball) or laser ablation (Nd:YAG) (conservative) versus AH	TCRE (loop) versus AH
Location	Royal Berkshire Hospital, Reading	Aberdeen Royal Infirmary, Aberdeen	St Michael's Hospital, Bristol
Pre-operative endometrial thinning	Medroxyprogesterone acetate for TCRE	Goserelin for conservative arm	None
Prophylactic antibiotics	To all women	Not stated	To all women
Recruitment and withdrawals	25 (TCRE) and 26 (AH). No withdrawals detailed	204 agreed to participate; 2 withdrew prior to surgery and 4 refused allocated treatment. Treatments given: 99 (AH), 52 (TCRE) and 51 (laser)	200 agreed to participate; 3 (AH) and 1 (TCRE) withdrew after randomisation and prior to surgery. This left 97 (AH) and 99 (TCRE) in trial
Age (years)	Median (range): AH 40 (29-47) TCRE 40 (29-51)	Mean (SD): AH 40.3 (5.2) TCRE 40.5 (4.5) Laser 39.6 (5.6)	Mean: 40.6 40.4
Duration of symptoms	Not detailed	% more than 2 years: AH 84% TCRE 88% Laser 91%	% more than 5 years: AH 39% TCRE 34%
Extent of symptoms	Not detailed	% more than 7 days bleeding per month: AH 58% TCRE 69% Laser 49% % with dysmenorrhoea AH 69% TCRE 75% Laser 75%	% more than 7 days bleeding per month: AH 37% TCRE 34% % with dysmenorrhoea AH 87% TCRE 86%

Table 6.2   Analysis of Generalisability I: parameters for which alternative estimates have been taken from the two other RCTs, together with their base-case values

Parameter	Alternative parameter estimates				Base-Case	
	Gannon *et al* [1991]		Pinion *et al* [1994]		Dwyer *et al* [1993]	
	AH	TCRE	AH	TCRE	AH	TCRE
Mean time in theatre (mins)*						
- Without complications	66.3$^{\text{s}}$	45.5$^{\text{s}}$	76.4$^{\text{s}}$	54.9$^{\text{s}}$	62.75	50.12
- With complications	66.3$^{\text{s}}$	45.5$^{\text{s}}$	76.4$^{\text{s}}$	54.9$^{\text{s}}$	63.16	63.75
Mean hospital stay (days)						
- Without complications	7.1$^{\text{s}}$	1.4$^{\text{s}}$	7.3$^{\text{s}}$	2.5$^{\text{s}}$	6.19	1.99
- With complications	7.1$^{\text{s}}$	1.4$^{\text{s}}$	7.3$^{\text{s}}$	2.5$^{\text{s}}$	6.64	3.25
Overall complication rate (%)	46.15	0.00	ID$^{\text{tt}}$	ID$^{\text{tt}}$	38.14	8.08
Haemhorrage (%)†	ID$^{\text{tt}}$	0.00	5.15	0.95	3.09	1.01
Uterine perforation (%)	ID$^{\text{tt}}$	0.00	0.00	0.95	0.00	4.04
Fluid overload (%)	ID$^{\text{tt}}$	0.00	1.03	11.43	0.00	1.01
Wound haematoma (%)	ID$^{\text{tt}}$	0.00	14.43	0.00	1.03	0.00
Pelvic haematoma (%)	ID$^{\text{tt}}$	0.00	11.34	0.00	8.25	1.01
Late complication (prior to discharge) (%)	ID$^{\text{tt}}$	0.00	4.12	2.86	4.12	1.01
One repeat TCRE (%)	-	16.00‡	-	10.48‡	-	12.06*
Hysterectomy (%)	-	0.00‡		16.19‡	-	12.17*
Cost of uterine pre-treatment	£0.00	£5.35¶	£0.00	£144‡‡	£0.00	£0.00
Mean time until work (days)						
- Without complications	67.6$^{\text{s}}$	14.9$^{\text{s}}$	70.00$^{\text{s}}$	21.00$^{\text{s}}$	81.41	16.24
- With complications	67.6$^{\text{s}}$	14.9$^{\text{s}}$	70.00$^{\text{s}}$	21.00$^{\text{s}}$	74.55	32.97

*	As for base-case analysis, 15 minutes are added to the operation length to allow for preparation and recovery of woman
s	No information available on distinction between with and without complications, hence overall mean used
†	Requiring blood transfusion
††	Base-case figures used
‡	Based on one-year follow-up
**	Based on two-year follow-up
¶	150mg Depo-Provera (Upjohn)
‡‡	3.6mg (as acetate) Zoladex (Zeneca)
ID	Insufficient detail provided

Table 6.3   **Analysis of Generalisability I: two-year results of the CUA using
               alternative parameter estimates from the two other RCTs**

Results	Alternative parameter estimates				Base-case	
	Gannon *et al* [1991]		Pinion *et al* [1994]		Dwyer *et al* [1993]	
	AH	TCRE	AH	TCRE	AH	TCRE
Expected costs						
Initial surgery						
- Theatre	£302	£219	£333	£244	£292	£234
- In-patient	£850	£168	£874	£299	£761	£250
- Complications	£39	£0	£49	£18	£32	£5
- Uterine pre-treatment	£0	£5	£0	£144	£0	£0
- Other*	£45	£34	£45	£34	£45	£34
Re-treatment	£0	£223	£0	£414	£0	£264
Longer term other†	£9	£7	£9	£8	£9	£7
Total costs	£1245	£656	£1310	£1161	£1139	£794
Expected QALYs	1.597	1.359	1.596	1.363	1.593	1.363
Incremental cost of AH per additional QALY	£2475		£639		£1500	

*   Includes pre-operative, post-operative and general practice costs (until 4 months after
        initial surgery) and other related costs until 2 years after surgery
†   Includes costs of cervical cytology and hormone replacement therapy

Using the criteria described above, the parameters within the CUA for which
alternative estimates have been taken from the two trials are detailed in Table
6.2. For purposes of comparison, the base-case value for each parameter, taken
from the Bristol RCT, are also presented.

Table 6.3 provides details of the results of the sensitivity analyses using
alternative parameter estimates from the other two trials; the two-year results
from the base-case analysis are presented again for comparison. The expected
health service costs of AH are broadly similar for the three trials. Compared to
the base-case results, taking parameters from the Gannon *et al* trial increases the
expected total cost of AH by 9%, due largely to the fact that women remained
in hospital for an average of about seven days compared to about six in the
Bristol trial. Using parameters from the Pinion *et al* RCT increases the expected
two-year cost of AH by 15%, again due to a longer length of stay in hospital and
a longer period of time in theatre.

The variation between the trials in the expected two-year costs of TCRE, on the basis of the alternative trial results, is more pronounced. Using data from the Gannon *et al* trial results in an 8% reduction in expected two-year costs relative to the base-case, due principally to a shorter mean length of stay in hospital. Using parameters from the Pinion *et al* trial, however, results in a 46% increase in two-year cost compared to the base-case. This is due to a number of factors, but most importantly to the 16% hysterectomy rate by one year in the Pinion *et al* trial, which is equivalent to a 20% rate when extrapolated to two years; and the fact that the women randomised to TCRE in the Pinion *et al* trial underwent uterine preparation using an expensive drug (ie. goserelin). The differential cost of AH over TCRE, varies between £149 (Pinion *et al*) and £589 (Gannon *et al*), compared to the base-case of £345.

Compared to the effect on expected costs, the impact of using parameters from the two alternative trials on expected two-year QALYs is modest (Table 6.3). Combining the expected costs and QALYs from the alternative parameter estimates results in a higher incremental cost per additional QALY of £2475 based on data from Gannon *et al*; and a lower ratio of £639 per additional QALY using parameters taken from the Pinion *et al* trial. Compared to a base-case estimated ratio of £1500, these sensitivity analyses show some degree of variation in costs and benefits which may partly reflect variation in routine clinical practice. However, if the illustrative cost per QALY thresholds suggested in Chapter 5 are considered acceptable, AH would remain the more cost-effective therapy on the basis of the alternative trial results.

6.5 Analysis of Generalisability II: routine clinical practice

6.5.1 Purpose

The use of two alternative trials to provide data for sensitivity analysis facilitates some assessment of the generalisability of the base-case analysis because they were undertaken in different hospitals, by different clinicians, using different processes and at different times to the Bristol RCT. However, as discussed in

Section 6.3 above, the high internal validity of RCTs often comes at the expense of limited generalisability, and it is possible that the sensitivity analysis presented in Section 6.4 may still not reflect true routine practice.

An alternative approach is to use non-trial data to provide alternative estimates for the key parameters in the cost-utility model. Ideally, these non-trial data would take the form of detailed observational data about the resource and non-resource consequences of a large number of procedures in a range of clinical settings. Although these sorts of data are rare in relation to most health care technologies, such data are beginning to emerge on treatments for menorrhagia.

In 1993, the Royal College of Obstetricians and Gynaecologists' (RCOG) Medical Audit Unit began a detailed survey of the use of non-hysterectomy forms of surgery for the treatment of menorrhagia (the Minimally Invasive Surgical Techniques Laser, Endothermal Or Endoresection (MISTLETOE) survey). Over a period of 18 months, data were collected on an estimated 80% of all cases undertaken, and this information related to the surgeon, procedure and post-operative phase. Questionnaires are currently being sent directly to the women in the survey to assess their health status and use of health service resources one year after surgery, and data are available for a sub-group of these women. The RCOG Medical Audit Unit has given access to MISTLETOE data which offers a further means of assessing the generalisability of the base-case CUA results.

The purpose of this second analysis of generalisability, therefore, is to use the MISTLETOE survey as an alternative source of parameter estimates in relation to TCRE for the cost-utility model, to retain the base-case estimates of the costs and effects of AH and to compare the resulting incremental cost per additional QALY of the two technologies with that of the base-case analysis.

6.5.2 Methods

The MISTLETOE survey collected data on 10,686 women. As shown in Table 6.4, the survey collected information on women undergoing a range of

Table 6.4   The non-hysterectomy surgical procedures undertaken on women with menorrhagia entered into the MISTLETOE survey

Procedure	Number of cases	%
TCRE with rollerball and loop diathermy	4279	40.04
TCRE with loop diathermy	3740	35.00
Laser	1785	16.70
TCRE with rollerball diathermy	644	6.03
Radiofrequency	136	1.27
Other	59	0.55
Cryoablation	36	0.34
Other combinations	7	0.07

non-hysterectomy surgical treatments for menorrhagia. The survey distinguished between two types of TCRE: that based on loop diathermy and that using rollerball diathermy. For the current analysis, data relating to loop resection alone are used, as this was the procedure employed in the Bristol RCT.

Although not offering the array of variables collected in a RCT, the MISTLETOE survey provides information on the key resource and non-resource consequences of a large number of TCREs, which is important for an economic analysis of the technology in comparison with AH. The criteria used to select which parameters to re-estimate using MISTLETOE data are the same as those described in Section 6.4.2 above in relation to the two alternative RCTs. As for the first analysis of generalisability, a multi-way sensitivity analysis is undertaken using all the re-estimated parameters.

The re-treatment data relate to one-year follow-up of all women in the survey using a postal questionnaire. That follow-up is not complete, and currently 1751 (47%) women, who initially underwent a TCRE with loop diathermy, have responded to the questionnaire. In order to estimate re-treatment rates at one year on the basis of the questionnaire data and using comparable women to those in the Bristol RCT, those women who had undergone previous uterine surgery are excluded. The hysterectomy rate at one year used in the analysis is

based only on those hysterectomies undertaken as the first re-treatment; that is, it excludes hysterectomies undertaken after a repeat TCRE. The probability of a woman having a hysterectomy given that she had previously undergone a repeat TCRE is assumed to retain its base-case value. The rates of repeat TCRE and of hysterectomy at one year are extrapolated to two years using the same methods as for the first analysis of generalisability. The hysterectomies undertaken on women who initially had TCREs are all assumed to be by the abdominal route.

The MISTLETOE survey collected detailed information on the rate of operative and early post-operative complications. For the purposes of the current analysis, a complication is defined as having occurred if either an immediate or post-operative complication prior to discharge was noted. In addition to the overall complication rate, information on only three specific complications was requested in the MISTLETOE survey: haemorrhage, perforation and post-operative complication (prior to discharge). The probabilities of the other possible complications retain their base-case values in the current analysis. As for the base-case analysis using data from the Bristol RCT, mean time in theatre, length of hospital stay and time until return to work are calculated separately according to whether or not complications were experienced.

The survey collected information on whether a woman was prescribed endometrial thinning medication prior to surgery, and these data are used to calculate the expected cost of endometrial thinning. For women who were prescribed such medication, the broad category of drug was noted: progestogen, danazol, LHRH analogue and other. These have been costed using standard British National Formulary dosages for the most frequently used drug in each category, based on clinical opinion. Other drugs are assumed to have the equivalent cost of the average of the other three categories.
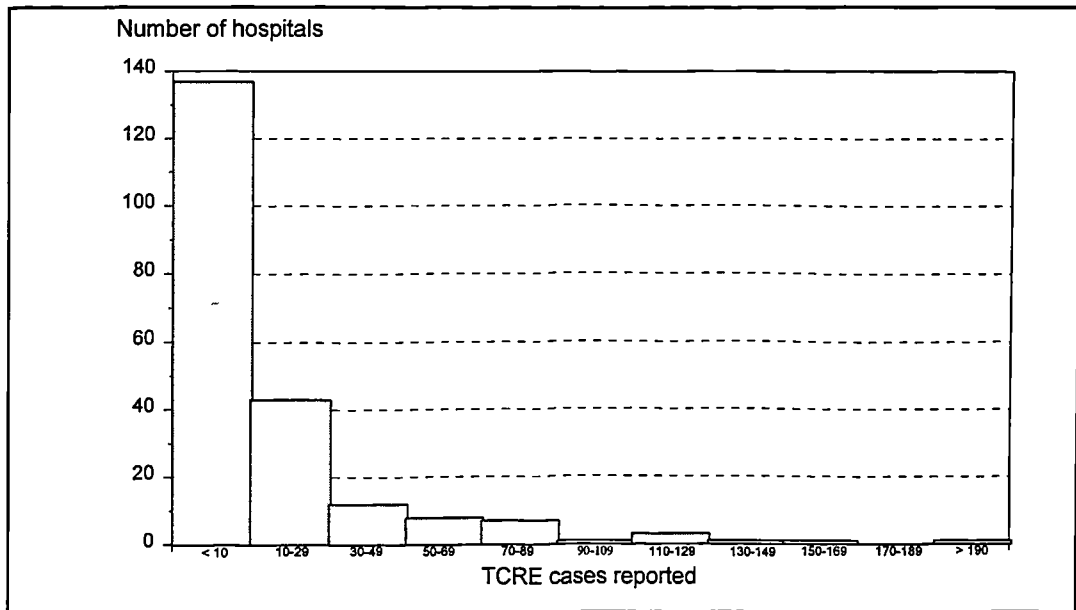
Figure 6.1   Distribution of cases reported by hospitals in the MISTLETOE survey

6.5.3 Results

The distribution of the number of cases of TCRE reported by hospitals in the
survey is shown in Figure 6.1. The figure indicates that the majority of hospitals
doing TCRE (64%) reported fewer than 10 cases in total over the period of the
survey. Given the large number of women being referred to hospital for
menorrhagia (see Chapter 1), this figure seems to suggest either that these
referrals are unevenly spread across units, that there was significant under-
reporting of cases from some centres in the MISTLETOE survey or that
hysterectomy was the first-line surgical treatment in most centres.

The key pre-operative characteristics of women undergoing TCRE in the
MISTLETOE are detailed in Table 6.5. Where similar data exist, the details of the
TCRE group in the Bristol RCT underlying the base-case analysis are presented
for comparison. The table shows that the survey and trial women are broadly
similar in terms of age, use of any type of drug therapy prior to surgery and use
of danazol in particular. The clearest difference between the two TCRE samples
is that the women in the trial had experienced symptoms of menorrhagia for
longer than those women in the MISTLETOE survey. This difference may be a

Table 6.5   Analysis of Generalisability II: pre-operative characteristics of the women
               in the MISTLETOE survey undergoing TCRE with loop diathermy
               compared to women randomised to TCRE in the Bristol RCT

Characteristic	MISTLETOE	Bristol RCT
Mean age (95% CI)	42.30 (42.1 - 42.5)	40.4 (39.4 - 41.5)
Mean duration of symptoms (mths) (95% CI)	35.88 (34.6 - 37.2)	61.08 (50.16 - 72.00)
Previous uterine surgery (%)	4.72	0.00
Previous medical therapy (%)*		
- None	15.57	14.29
- Progestogens	49.67	-
- Danazol	16.38	17.50
- NSAID	12.23	30.30
- Contraceptive pill	3.14	11.25
- Other	3.01	21.33
- Norethisterone	-	80.21

*   In Bristol trial women may have taken more than one drug

result of the observed shifting of thresholds in referral for surgical treatment for
menorrhagia following the diffusion of minimal access methods [Coulter, 1994;
Bridgman, 1994]. In other words, since the Bristol RCT took place, women and
their clinicians may be willing to accept surgical treatment earlier than when
hysterectomy was the only surgical option available.

The key point about the comparability of the Bristol RCT and MISTLETOE
samples, however, is that the latter survey focuses on women who are *currently*
undergoing TCRE for menorrhagia. It is quite feasible that the type of woman
receiving TCRE has changed between 1991 and 1994, but this is the very point
of undertaking an analysis of generalisability: to assess whether the conclusions
of the base-case (RCT-based) analysis are robust to alternative parameter
estimates generated from data relating to current routine practice.

The parameters that have been re-estimated from the MISTLETOE database for
the sensitivity analysis are detailed in Table 6.6, together with their base-case
values. The table shows a shorter mean period of stay in hospital in MISTLETOE
compared to the Bristol RCT. Figure 6.2 focuses on this variable in particular by

Table 6.6   **Analysis of Generalisability II: parameters for which alternative estimates have been taken from MISTLETOE survey for TCRE, together with their base-case values**

Parameter	Source of Estimates	
	MISTLETOE	Base-Case
Mean time in theatre (mins)*		
- Without complications	39.89	50.12
- With complications	49.05	63.75
Mean hospital stay (days)§		
- Without complications	1.18	1.99
- With complications	2.25	3.25
Overall complication rate (%)	7.06	8.08
Haemhorrage (%)†	0.17	1.01
Uterine perforation (%)	2.27	4.04
Late complication (prior to discharge) (%)	1.53	1.01
One repeat TCRE (%)	1.87**	12.06‡
Hysterectomy (%)¶	9.00**	12.17‡
Cost of uterine pre-treatment	£74	£0

*	As for base-case analysis, 15 minutes are added to the operation length to allow for preparation and recovery
§	Day-case surgery is taken as 0.5 days
†	Requiring blood transfusion
¶	Only those hysterectomies undertaken as first re-treatment
**	Based on one-year follow-up
‡	Based on two-year follow-up

presenting the distribution of lengths of stays in MISTLETOE compared to the trial. The major difference between the two data sources is that TCREs are now routinely undertaken as day-case procedures: 35% of TCRE cases in MISTLETOE were undertaken on this basis compared to none in the Bristol RCT. The mean stay in theatre was also shorter in the MISTLETOE survey, which may partly reflect further development of the procedure. Another notable difference is the lower rates of perforation and haemorrhage, but a higher rate of late complications in MISTLETOE.
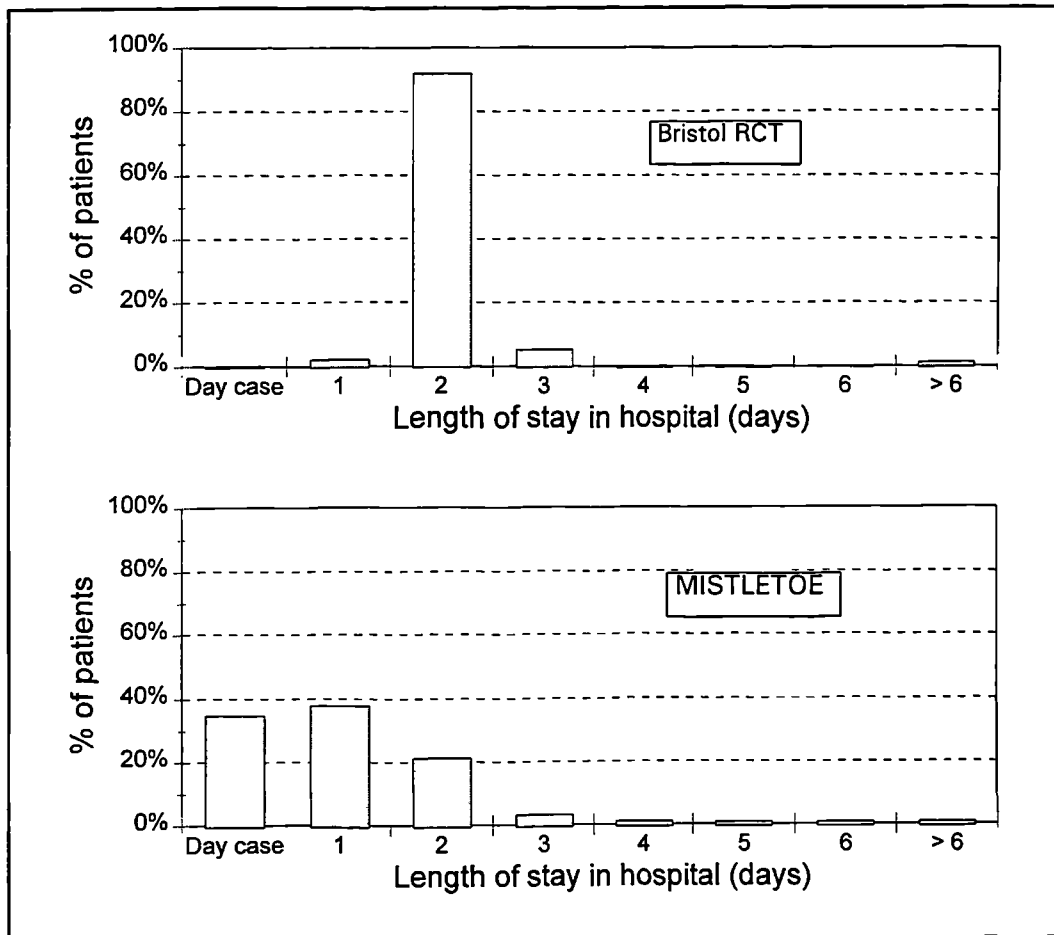
Figure 6.2   Distribution of lengths of stay in hospital following TCRE in the Bristol RCT and in the MISTLETOE survey

As regards re-treatment rates, the rate of second TCRE in the MISTLETOE survey was 1.87% at one year which, when extrapolated to two years, is 3.29%; the hysterectomy rate at one year is 9.00% which translates to 13.01% at two years. This compares with 12.06% and 12.17%, respectively, at two years in the Bristol RCT. This may suggest that a learning process has gone on since the Bristol RCT: if the first TCRE fails, then a second one is generally not felt to be worthwhile and a hysterectomy is, for many women, the sensible option.

The cost of uterine pre-treatment is based on the following rates of drug use: 17% of women used no thinning agent; 6% used a progestogen; 55% used danazol; 19% used a LHRH analogue and 2% used other agents. The overall

Table 6.7   Analysis of Generalisability II: two-year results of the CUA using
                 alternative parameter estimates from the MISTLETOE survey. Changes
                 relate only to TCRE with no re-estimation for AH

Results	Source of Estimates	
	MISTLETOE	Base-case
Expected costs		
Initial surgery		
- Theatre	£206	£234
- In-patient	£150	£250
- Complications	£5	£5
- Uterine pre-treatment	£74	£0
- Other*	£34	£34
Re-treatment	£182	£264
Longer term other†	£7	£7
Total costs	£658	£794
Expected QALYs	*1.371*	*1.363*
Incremental cost of AH per additional QALY¶	£2167	£1500

*   Includes pre-operative, post-operative and general practice costs (until 4 months after
       initial surgery) and other related costs until 2 years after surgery
†   Includes costs of cervical cytology and hormone replacement therapy
¶   Using base-case results for AH

expected cost of £74 is all additional to the base-case, as women in the Bristol

RCT were not prescribed these drugs.

Table 6.7 presents the revised estimates, using MISTLETOE data, of the

expected costs and QALYs associated with TCRE two years after surgery.

Expected costs clearly reflect the parameters in Table 6.6: using MISTLETOE

data, the costs of theatre and in-patient stay are lower than the base-case,

reflecting shorter mean durations in theatre and in hospital. Other costs

associated with the initial procedure are higher as a result of the fact that 93%

of women undergoing TCRE with loop diathermy in MISTLETOE were prescribed

some form of uterine thinning agent prior to surgery; and re-treatment costs are

lower because fewer women underwent repeat TCRE.

Expected QALYs are slightly higher on the basis of MISTLETOE data, reflecting the lower mean complication and failure rates. When the expected two-year costs of TCRE are combined with the cost of AH estimated in the base-case, the incremental cost of an additional QALY with AH is higher on the basis of MISTLETOE data: £2167 compared to £1500 in the base-case. This revised ratio remains below the lower illustrative cost per QALY threshold introduced in Chapter 5, suggesting that AH would remain the more cost-effective option if the threshold ratio were considered acceptable.

6.6 Analysis of Generalisability III: high and low resource use in routine clinical practice

6.6.1 Purpose

The last sensitivity analysis, using mean values relating to TCRE with loop diathermy from the MISTLETOE survey, does not fully reflect the *variation* in the resource and non-resource consequences of treatment within routine clinical practice. For example, although the overall hysterectomy rate following TCRE in the survey was 9% at one year, some hospitals would have had higher rates and some lower rates than this. This third analysis of generalisability, therefore, uses MISTLETOE data to explore the implications, for the relative value for money of TCRE and AH, of these variations in clinical practice.

6.6.2 Methods

The general approach is the same as for the previous two; namely, to re-estimate key parameters in the cost-utility model using alternative data sources. As the MISTLETOE survey is again the alternative data source, the same parameters are re-estimated as in the second analysis of generalisability (Table 6.6). However, instead of using mean values and overall proportions, the results at the high and low ends of the resource use distributions are the focus. The mean values or rates are calculated *by hospital*. For an analysis of resource intensive clinical practice, the upper quartiles of the mean values or rates are incorporated into the

model; for an analysis of resource sparing clinical practice, the lower quartile of the mean values or rates are used[1].

In general, the methods used within the analysis are the same as for the second analysis of generalisability detailed above, apart from the following differences. Firstly, because complications are generally rare, it is not possible to get a good estimate, by hospital, of duration in theatre and in hospital for cases with complications because the numbers are too few. Therefore, the upper and lower quartile mean values of these durations by hospital do not differentiate according to complication status.

A second difference between the methods used in the last analysis of generalisability and this one relates to the use of endometrial thinning agents. The second analysis of generalisability showed that, of those women who were prescribed some form of endometrial thinning agent, 89% used either danazol or a LHRH analogue. The more expensive of these two agents is usually the LHRH analogue, which would normally be goserelin (typically, £144 per patient versus £78 per patient with danazol). Therefore, the rate of LHRH analogue use is calculated by hospital. For the resource intensive analysis, the expected cost of the endometrial thinning agent is calculated using the cost of goserelin for the upper quartile proportion of cases using LHRH analogues, and the average cost of the other three categories of drug (progestogen, danazol and other) for that proportion of cases for which LHRH analogues were not used. For the resource sparing analysis, the lower quartile rate of uterine pre-treatment by hospital is used, and an expected cost is calculated by multiplying that rate by the expected cost of uterine pre-treatment estimated on the basis of all MISTLETOE TCRE data in Section 6.5.2.

---

[1] It would be possible to estimate a more extreme scenario of high and low resource use, by incorporating maximum and minimum mean values and rates by hospital. However, given that these rates are incorporated individually (the values and rates of the hospitals with, for example, high lengths of stay, times in theater and complication rates are used simultaneously to represent a notional resource intensive hospital), and that some hospitals reported very few cases which may have led to their having extreme and unrepresentative mean values, the use of upper and lower quartile values and rates are considered more appropriate.

Table 6.8   **Analysis of Generalisability III: parameters for which alternative estimates have been taken from the MISTLETOE survey for the analyses of resource intensive and resource sparing clinical practice related to TCRE, together with their base-case values**

Parameter	Sources of Estimates					
	MISTLETOE					Base-case
	No. of hospitals	Mean no. of cases	Mean of the means	Upper quartile of the means	Lower quartile of the means	
Mean time in theatre (mins)*						
- Without complications	198	17.47	46.23	53.33	36.88	50.12
- With complications	198	17.47	46.23	53.33	36.88	63.75
Mean hospital stay (days)°						
- Without complications	205	17.48	1.46	2.00	0.93	1.99
- With complications	205	17.48	1.46	2.00	0.93	3.25
Overall complication rate (%)	212	17.48	9.16	12.50	0.00	8.08
Haemhorrage (%)†§	212	16.65	0.15	0.00	0.00	1.01
Uterine perforation (%)§	212	16.65	2.99	0.00	0.00	4.04
Late complication (prior to discharge) (%)§	212	16.65	2.09	0.00	0.00	1.01
One repeat TCRE (%)	174	9.20	1.57**	0.00**	0.00**	12.06‡
Hysterectomy (%)¶	174	9.20	9.83**	13.64**	0.00**	12.17‡
No uterine pre-treatment (%)	214	17.48	16.07	20.00	0.00	100.00
LHRH analogue for uterine pre-treatment (%)	212	17.48	21.09	33.33	0.00	0.00

*	As for base-case analysis, 15 minutes are added to the operation length to allow for preparation and recovery
ø	Day-cases have a length of stay of 0.5 days
†	Requiring blood transfusion
¶	Only those hysterectomies undertaken as first re-treatment
§	Only calculated for cases where data on whether or not there were any complications are available
**	Based on one-year follow-up
‡	Based on two-year follow-up

It is important to note that data on the high and low resource using hospitals only relate to TCRE procedures. Hence, comparing these extremes with the base-case estimates of the costs and effects of AH masks the fact that there is undoubtedly significant variation in clinical practice relating to hysterectomy.

Table 6.9   **Analysis of Generalisability III: two-year results of the cost-utility analysis using alternative parameter estimates from the MISTLETOE survey relating to resource intensive and resource sparing clinical practice. Changes relate only to TCRE with no re-estimation for AH**

Results	MISTLETOE		Base-case
	Resource sparing	Resource Intensive	
Expected costs			
Initial surgery			
- Theatre	£196	£239	£234
- In-patient	£111	£239	£250
- Complications	£0	£2	£5
- Uterine pre-treatment	£59	£93	£0
- Other*	£34	£34	£34
Re-treatment	£51	£200	£264
Longer term other[t]	£7	£7	£7
Total costs	£458	£814	£794
Expected QALYs	1.377	1.374	1.363
Incremental cost of AH per additional QALY[1]	£3153	£1484	£1500

*   Includes pre-operation, post-operative and general practice costs (until 4 months after
           initial surgery) and other related costs until 2 years after surgery
t   Includes costs of cervical cytology and hormone replacement therapy
1   Using base-case results for AH

6.6.3 Results

Table 6.8 provides details of the parameters that have been re-estimated using

MISTLETOE survey data and incorporated into the cost-utility model to assess

the costs and effects of TCRE in resource intensive and resource sparing clinical

practice. The table shows, for each mean value and rate analysed by hospital,

the mean, upper and lower quartile, as well as the number of hospitals and mean

number of cases per hospital providing the estimates; the base-case estimate of

each parameter is also shown. Considerable variation between hospitals is

evident in all areas of resource use, but perhaps the most significant variation is

in the rate of complications: the proportion of cases experiencing immediate or

post-operative (pre-discharge) complications has an inter-quartile range of 0% to 12.5%; although the incidence of costly complications is small. The table shows that, within a year of initial surgery, most hospitals had no repeat TCREs, but that the upper quartile rate of hysterectomy was 13.64%. This seems to emphasise the point made in Section 6.5.3 above that clinicians and women are increasingly eschewing repeat TCREs given initial failure, with many moving straight to hysterectomy.

The implications for this variation in the process and results of clinical practice for the costs and benefits of TCRE are shown in Table 6.9, together with the base-case results. The resource sparing analysis incorporates all the lower quartile values and rates into the cost-utility model, which results in a total two-year expected cost of only £458, compared to £794 in the base-case. These cost reductions come largely from a shorter length of hospital stay and less re-treatment. If the costs and benefits of the resource sparing hospitals are compared with the base-case estimates for AH, the incremental cost of AH per additional QALY is £3158, compared to £1500 in the base-case.

In the case of the resource intensive hospitals, the total two-year expected cost is £814. This is only 2.5% more than the base-case estimate, reflecting the modest differences, relative to the base-case, in length of stay in hospital and theatre, and the limited number of second TCREs and expensive complications. The higher cost of this resource intensive scenario is largely generated by the higher hysterectomy rate and the use of LHRH analogues for uterine pre-treatment. Compared to the base-case costs and benefits of AH, the incremental cost of each additional QALY under the resource intensive scenario falls modestly to £1484.

If the illustrative cost per QALY thresholds introduced in Chapter 5 are acceptable, then the revised cost and benefit results under the resource intensive and sparing scenarios would fail to alter the conclusion that, under most circumstances, AH is more cost-effective than TCRE.

6.7   Analysis of Generalisability IV: variation in unit costs

6.7.1 Purpose

The three analyses of generalisability described above all focus on the implications of variations in the resource and non-resource consequences of TCRE and AH compared to the base-case analysis. However, as shown in Chapter 5, a major source of uncertainty relating to the relative value for money of TCRE and AH is the unit cost of particular resources. This analysis uses unit cost estimates from specific hospitals and explores their impact on the relative cost of the two interventions and, in turn, on their relative cost-effectiveness.

6.7.2 Methods

Two unit costs are the focus of this sensitivity analysis. The first of these is the ward cost per day which, as shown in Chapter 5, has an important impact on the differential cost of TCRE and AH and is known to vary considerably between hospitals [CIPFA, 1990]. The second unit cost considered is the variable cost of a minute in theatre, excluding the cost of anaesthetics and staff. Together, these two unit costs determine the bulk of the two-year expected costs of both procedures.

The base-case values of these two unit costs were taken from available data sources. The ward cost per day was based on estimates by the hospital in which the Bristol RCT was based; the theatre cost came from a national study [Bevan 1989] (see Chapter 3). In seeking to assess the robustness of the base-case conclusions to the value of these unit costs, it is important to be aware that differences in unit costs between hospitals may reflect divergence in costing methods more than genuine cost differences. When the unit costs were estimated for the base-case analysis, the quality of unit cost data available in the NHS was generally poor. Since then, improvements in information systems and national guidelines [NHS Executive, 1993] have resulted in improvements in cost data, but it is difficult to identify hospital-specific unit costs which are comparable, in terms of costing methods, to those used in the base-case. Therefore, the approach taken has been to acquire three sets of alternative unit

Table 6.10   Analysis of Generalisability IV: estimates of alternative hospital-specific
                unit costs for an in-patient day in hospital and a minute in theatre (June
                1994 prices)

Unit cost	Alternative unit costs			Base-case
	Hospital 1	Hospital 2	Hospital 3	
In-patient day in gynaecology ward	£83.33	£89.72	£104.00	£120.00
Minute in theatre	£2.04	£1.96	£2.02	£1.08

costs from specific hospitals which use similar information systems and costing
methods, and to compare the implications of these unit costs for the expected
two-year costs of the two procedures.

The three hospitals with broadly similar costing methods were identified through
their use of the same financial information software. Out of four hospitals
contacted with a view to acquiring unit cost data from them, three were able and
willing to provide an estimate of the cost of a day on a gynaecological ward and
of a minute in theatre. These three hospitals were large teaching hospitals, one
based in London the other two in northern England. In requesting the estimates
of unit costs, a set of guidelines were given to the finance departments, and
these are shown in Appendix 6.1.

As for the other three analyses of generalisability, the sensitivity analysis
involved incorporating the alternative unit cost estimates into the model jointly.

6.7.3 Results

Table 6.10 presents the unit cost estimates from the three hospitals, together
with the values used in the base-case analysis. The table shows remarkable
consistency between the three alternative sources of hospital-specific unit costs,
with ward costs per day ranging between £83 and £104, and theatre costs per
minute ranging between £1.96 and £2.04. Compared to the base-case unit
costs, the most obvious difference is in the cost of a minute in theatre, with the
base-case value being only 54% of the mean cost of the three hospitals. This

Table 6.11   Analysis of Generalisability IV: impact of alternative unit costs on
              expected costs of AH and TCRE and on cost-utility ratios at two years

Results	Alternative unit costs						Base-case	
	Hospital 1		Hospital 2		Hospital 3			
	AH	TCRE	AH	TCRE	AH	TCRE	AH	TCRE
Expected costs								
Initial surgery								
- Theatre	£352	£283	£347	£279	£351	£282	£292	£234
- In-patient	£530	£174	£570	£187	£661	£217	£761	£250
- Complications	£32	£5	£32	£5	£32	£5	£32	£5
- Other*	£45	£34	£45	£34	£45	£34	£45	£34
Re-treatment	£0	£234	£0	£240	£0	£259	£0	£264
Longer term other†	£9	£7	£9	£7	£9	£7	£9	£7
Total costs	£968	£737	£1003	£752	£1098	£804	£1139	£794
Incremental cost of AH per additional QALY	£1004		£1091		£1278		£1500	

*   Includes pre-operation and general practice (until 4 months after initial surgery) and other
         related costs until 2 years after initial surgery
†   Includes costs of cervical cytology and hormone replacement therapy

difference reflects the fact that the hospital-specific theatre costs have been
estimated by hospitals using the same financial information system and very
similar costing methods based on NHS Executive guidelines. The base-case
estimate came from a specific study undertaken in the late 1980s, across all
types of theatre and probably using rather different costing methods.

The implications for expected costs and cost per QALY ratios of incorporating
these alternative sets of unit costs are shown in Table 6.11. The unit costs from
Hospitals 1 and 2 generate very similar total two-year expected costs; the
somewhat higher ward cost for Hospital 3 results in higher expected costs than
for the other two hospitals. Compared to the base-case, the alternative unit
costs result in lower total expected costs for both treatments with the one
exception of TCRE using unit costs from Hospital 3, where costs increase
slightly. The general reduction in expected costs using the alternative unit costs,

is slightly more pronounced for AH because a higher proportion of its costs consist of ward cost, which is lower for the three hospitals. The effect of this is to reduce the incremental cost of AH over TCRE, and the cost-utility ratios fall to between £1004 and £1278 from the base-case ratio of £1500. Therefore, this analysis of generalisability indicates that base-case conclusions are robust to unit costs from specific clinical centres.

6.8   Analysis of Generalisability V: alternative surgical methods

6.8.1 Purpose

The base-case analysis compares the costs and benefits of TCRE and AH. Although these procedures represent the main surgical treatments for menorrhagia, other forms of surgery - both non-hysterectomy and hysterectomy - are used in the UK. Table 6.4 shows that, on the basis of MISTLETOE survey data, TCRE using loop diathermy was used in 35% of non-hysterectomy MAS procedures. The table indicates that several other non-hysterectomy MAS techniques were used, including TCRE where rollerball diathermy was used in addition to or instead of a loop, laser ablation and radiofrequency.

Although, until recently, AH was used in about 88% of hysterectomies [Vessey *et al*, 1992], vaginal hysterectomy and laparoscopic-assisted vaginal hysterectomy are now being used in some centres [RCOG Medical Audit Unit, personal communication]. Therefore, the case-study of surgical treatment for menorrhagia emphasises the general point made in Chapter 2 about MAS technologies: that the comparator against which new MAS techniques need to be assessed will change over time and may not be conventional open surgery. In the case of menorrhagia, TCRE is becoming the old from of MAS which should be assessed against the new hysterectomy forms of MAS.

The focus of this thesis is the comparison of TCRE and AH, but it is of interest to consider if the results of this evaluation can be extended to a more general comparison of non-hysterectomy and hysterectomy forms of surgery for

menorrhagia. In order to do this, it is necessary to consider the key differences between TCRE and other non-hysterectomy MAS techniques, and between AH and other forms of hysterectomy, in terms of resource and non-resource consequences. Therefore, this final analysis of generalisability adjusts the parameters in the base-case model in order to estimate the expected costs and benefits of the alternative surgical treatments for menorrhagia.

It should be emphasised that the comparison of these technologies is indicative only and the results should be interpreted with care. Unlike the TCRE and AH comparison in the base-case, which is based on the results of a RCT, there is no source of data that compares the use of these other surgical treatments using a homogenous population and randomised treatment allocation. However, the analysis provides a useful 'broad-brush' picture of the relative costs and potential cost-effectiveness of these alternative techniques, and helps to highlight the priorities for further research.

6.8.2 Methods

In addition to the treatment options considered in the base-case analysis, the following alternative surgical interventions are considered.

(a)   *Non-hysterectomy forms of MAS other than TCRE with loop diathermy*
  (i)   TCRE using a combination of loop and rollerball diathermy;
  (ii)   TCRE using rollerball diathermy alone;
  (iii)   laser ablation; and
  (iv)   radiofrequency (RF) ablation.

(b)   *Types of hysterectomy other than AH*
  (I)   vaginal hysterectomy (VH); and
  (ii)   laparoscopic-assisted vaginal hysterectomy (LAVH)[1].

---

[1]A totally laparoscopic procedure is used in some centres but this is rare and is not considered here.

Two sets of parameters within the cost-utility model are adjusted to estimate the costs and benefits of these alternative treatments; namely, resource and non-resource consequences; and costs of procedures. In order to adjust the key parameters relating to non-hysterectomy forms of MAS, the MISTLETOE survey is used as the principal data source. Although not a RCT, the survey offers standardised data collection on a large number of women who have undergone treatment for menorrhagia with one of the four technologies listed above, as well as TCRE with loop diathermy alone.

The choice of parameters to re-estimate for the model is based on the same criteria detailed for the first analysis in Section 6.3, and the re-estimated parameters are the same as those used in the second and third analyses of generalisability. The non-resource consequences of these alternative treatments are only adjusted in terms of complications and failure rates; it is likely that the treatments will differ in terms of process and outcomes in ways that are not allowed for in this analysis. The assumptions used to re-estimate the resource use and non-resource use parameters are the same as those used in Analyses II and III above. It should be emphasised that, for women undergoing hysterectomy because their non-hysterectomy treatment has failed, it is assumed that an AH is performed.

No equivalent data source to the MISTLETOE survey yet exists to estimate the resource and non-resource consequences of the alternative types of hysterectomy in routine clinical practice for the model. Published literature and clinical opinion has, therefore, been used in order to estimate the key parameters - length of hospital stay and length of time in theatre - for the two alternative types of hysterectomy. The non-resource consequences of VH and LAVH, including complication rates, are assumed to be the same as for AH.

The costs of the alternative procedures have been estimated by adjusting the unit cost of TCRE with loop diathermy, for the non-hysterectomy forms of MAS,

Table 6.12   Analysis of Generalisability V: pre-operative characteristics of women
undergoing various forms of non-hysterectomy surgery for menorrhagia
in the MISTLETOE survey. Where available, the characteristics of
women in the Bristol RCT are also detailed

Characteristic	Taken from MISTLETOE					TCRE (L) (Bristol RCT) (n = 99)
	TCRE (L) (n = 3740)	TCRE (L + R) (n = 4279)	TCRE (R) (n = 644)	Laser (n = 1785)	RF (n = 136)	
Mean age (95% CI)	42.30 (42.1-42.5)	41.98 (41.8-42.2)	41.71 (41.2-42.2)	41.79 (41.5-42.1)	42.32 (41.4-43.2)	40.47 (39.4-41.5)
Mean duration of symptoms (mths) (95% CI)	35.88 (34.6-37.2)	35.73 (34.6-36.8)	38.56 (35.1-42.0)	38.08 (36.2-40.0)	57.13 (47.9-66.4)	61.03 (50.2-72.0)
Previous uterine surgery (%)	4.72	3.89	11.99	10.17	3.64	0.00
Current smokers (%)	25.69	27.17	28.87	27.41	25.00	-
Previous medical therapy (%)*						
- None	15.57	13.52	14.83	13.49	4.76	14.29
- Progestogers	49.67	51.92	47.03	49.93	52.38	-
- Danazol	16.38	15.87	14.83	19.34	15.87	17.50
- NSAID	12.23	11.96	18.22	8.55	11.11	30.30
- Contraceptive pill	3.14	3.13	2.12	2.99	4.76	11.25
- Other	3.01	3.61	2.97	5.69	11.11	21.33
- Norethisterone	-	-	-	-	-	80.21
Hysteroscopy (%)						
- No test	58.60	63.38	54.62	37.38	86.96	-
- Normal	32.02	28.93	38.46	50.31	8.69	-
- Polyps	3.13	3.42	3.08	4.92	1.45	-
- Fibroids	6.25	4.27	3.85	7.38	2.89	-
Normal histology (%)	95.31	95.58	97.71	93.19	86.11	-
Uterine pre-treatment (%)	82.59	86.27	89.25	97.32	98.39	0.00
One year follow-up (%)	46.82	43.05	42.24	42.58	32.35	-

*	In Bristol trial women may have taken more than one drug
TCRE(L)	Transcervical endometrial resection using loop diathermy
TCRE (L&R)	Transcervical endometrial resection using loop and rollerball diathermy
TCRE (R)	Transcervical endometrial resection using rollerball diathermy

and of AH, for the alternative forms of hysterectomy. The details of the
adjustment are shown in Appendix 6.2. The appendix indicates that, because of

important variation in the mix of disposable consumables versus reusable equipment as part of LAVH, two forms of the procedure are costed: 'disposable' and 'reusable' versions.

6.8.3 Results

As the MISTLETOE survey is used as a single data source to estimate the resource and non-resource consequences in routine practice of the non-hysterectomy forms of MAS for menorrhagia for this analysis of generalisability, Table 6.12 details the pre-operative characteristics of the women who underwent the procedures considered. In addition to the laser and RF ablation groups, three TCRE groups are shown: loop alone, loop plus rollerball and rollerball alone. Where available, the pre-operative characteristics of the women randomised to TCRE in the Bristol RCT are also detailed. The proportions of women in each of the treatment groups who had been sent and returned a one-year follow-up questionnaire are also detailed.

The table shows that the five treatment groups from the MISTLETOE survey were broadly similar. The main differences appear to be that a higher proportion of women in the laser and RF groups had undergone previous uterine surgery; that women in the RF group had experienced symptoms for a longer period than women in the other MISTLETOE groups, and were closer to women in the Bristol RCT in this respect; that more women in the RF group had tried medical therapy for menorrhagia prior to surgery; and that women undergoing laser and RF ablation were more likely to have been prescribed uterine thinning agents prior to surgery. The relatively small numbers of women in the RF group may explain its asymmetry compared the other groups, and the difference in the use of uterine pre-treatment is part of the process of care (ie. such preparatory therapy is effectively mandatory for laser and RF techniques). The main difference

Table 6.13   Analysis of Generalisability V: parameters re-estimated in the cost-utility model to assess the costs and benefits of alternative surgical treatments. The base-case values are also reported

Parameter	Non-hysterectomy*					Hysterectomy		Base-case†	
	TCRE (L)	TCRE (R&L)	TCRE (R)	Laser	RF	VH	LAVH	AH	TCRE
Mean time in theatre (mins)°									
- without complications	39.89	43.19	45.89	46.83	32.71	NC	120.00‡	62.75	50.12
- with complications	49.05	48.81	49.00	54.49	38.75	NC	120.00‡	63.16	63.75
Mean hospital stay (days)°°									
- without complications	1.18	1.16	1.45	1.07	1.47	4.00**	4.00**	6.19	1.99
- with complications	2.25	2.19	3.20	2.00	4.00	4.00**	4.00**	6.64	3.25
Overall complication rate (%)	7.06	5.35	3.10	3.36	4.80	NC	NC	38.14	8.08
Haemorrhage (%)§	0.17	0.07	0.00	0.00	0.00	NC	NC	3.09	1.01
Uterine perforation (%)	2.27	1.19	0.33	0.59	0.00	NC	NC	0.00	4.04
Late complication (%)	1.53	1.72	1.15	1.12	4.80	NC	NC	4.12	1.01
One repeat procedure (%)	1.87¶	1.78¶	1.72¶	5.70¶	10.26¶	-	-	-	12.06††
Hysterectomy (%)	9.00¶	8.44¶	12.87¶	9.60¶	23.08¶	-	-	-	12.17††
Cost of uterine pre-treatment	£74	£79	£92	£99	£99	£0	£0	£0	£0

*	All data taken from MISTLETOE survey
†	All data taken from Dwyer et al [1993]
‡	Taken from Summitt et al [1992]
††	Based on two-year follow-up
°	As for base-case analysis, 15 minutes added to the operation length to allow for preparation and recovery
°°	For non-hysterectomy options based on MISTLETOE data, a day-case procedure is taken as a 0.5 day stay
**	Clinical advice

§	Requiring blood transfusion
¶	Based on one-year follow-up

TCRE(L)	Transcervical endometrial resection using loop diathermy
TCRE (L&R)	Transcervical endometrial resection using loop and rollerball diathermy
TCRE (R)	Transcervical endometrial resection using rollerball diathermy
NC	No change from base-case (AH)

between the MISTLETOE groups and the Bristol RCT is that the trial women had experienced their condition for a longer mean time period. The possible reasons for this were discussed in Section 6.4 above.

Table 6.13 details the parameters which have been re-estimated within the model to assess the relative costs and benefits of the alternative non-hysterectomy and hysterectomy forms of surgery. TCRE using loop diathermy is the same as the technology considered in Analysis II. As regards the non-hysterectomy surgical options, the largest adjustments to the base-case analysis are the lower rates of complications with all treatments, but most notably with laser and TCRE with rollerball alone; the shorter lengths of hospital stay which, for women not experiencing complications, are day-case or one night; and the re-treatments rates which, at one year follow-up in the MISTLETOE survey, vary between 1.72% (TCRE with rollerball) and 10.26% (RF) for repeat procedures and between 9% (TCRE with loop diathermy alone) and 23.08% (RF) for hysterectomy. Only modest adjustments are made to the base-case parameters of AH to estimate the costs of VH and LAVH: a reduced length of stay and a longer period of time in theatre.

Table 6.14 details the two-year expected costs and QALYs of the various surgical options on the basis of the adjustment made to the base-case parameters. As regards the non-hysterectomy forms of surgery considered using MISTLETOE survey data, the TCRE procedures and laser ablation have similar expected costs, and they are not greatly different to the base-case estimates for TCRE using loop diathermy; but RF has a considerably higher expected cost, due largely to the higher repeat surgery rates shown in the survey.

As regards hysterectomy, the lower length of hospital stay that is assumed to be associated with VH and the 'reusable' version of LAVH results in a marked reduction in their expected total costs compared to the base-case estimates for AH. The significant additional cost of consumables associated with 'disposable' LAVH results in the procedure being by far the most expensive of the surgical procedures considered.

Table 6.14   Analysis of Generalisability V: expected costs and benefits of alternative surgical treatments for menorrhagia at two years

Results	Non-hysterectomy					Hysterectomy			Base-case	
	TCRE (L)	TCRE (R&L)	TCRE (R)	Laser	RF	VH	LAVH (disposable)	LAVH (reusable)	AH	TCRE
Expected costs										
Initial surgery										
- Theatre	£206	£214	£220	£274	£312	£292	£1462	£505	£292	£234
- In-patient	£150	£145	£180	£132	£190	£479	£479	£479	£761	£250
- Complications	£5	£4	£2	£2	£8	£32	£32	£32	£32	£5
- Uterine pre-treatment	£74	£79	£92	£99	£99	£0	£0	£0	£0	£0
- Other*	£34	£34	£34	£34	£34	£45	£45	£45	£45	£34
Re-treatment	£182	£175	£224	£230	£437	£0	£0	£0	£0	£264
Longer term other†	£7	£7	£7	£7	£8	£9	£9	£9	£9	£7
Total costs	£658	£658	£759	£778	£1088	£857	£2027	£1070	£1139	£794
Expected QALYs	1.371	1.372	1.371	1.367	1.360	1.593	1.593	1.593	1.593	1.363

*   Includes pre-operative, post-operative and general practice (until 4 months after initial surgery) costs
†   Includes costs of cervical cytology and hormone replacement therapy
TCRE(L)   Transcervical endometrial resection using loop diathermy
TCRE (L&R)   Transcervical endometrial resection using loop and rollerball diathermy
TCRE (R)   Transcervical endometrial resection using rollerball diathermy

LAVH (D)   Laparoscopic-assisted vaginal hysterectomy (disposable)
LAVH (R)   Laparoscopic-assisted vaginal hysterectomy (reusable)

VH   Vaginal hysterectomy
AH   Abdominal hysterectomy
RF   Radiofrequency ablation

The two-year expected QALYs associated with the alternative non-hysterectomy treatments differ little from the base-case estimates for TCRE. However, this is because complications and re-treatment rates are the only parameters related to non-resource consequences that are altered relative to the base-case. By assumption, the expected QALYs associated with VH and the two forms of LAVH are the same as the base-case estimates for AH.

Table 6.15 compares the expected costs and QALYs of each of the non-hysterectomy options with those of each of the hysterectomy options. The non-hysterectomy procedures include two estimates for TCRE with loop diathermy: the base-case results and those based on parameters estimated from the MISTLETOE survey. These comparisons are based on the assumption that the non-resource consequences of each of the forms of hysterectomy are the same as for AH (ie. the expected QALYs are identical), and that the only differences between the non-hysterectomy options in terms of non-resource consequences is in the incidence of complications and re-treatment.

On this basis, AH - the standard form of hysterectomy - has an incremental cost over the non-hysterectomy forms of surgery of between £219 and £2176 per additional QALY generated, compared to a base-case ratio of £1500 relative to TCRE. The incremental ratios resulting from the comparison of the various non-hysterectomy options with 'reusable' LAVH are similar to those relating to AH, although this form of hysterectomy dominates RF. The consumable cost with 'disposable' LAVH results in appreciably higher incremental costs per QALY in comparison with the non-hysterectomy options. However, the lower lengths of hospital stay with VH reduces the incremental costs per additional QALY, and the analysis indicates that VH dominates RF.

With reference to the illustrative cost per QALY thresholds suggested in Chapter 5, it would again seem that the broad conclusions of that chapter are generalisable. The data presented here suggest that, in comparison with AH and assuming the outcome differences are solely reflected in complication and re-treatment rates, the use of types of TCRE other than that relying only on the

Table 6.15   Analysis of Generalisability V: two year expected incremental cost per additional QALY of non-hysterectomy versus hysterectomy surgical procedures

Non-hysterectomy surgical procedures	Hysterectomy surgical procedures			
	AH (base-case)	VH	LAVH (D)	LAVH (R)
TCRE (L)	£2167	£896	£6167	£1856
TCRE (R&L)	£2176	£900	£6195	£1864
TCRE (R)	£1712	£441	£5712	£1401
Laser	£1597	£350	£5527	£1292
RF	£219	(VH) dom	£4030	(LAVH(R)) dom
TCRE (base-case)	£1500	£274	£5361	£1200

TCRE(L)	Transcervical endometrial resection using loop diathermy
TCRE (L&R)	Transcervical endometrial resection using loop and rollerball diathermy
TCRE (R)	Transcervical endometrial resection using rollerball diathermy
RF	Radiofrequency ablation
VH	Vaginal hysterectomy
AH	Abdominal hysterectomy
LAVH (D)	Laparoscopic-assisted vaginal hysterectomy (disposable)
LAVH (R)	Laparoscopic-assisted vaginal hysterectomy (reusable)
(*) dom	* dominates (ie. is less costly and more effective)

loop, or the use of laser or RF ablation, is unlikely to alter the base-case conclusion that AH is the more cost-effective treatment. If VH or 'reusable' LAVH rather than AH were the type of hysterectomy to be used, the base-case conclusion would be even firmer, unless these procedures are significantly less effective (eg. in terms of complications) than AH. However, if 'disposable' LAVH were compared to the non-hysterectomy forms of surgery, the base-case conclusion would be less firm, although the cost per QALY ratios would remain less than the lower illustrative threshold.

6.9   Discussion

6.9.1 Assessing generalisability in economic evaluation

It is rare to see a detailed consideration of the level of external validity characterising an economic evaluation. This is surprising given that so many of the data inputs in a typical analysis are specific to the context of their use, and hence likely to be influenced by the known variations in clinical practice. The dearth of formal analyses of generalisability is partly due to the widespread view that economic evaluation is a 'once and for all' process; that a technology is either cost-effective or not, and that the best way to test this is to attach an economic analysis to a 'definitive' RCT.

In contrast, economic evaluation should be seen as an iterative process. Sculpher *et al* (forthcoming) argued that economic evaluation should typically consist of four stages. This would begin with Stage I analysis, when the technology is first used on patients in experimental centres. This form of analysis focuses largely on the costs and effectiveness of the standard intervention that the new technology may seek to replace, assesses the likelihood of the innovation proving cost-effective and hence worthy of further comparative evaluation, and considers the key variables that would need to be measured in any subsequent study. Stage II and III analyses re-visit the economic assessment of the developing technology as it is used more widely and as the volume and quality of patient-specific clinical data increase, to give a firmer estimate of cost-effectiveness.

Stage IV economic analysis takes as a starting point that the value for money of the new technology has been estimated on the basis Stage III studies, but considers whether the conclusions of that work are generalisable. Hence knowledge about the economic impact of health care technologies will take shape over time, and it is inappropriate to cease economic evaluation at Stage III, if the results of that analysis are likely to be sensitive to variation in parameters which differ by location and context.

Using the terminology of the four stages of economic evaluation, the analysis in this chapter can be described as Stage IV assessment. The methods proposed here to analyse the generalisability of a study would invariably use a decision analytic framework to build upon the firm evidential basis and high internal validity of a RCT, augmented with data taken from other sources which are likely to reflect variation in clinical practice. The approach explicitly recognises the trade-off inherent in clinical and economic evaluation between internal and external validity. The RCT typically maximises the former, but often falls down on the latter, and increasingly provides the pivotal evidence on effectiveness and resource use for Stage III analysis (in the context of this study, the base-case analysis in Chapter 5). Analysis of generalisability (or Stage IV analysis), seeks to assess the external validity of a study, but will often use observational data to achieve this.

A source of variability that particularly affects MAS technologies relates to the swift development of the procedures over time. Given that new RCTs are unlikely to be staged unless the changes to interventions are major, analysis of generalisability must be flexible enough to be able to explore the economic implications of technological developments shortly after they occur.

The extent of generalisability is one of four areas of uncertainty in economic evaluation [Briggs *et al*, 1994], and this chapter can be viewed as a further series of sensitivity analyses to those described in Chapter 5. The ultimate aim of analysis of generalisability is to assess the robustness of the conclusions coming out of the base-case analysis. In Chapter 5, two illustrative cost per QALY thresholds were defined: the lower threshold (£6,500) was assumed to represent the ratio below which most technologies would be considered cost-effective; the higher threshold (£33,000) was taken as the ratio above which few technologies would be considered cost-effective. Against these thresholds, on the basis of base-case parameters, AH would probably be considered cost-effective relative to TCRE. However, Chapter 5 showed that, although this conclusion was robust to variation in individual parameters, this did not apply to

simultaneous extreme variation. The aim of this chapter is to assess whether the base-case conclusions are robust to uncertainty related to generalisability.

6.9.2 Alternative RCT results

The first analysis of generalisability considers the relative cost-effectiveness of TCRE and AH based on two alternative sets of trial results. As discussed in Section 6.3, for analysis of generalisability which explores the importance of variation in clinical practice, there is value in keeping the data from the two trials separate rather than undertaking a meta-analysis which risks masking the variations between the studies.

The major source of variation between these trials and the Bristol RCT is the rate of re-treatment of women randomised to TCRE. The alternative trials reported between a 0% and 16% hysterectomy rate in these women at one year follow-up, which can be extrapolated to 4% and 20%, respectively, at two years, compared to 12% in the base-case analysis. This variation is not apparently due to any differences in case-mix, and is more likely to reflect differences in clinical practice and in patients' attitudes. The effect of the differences in the trial results is to generate cost per QALY ratios which span the base-case estimate of £1500 per additional QALY, with a range of £639 to £2475. If the two illustrative cost per QALY thresholds are acceptable to purchasers, the alternative trial would not seem to be sufficiently at odds with those in the base-case to alter the conclusion that AH is the more cost-effective treatment.

6.9.3 Routine clinical practice

The second analysis of generalisability moves away from the frequently atypical practice in clinical trials, and seeks to explore whether the process and outcomes of routine care result in different conclusions about the relative cost-effectiveness of TCRE and AH. The MISTLETOE survey offers a valuable source of data on the routine use of TCRE; however, it is a limitation of the analysis of generalisability that no comparable data are available for hysterectomy. The Vaginal, Abdominal or Laparoscopic Uterine Excision (VALUE) survey currently underway, will provide these data in due course.

In comparing the resource and non-resource consequences of TCRE using loop diathermy in the base-case analysis (the Bristol RCT) and that in the MISTLETOE survey, two points can be made. Firstly, the use of TCRE seems to have developed since the Bristol trial in 1990-1991. On average, length of hospital stay, time in theatre and complication rates all appear to be lower in the MISTLETOE survey. This may be partly due to the fact that case mix has changed as clinicians make judgments about which women are more likely to benefit from TCRE.

The second point to note is that the use of repeat TCRE was lower overall in MISTLETOE than in the Bristol RCT although hysterectomy rates are similar, and it may be the case that, given failure with an initial TCRE, clinicians are more likely now to advise women to undergo a hysterectomy rather than a repeat resection. Both of these findings affect the expected total cost of TCRE, with the sensitivity analysis showing a 21% reduction compared to the base-case estimates for TCRE. Comparing MISTLETOE-based cost estimates of TCRE with the base-case (trial-based) cost estimates for AH results in an increase in the incremental cost per additional QALY to £2167 from £1500 in the base-case, an increase which is unlikely to be crucial for decision makers. It remains to be seen how routine practice regarding AH influences the relative cost-effectiveness of the two treatments.

6.9.4 Resource sparing and intensive clinical practice

The third analysis of generalisability deals with variation in routine practice in relation to TCRE. Focusing on mean values and proportions tends to mask the significant range in process and outcomes across hospitals. The coverage and size of the MISTLETOE survey allows a by-hospital analysis which only the largest multi-centre RCTs can offer. The analysis is again limited by the fact that this detailed information on the routine use of TCRE is not yet mirrored by similar data for hysterectomy.

MISTLETOE shows significant between-hospital variation in the process and outcomes of TCRE: lengths of stay in hospital range from 0 (day-case) to 2 days

for patients without complications; overall complication rates range from 0% to 12.5%; and, at one year, hysterectomy rates range between 0% and 13.64%. Translated into expected costs, the analysis provides estimates at two-years ranging between £458, in the most 'resource sparing' hospitals, and £814, in the most 'resource intensive' hospitals. Although part of this range may reflect variation in case-mix, requiring the exercise of caution when interpreting these estimates, many of the women in the MISTLETOE survey would have undergone hysterectomy if conservative surgical methods had not been available, and the analysis suggests that, at some hospitals, TCRE is significantly less costly than the base-case costs of AH over two-years. On the other hand, resource sparing hospitals seem to be generating expected two year costs only sightly greater than the base-case, again re-enforcing the importance of apparent recent developments in the use of TCRE.

It is possible that, in due course, the VALUE survey will identify similar variations in the process and outcomes of hysterectomy. For example, VALUE might show a large proportion of women undergoing surgery at some centres continuing to consume health service resources for years after their hysterectomy. However, clinical opinion suggests that this is unlikely to be the case and that, for most women, hysterectomy is a 'once and for all' solution to menorrhagia. The RCT results show a smaller variation in per patient costs of AH compared to those for TCRE, and it is difficult to see what would drive a similar variation in the two-year costs of AH. Overall, the resource sparing and intensive analyses do not generate cost per QALY ratios higher than the lower illustrative threshold. If the illustrative thresholds are acceptable to purchasers, this would suggest that the base-case conclusions are likely to be robust to the variation in routine resource use associated with these procedures.

6.9.5 Unit costs

Analysis of Generalisability IV looks at how robust the base-case conclusions are to variation in the unit costs of key resources. The sensitivity analysis in Chapter 2 showed that the unit cost of a day in hospital had a marked influence on total costs. Recently, the quality of cost data in the NHS has improved and it

has been possible to compare the implications of using unit costs from three different hospitals using similar costing methods. These hospitals showed remarkable consistency in their costs: ward costs per day ranging from £83 to £104 and theatre overheads per minute ranging from £1.96 to £2.04. Compared to the base-case values, the hospital-specific ward costs per day show between a 13% and 68% reduction; and the theatre costs per minute show between a 81% and 89% increase. This has a larger downward impact on the expected two-year costs of AH than of TCRE, but its effect is not sufficient significantly to alter the cost per QALY ratios.

The three hospitals supplying the alternative unit costs may not be representative of UK hospitals undertaking TCRE and AH: all three are teaching hospitals based in large cities. As yet, however, reliable and similarly estimated unit cost data are not available in a large number of hospitals. In time, an important element of analysis of generalisability will be to study the robustness of base-case results to variation in unit costs taken from a large and representative sample of UK hospitals. As regards the AH versus TCRE comparison, on the basis of the data considered here, the base-case conclusions would seem to be robust to this source of variation.

6.9.6 Alternative types of surgical procedure

For many years, AH was the mainstay of surgical treatment for women with menorrhagia whose condition had not improved adequately on medical therapy. In the late 1980s, a range of conservative non-hysterectomy MAS options began to be used in the NHS, most notably TCRE. The rationale for this economic evaluation, as well as for the three published RCTs comparing TCRE and AH, is, therefore, clear. However, inevitably, the scenario of AH as the conventional surgical treatment and TCRE as 'the new technology' is over-simplistic, for several reasons. Firstly, the speed with which TCRE diffused in the UK meant that, by the time clinical evaluation using an RCT and economic analysis began, TCRE had become a widely used procedure and, in many centres, the first-line surgical treatment for menorrhagia. Secondly, TCRE was only one of several conservative surgical options being used: TCRE itself took several forms (loop or

rollerball diathermy, with or without uterine pre-thinning); laser and RF ablation were also used in some centres. Thirdly, although the abdominal route is used for the majority of hysterectomies, VH is used in some centres; recently, laparoscopic methods, usually to complement VH, have been developed in a limited number of hospitals.

It is an important component of analysis of generalisability to consider the economic implications of the variation in the way technologies are used in practice, as well as of the developments in procedures over time. The RCT upon which the base-case CUA is based was undertaken in 1990-91. Since then, TCRE has developed and new non-hysterectomy MAS procedures have diffused. Analysis of Generalisability V, therefore, uses sensitivity analysis to adjust the base-case parameters to estimate the costs and benefits of a range of surgical treatments for menorrhagia. This analysis moves some way from the firm evidential basis of the RCT underlying the base-case analysis.

The costs of the alternative forms of hysterectomy are estimated using assumptions based on published results and clinical opinion, and their benefits in terms of two-year QALYs are assumed to be equivalent to AH. The five non-hysterectomy MAS procedures, although assessed using data taken from a large survey collecting standardised information, were not necessarily used on homogenous groups of patients, and the full array of process differences and outcomes over which women may have preferences is not considered. This final analysis of generalisability should not, therefore, be seen as a full economic evaluation of all the important surgical treatments in menorrhagia; rather, it is a sensitivity analysis to the AH-TCRE comparison in the base-case analysis, and seeks to provide a broad-brush indication of whether variations in the two forms of surgery will substantively alter the conclusions of the base-case.

A number of conclusions are possible from Analysis V. Firstly, the additional costs of equipment and consumables with laser and RF ablation results in higher overall procedure costs relative to TCRE. Furthermore, on the basis of the MISTLETOE survey, there are no resource cost savings associated with these

ablative treatments (eg. complications, re-treatment) to offset these additional costs. Indeed, on the contrary, women treated by laser or RF in the MISTLETOE survey had higher re-treatment costs than women undergoing TCRE, although this finding must be tentative given possible differences in case-mix between the alternative types of procedure. Hence laser and RF are unlikely to tip the economic argument in favour of non-hysterectomy forms MAS; if anything, these modalities are less likely to represent cost-effective alternatives to AH.

The second conclusion form Analysis V is that the three approaches to diathermy delivery as part of TCRE (loop, rollerball or a combination), on the basis of MISTLETOE data, do not differ sufficiently in terms of two-year expected costs to affect the economic comparison markedly between TCRE and AH. The range of cost per QALY ratios between AH and the different forms of TCRE is relatively narrow (incremental cost of AH per additional QALY £1500 to £2176).

The third conclusion prompted by Analysis V is that the type of hysterectomy undertaken may alter the economic balance between hysterectomy and non-hysterectomy surgical procedures. The shorter length of stay in hospital generally associated with VH reduces the expected cost of this form of hysterectomy relative to AH. On the basis of the data presented here, this would suggest that VH is likely to make clearer the base-case conclusion that hysterectomy is more cost-effective than non-hysterectomy forms of MAS if the illustrative cost per QALY thresholds are acceptable: VH dominates RF and has an incremental cost per QALY gained of between £350 and £900 relative to the other non-hysterectomy forms of MAS. A similar conclusion is likely to apply if 'reusable' LAVH is used, although the expected two-year cost of the procedure is similar to that of AH. On the other hand, 'disposable' LAVH is likely to result in a higher cost of treatment than AH: the cost reduction due to a shorter length of stay is more than offset by the additional cost of equipment and/or consumables.

On the basis of data presented in Analysis V, 'disposable' LAVH is less likely than VH, 'reusable' LAVH and AH to be considered more cost-effective than

TCRE. However, all three forms of hysterectomy are likely to be considered cost-effective relative to the non-hysterectomy MAS options if the illustrative cost per QALY thresholds are acceptable. An important caveat is necessary here. The analysis assumes that the process and outcomes of VH and of LAVH would not be valued differently by women with menorrhagia. Although it is likely that any differences in health state values between the forms of hysterectomy are likely to be in the short-term (ie. period of convalescence) which has only a modest effect on expected QALYs, firm conclusions about the relative effectiveness of the procedures must await good comparative trials.

6.10 Conclusions

Figure 6.3 compares the differential cost and QALYs of non-hysterectomy forms of surgery and AH, with the lower dotted line representing the base-case estimate and the higher dotted line the lower of the two illustrative cost per QALY thresholds suggested in Chapter 5. The figure shows considerable variation in differential costs and benefits of the alternative estimates compared to the base-case estimates. If the lower illustrative cost per QALY threshold is acceptable to purchasers, the base-case conclusions can be considered robust to this variation, and AH would remain a more cost-effective treatment than TCRE (or the other non-hysterectomy forms of surgery).

The validity of this finding does depend crucially, however, on whether the illustrative ratios have any meaning to purchasers. If these decision makers are more concerned with minimising costs, their meaningful threshold will be the point where one of the two treatments saves money relative to the other. Against this threshold, TCRE would be the preferred intervention in the base-case, and this would be generalisable based on the analyses in this chapter.

This chapter has addressed one major source of uncertainty relating to the relative cost-effectiveness of AH and TCRE. The next chapter considers the validity of the QALY as a measure of benefit.

Difference in costs (AH minus TCRE)
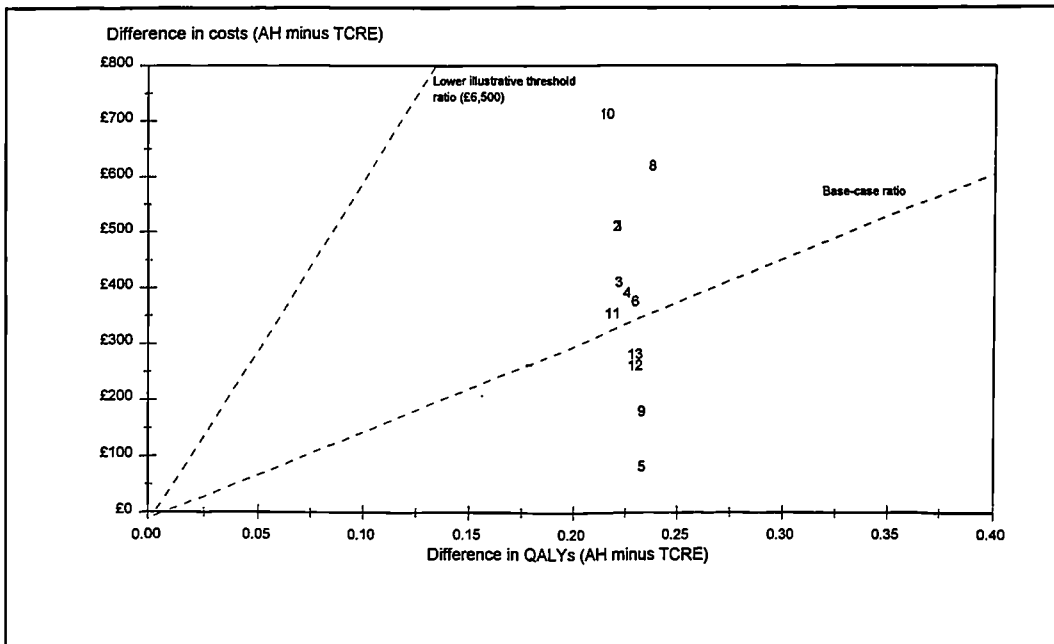
Figure 6.3   Graphical representation of the sensitivity analyses undertaken as part of
the analysis of generalisability using the cost-effectiveness plane. The
analyses all compare minimal access (non-hysterectomy) forms of
surgery with abdominal hysterectomy. The numbers relate to the
following comparisons: 1 = TCRE with loop (MISTLETOE) vs AH (base-
case); 2 = TCRE with loop and rollerball (MISTLETOE) vs AH (base-case);
3 = TCRE with rollerball (MISTLETOE) vs AH (base-case); 4 = laser
(MISTLETOE) vs AH (base-case); 5 = RF (MISTLETOE) vs AH (base-case);
*6 = TCRE (base-case) vs AH (base-case); 7 = TCRE vs AH (both Gannon
et al RCT)*; 8 = TCRE vs AH (both Pinion *et al* RCT); 9 = Resource sparing
TCRE (MISTLETOE) vs AH (base-case); 10 = Resource intensive TCRE
(MISTLETOE) vs AH (base-case); 11 = TCRE vs AH (both with Hospital 1
unit costs); 12 = TCRE vs AH (both with Hospital 2 unit costs);
13 = TCRE vs AH (both with Hospital 3 unit costs).

Appendix 6.1   **Costing guidelines given to the hospitals**

providing alternative unit cost estimates

(a) *Ward cost per day.* The hospitals were asked to include the following in their estimates:

(I)   nursing costs;

(ii)   ward disposables;

(iii)   ward overheads (eg. heating, laundry) apportioned on a reasonable basis;

(iv)   hospital overheads (eg. chief executive costs) apportioned on a reasonable basis; and

(v)   ward-related capital cost.

Each finance department was asked to estimate the annual cost of a gynaecological ward using these cost components and to calculate a cost per day using the annual throughput of the ward in terms of bed days.

(b) *Theatre cost per minute.* The following components of cost were to be included in the estimation of theatre costs:

(I)   theatre-related capital costs;

(ii)   theatre overheads (eg. heating) apportioned on a reasonable basis; and

(iii)   hospital overheads apportioned on a reasonable basis.

As described in Chapter 3, the remainder of the major theatre costs - for example, the cost of medical and nursing staff, anaesthetics, consumables and non-standard equipment required specifically for TCRE - were estimated separately, and are not subject to the same variation between hospitals. Again, the hospitals were asked to estimate an annual cost and to calculate the cost of a theatre minute by dividing the annual cost by a measure of throughput in patient minutes.

Appendix 6.2   **Details of the adjustments made to the unit**

costs of TCRE with the loop diathermy and AH

to estimate the unit costs of the alternative

procedures assessed as part of Analysis of

Generalisability V

The costs of the alternative procedures have been estimated by adjusting the

unit cost of TCRE with loop diathermy, for the non-hysterectomy forms of MAS,

and of AH, for the alternative forms of hysterectomy. As explained in Chapter 3,

the base-case cost analysis makes a distinction between standard equipment

that will be available in most theatres as a matter of course (eg. a diathermy

generator), and non-standard equipment which will not, because it is related

specifically to the new procedure being evaluated. In the base-case analysis, the

non-standard equipment, such as the telescope, camera and xenon light source

used as part of TCRE, were costed separately and added to the capital cost of

standard equipment which was included in the theatre overheads. For the

purposes of this analysis of generalisability, the distinction between standard and

non-standard equipment is maintained.

Table A6.1 details the assumptions that have been used to make the

adjustments to the base-case procedure unit costs. It is assumed that the fixed

cost (non-standard equipment and consumables) and the variable cost per minute

(staff, overheads and anaesthetics) of TCRE are the same whether the loop,

rollerball or a combination of diathermy methods is used. Similarly, it is assumed

that the fixed and variable costs per minute of VH are the same as for AH.

Table A6.1 shows that the main adjustments to the cost of a TCRE procedure, in

order to estimate the cost of laser and RF ablation procedures, are the changes

in equipment and consumables. The adjusted cost for a laser ablation procedure

involves the added cost of an Nd:YAG laser and laser fibres, but no cost of loops

is incurred. The adjusted cost of a RF ablation procedure requires the additional

costs of a generator, probe set and abdominal guard, but the endoscopic

Table A6.1   **Analysis of Generalisability V: adjustments made to the base-case unit costs of TCRE and AH procedures in order to estimate the unit costs of alternative surgical procedures**

Cost component	Non-hysterectomy surgery*		Alternative forms of hysterectomy[†]	
	Laser	RF	LAVH (Disposable)	LAVH (Reusable)
Staff	Addition of ODA to power laser	No change	No change	No change
Anaesthetics	No change	No change	No change	No change
Overheads	No change	No change	No change	No change
Consumables	Use of saline instead of glycine; use of re-usable laser fibres; no loop required	Use of probe set and abdominal belt; no loop irrigation tubing, catheter or glycine	Addition of primary trocar and canula; 2 more trocars; 5.5mm converters x 3; endograspers ; endosheers; endo-GIA; 3 extra staples; suction tubing	No change
Equipment	Use of Nd:YAG laser (costing £56,400, annual maintenance of £4,250; an expected useful life of 10 years and an estimated annual throughput of 250); no working element or sheath	Use of memostat generator and cable set (cost £47,000, annual maintenance of £4759, an expected useful life of 10 years and an estimated annual throughput of 250); no other specialised equipment	Use of laparoscope, telescope; camera; beam spitter; cable; xenon light source; monitor (all assumed to last 5 years with an estimated annual throughput of 1500)	Addition of trocars and portals x 4; bipolar forceps x 2; scissors x 2; graspers x 2; cannulae (all assumed to last 5 years with an estimated annual throughput of 250). Use of telescope laparoscope; camera; beam spitter; cable; xenon light source; monitor (all assumed to last 5 years with throughput of 1500)

*   All adjustments made are to TCRE with loop diathermy (as in base-case)
†   All adjustments made are to AH (as in base-case)

equipment used as part of TCRE is not required, as the procedure is undertaken 'blindly'. Apart from the cost of an ODA to power the laser, the variable costs per minute of staff, theatre overheads and anaesthetics are assumed to be the same for a laser and a RF ablation procedure as for a TCRE.

Adjusting the cost of AH to estimate the cost of a LAVH procedure involves the added costs of non-standard equipment and of consumables. On the basis of clinical advice, there appears to be considerable variation in the specific consumables and equipment used as part of LAVH. Perhaps the most important source of variation is in the choice of re-usable equipment rather than disposable consumables for specific parts of the procedure. In practice, hospitals are likely to use a mix of re-usable and disposable hardware. However, for the purposes of the current analysis, two indicative procedure costs are estimated: a 're-usable' LAVH and a 'disposable' LAVH. In order that these two procedure costs reflect clinical practice in terms of assumptions about hardware, two clinicians were interviewed to acquire details of the equipment and consumables used during a typical LAVH in their hospitals, one describing their hospital's policy as 're-usable', the other describing their hospital's practice as 'disposable'. Table A6.1 describes the equipment and consumables used as part of these two forms of LAVH, which are additional to AH. As for the base-case analysis, all equipment and consumables have been costed using manufacturers' list prices including VAT.

Chapter 7

A Cost-Utility Analysis of AH Versus TCRE Using An Alternative Measure of Benefit

7.1   Introduction

The existence of clear trade-offs between MAS and conventional open surgery in the process and outcomes of care, and the fact that patients are likely to have preferences over these trade-offs, indicates that CUA is likely to be the most appropriate framework within which to assess these two categories of technology. However, it is unclear whether CUA using standard QALYs can adequately reflect patients' preferences about the process and outcomes of care. In recent years, alternative benefit measures have been proposed for use in CUA which may more adequately reflect patients' preferences.

The purpose of this chapter is to review the literature on the theory of benefit measurement in CUA and to consider one of these alternative measures of benefit in the context of AH versus TCRE. This element of the thesis can be

viewed as a further assessment of the uncertainty associated with the base-case
analysis of the alternative surgical treatments for menorrhagia detailed in Chapter
5, this time focusing on the analytical uncertainty related to the measure of
benefit adopted.

Section 7.2 of this chapter reviews the literature which has recently developed in
this field. The remainder of the chapter reports the results of an exercise
undertaken to elicit one alternative benefit measure from a sample of women
with menorrhagia - *ex ante* healthy years equivalents (HYEs). The aim of the
exercise was to assess whether *ex ante* HYEs are consistent with individual
preferences, and whether their use alters the conclusions of the base-case
results of the CUA in Chapter 5. Section 7.3 details the methods used to elicit
these values, to assess their consistency with women's descriptive preferences
and to incorporate them into the CUA. Section 7.4 reports the results of the
analysis, Section 7.5 provides a discussion and Section 7.6 offers some
conclusions.

7.2   The theory of benefit measures in cost-utility analysis

7.2.1 The standard QALY model based on TTO values

An important aspect of the QALY has been its relationship to individuals'
preferences about the relative desirability of different health states [Drummond
et al, 1987], and hence of alternative technologies. To reflect individuals'
preferences it is necessary for treatments generating more QALYs to be
preferred by individuals' over those producing fewer QALYs. However, this link
between the standard QALY and preferences is based on some important
assumptions about the individual's utility function [Pliskin *et al*, 1980; Loomes
and McKenzie, 1989; Johannesson, 1995].

In Chapter 5, it was argued that evidence is now available indicating that, in
practical terms, the time trade-off (TTO) is preferable to the standard gamble
(SG) as a choice-based measure of the health state values necessary to estimate

QALYs. Much of this evidence has been generated by the York Measurement and Valuation of Health project [Williams, 1995], which is the largest health state valuation exercise undertaken in the UK. The TTO has probably been the most widely used of the choice-based instruments in the UK, based largely on its perceived practical advantages. In general terms, the standard TTO-based QALY is derived in the following way.

(a)   Plausible health profiles (prognoses), following from alternative forms of patient management, are decomposed into a series of discrete health states.

(b)   These health states are valued independently. Sometimes a distinction is made during the valuation process between temporary and chronic health states.

(c)   The period of time in a given health state is multiplied by its value. QALYs are calculated by summing these products over the duration of the patient's survival (or over the time horizon of the study). This process may involve discounting QALYs generated in future years.

(d)   Uncertainty may be incorporated into this process by calculating a series of QALY profiles and attaching a probability to each. Expected QALYs are calculated by multiplying each QALY profile by its relevant probability and summing across all the profiles.

The QALY estimates in Chapter 5 were derived in this way, as were the benefit estimates of most other CUAs using the TTO instrument [Gerard, 1992; Daly, 1993; Sculpher *et al*, 1996B; Cook *et al*, 1994]. For the standard QALY based on TTO values to represent individual preferences adequately, however, the following assumptions are necessary.

Risk neutrality with respect to life-years for all health states. The TTO instrument measures health state values under conditions of certainty - that is, the outcomes are known for sure. In practice, uncertainty exists in most areas of medical practice. In the context of the treatment of menorrhagia, women face

uncertainty in relation to such factors as the mortality associated with surgery and the need for re-treatment following TCRE. If uncertainty exists in the decision context, the only way the standard QALY derived using the TTO can reflect preferences is if individuals are risk neutral over life-years *for all health states* [Johannesson, 1995].

Several studies in the area of health care, however, have shown that patients do not exhibit this risk neutrality. McNeil *et al* [1978] interviewed 14 patients with operable lung cancer and explored their attitudes to treatments with different life expectancy probabilities. On the basis of a series of standard gamble exercises, which were used to elicit patients' certainty equivalents, the authors found the patients were generally highly risk averse.

Eraker and Sox [1981] looked at a series of hypothetical decisions individuals made related to drug therapy. Individuals were asked to choose between two drugs the outcomes of which were described in terms of life expectancy, one had a certain outcome and the other an uncertain outcome. Using a series of gambles, the investigators found that, even when the expected outcome detailed in the scenarios was the same for both drugs, the respondents chose the therapy with the certain outcome, indicating risk aversion. However, when the scenarios were framed differently, in terms of loss in health status rather than gain, the individuals were found to be risk loving, choosing the option with the uncertain losses. Either way, the study found no evidence of risk neutrality.

Individuals' preferences regarding future survival and health status exhibit constant proportional trade-off. This means that an individual is willing to sacrifice a *constant* proportion of their remaining period of survival to acquire a given improvement in health status, whatever the absolute number of life-years that remain. For example, a person who is indifferent between 20 years in their present health state and 10 years in perfect health would be assumed also to be indifferent between 10 years in their present health state and 5 years in perfect health.

Little evidence exists to support this assumption. McNeil *et al* [1981] investigated the attitudes of 37 healthy volunteers towards the trade-off between quality and quantity of life associated with laryngeal cancer. They found evidence to contradict the constant proportional trade-off assumption: although respondents accepted some trade-offs, they refused to trade quantity for quality when survival was down to five years.

As part of their theoretical analysis of the QALY concept, Pliskin *et al* [1980] used a questionnaire to explore how 10 academics traded-off survival and quality of life associated with angina. Various questions were asked to assess the minimum number of years of baseline survival respondents would sacrifice for a given improvement in quality of life. Out of 30 questions asked in the questionnaire, only nine answers supported an assumption of constant proportional trade-off. Furthermore, five of these nine answers indicated that the respondent was not willing to trade-off any survival time for an improvement in quality of life. When these respondents are removed, only four out of 25 answers were consistent with constant proportional trade-off.

Another way of conceptualising this assumption is that the value an individual attaches to a health state is independent of the time spent in that state. Although empirical studies have struggled to distinguish the separate effects of time preference and duration on health state values [Dolan and Gudex, 1995], this is again generally not supported by the evidence. On the basis of interviews with 246 members of the general public and 29 individuals undergoing home dialysis, Sackett and Torrance [1978] elicited TTO values for 10 health states, each of which respondents were to imagine lasting for between one and three time durations. They found a strong statistical relationship between value and duration: the mean value of each health state declined as duration in that state increased. The authors concluded that 'the duration of time that patients will spend in a specific health state must be considered when assessing the utility, as well as the cost, of health care programs' (p703).

A study by Sutherland *et al* [1982] also casts doubt on the assumption of health state values being independent of time. A sample of 20 physicians and scientists was asked to rate five health state scenarios using a simple preference statement and a certainty equivalence exercise. The authors found that, as the duration of time in some states is increased, the values of these states change dramatically. Their findings caused them to suggest the concept of 'maximal endurable time'; that is, a time period within a health state beyond which individuals radically change their attitude towards that state.

A more recent study also raises questions about the validity of this assumption. As a sub-study within the York Measurement and Valuation of Health Study, 234 members of the public were interviewed and presented with details of 15 health states based on the EuroQol classification [Dolan, forthcoming]. The respondents were asked to value each of the states using a visual analogue scale, imagining the states to last for three alternative durations: 10 years, one year and one month. The values individuals attached to dysfunctional health states were found to decrease as duration increased; that is, the states became increasingly intolerable as the time spent in them increased.

A study by Hall *et al* [1992] contradicts this evidence, however. On the basis of 104 interviews with healthy women and women with breast cancer, the authors used the TTO to explore the effect of life expectancy on the values women attached to life time health profiles associated with breast cancer. They found no association between the values and life-expectancy.

Individuals' valuations of a given health state are independent of the health states that precede or follow it. An important characteristic of the standard QALY model is that, to represent a given prognosis, a QALY estimate is based on a summation of health state values over time. This additive model can be expressed as in Equation 7.1 below:

$$\text{Standard QALY} = \sum_{t=1}^{T} u(q_t) \qquad\qquad (7.1)$$

where T is the time horizon of the analysis and $u(q_t)$ is the value associated with the health state q in time period t. This process of dividing a prognosis (or lifetime health profile) into a number of independently valued units is most obviously a key part of valuation systems such as the Quality of Well Being scale [Kaplan and Anderson, 1990], the Rosser matrix [Kind et al, 1982)], the EuroQol (EuroQol Group, 1990; Brooks, 1996) and the Health Utilities Index [Torrance et al, 1982; 1996]. However, virtually all CUAs using QALY are founded on this assumption. In the CUA in Chapter 5, for example, the estimated QALYs for TCRE are the sum of several health states over time.

The assumption of additive independence too must be doubted if the value an individual attaches to their current health state is affected by the sequence of health states they experience; that is, by what comes before and after the current health state. Related to this, an individual may attach a different value to a health state if they feel that their current health state will affect future health status. For example, the value attached to a health state involving severe pain after surgery, but which is followed by recovery, is likely to be quite different to the value associated with the same health state followed by death. Although additive independence may have little intuitive appeal, there is little evidence to support or to reject it. The Sutherland et al [1982] study referred to above does cast doubt on the assumption, as the concept of 'maximal endurable time' means that health states cannot be valued in isolation without considering health status in other periods [Bleichrodt, 1995].

A study by Richardson et al (1989 and 1996) casts doubt on both the time and sequence assumptions of the standard QALY. The authors constructed a series of scenarios related to breast cancer that referred to a relatively short time duration, and a health state profile which effectively linked the three and introduced a time element into a single scenario. Using the TTO, SG and a visual analogue scale with a sample of 63 women, the authors elicited values for each health state and for the profile. They then tested whether the construction of a standard (or composite) QALY, by multiplying the individual health state values

229

by their relevant duration and aggregating those products, produced a similar result to that implied by the answer to the TTO for the profile. They found that, using some typical discount rates, the composite or standard approach to QALY estimation produced quite different results to that for the profile (31% to 57% discrepancy). Furthermore, it was not possible to identify a plausible discount rate that would produce the same QALY estimates for both approaches. This result could indicate that duration and/or sequence affects health state values.

7.2.2 Alternative approaches to CUA I: the standard QALY model based on standard gamble values

If doubts are cast on the link between the standard TTO-based QALY employed in Chapter 5 and patients' preferences, alternative approaches to benefit estimation for CUA might be considered. One alternative is to use the SG technique to elicit health state values (or utilities in the case of the SG). The rationale for the use of the TTO in the base-case CUA was detailed in Chapter 5, and centred on the practical advantages of the TTO identified in the York Measurement and Valuation of Health study [Williams et al, 1995], as well as the fact that evidence indicates that the TTO and SG seem to generate similar values [Krabbe et al, 1996]. However, if the SG overcomes the likely difficulties of relating QALYs with individual preferences, its use may be more appropriate than the TTO.

The SG has been used widely in CUA, largely because it is seen as the technique with the strongest theoretical foundation, being based on the axioms of von Neumann-Morgenstern expected utility theory [Gafni, 1994]. Indeed, it would seem reasonable to retain the term 'standard QALY model' if the SG is merely substituted for the TTO.

However, there are a number of problems associated with the SG. It is true that its use to calculate QALYs will, in theory, enable part of the assumption of risk neutrality discussed in Section 6.2.1 to be removed from the link between that standard QALY and individual preferences. As the SG is based on von-Neumann-

230

Morgenstern axioms, which is a theory of decision making under uncertainty, it is frequently argued that the SG health state values reflect individuals' attitude to risk and uncertainty in a way that TTO values do not [Gafni *et al*, 1993]. However, although the SG, in theory, generates health state values that reflect individuals' risk attitudes, the life-years element of the standard QALY is not incorporated into the valuation process, and, in practice, risk and uncertainty clearly relates to survival as well as to health-related quality of life (HRQL). So the standard QALY based on the SG, at best, only partially reflects individuals' risk attitude. The standard QALY is a product of a value (or a utility) function where health status is the only argument, and an estimate of life-years is left in natural units.

Pliskin *et al* [1980] referred to this version of the standard QALY model as the risk neutral (RN) QALY. Whether the TTO or the SG is used to value health status, the life-years (quantity) element of the QALY will not reflect individuals' attitude to risk. Hence, the only way individual preferences can coincide with the RN QALY is if the individual is risk neutral with respect to life years. The evidence reviewed in Section 7.2.1 suggests risk neutrality as regards decisions in the health area is likely to be the exception rather than the rule. Furthermore, in order to reflect individual preferences, the SG-based QALY still requires the same assumptions as TTO-based QALYs. In addition to risk neutrality with respect to life-years, constant proportional trade-off must exist, and there must be independence between health state values and the sequence of health states.

Moreover, a crucial further assumption is required; namely, that the theoretical foundations of the standard gamble are valid. The strength of the SG is usually considered to be its strong links with economic theory, namely the axioms of von-Neumann-Morgenstern expected utility theory. However, this theoretical foundation has been criticised on several levels. At a theoretical level, it has been claimed that the SG does not, in fact, incorporate risk attitude and is theoretically equivalent to values elicited under conditions of certainty [Bouyssou and Vansnick, 1988]. A further criticism at a theoretical level was made by

Richardson [1994] who argued that, although risks are used as part of the SG technique, this is not adequate fully to elicit individuals' risk attitude. This is because the risk of instant death in the SG is unrealistic, with the real risks faced by patients relating to such things as surgical complications. Richardson further criticised the key outcome of the SG, namely the value of p; that is, the threshold probability of death which makes an individual indifferent between a gamble involving immediate death and perfect health and a certainty of a dysfunctional health state for the remainder of their life. Richardson argued:

> 'The value of p in the SG depends primarily upon the unpleasantness of the health state, S, which is described under conditions of certainty. In reality, S may occur in conjunction with very significant uncertainty or with negligible uncertainty. Yet the same SG is believed to capture the essence of both risk contexts. Clearly p cannot reflect real-world uncertainty when information about the nature and magnitude of this is not given to subjects' (p17).

Much attention has been given to exploring, empirically, the performance of the von Neumann-Morgenstern expected utility theory in explaining individuals' observed behaviour under conditions of uncertainty. In his extensive review of the theory and evidence relating to expected utility theory, Schoemaker [1982] considered four areas of empirical evidence on the theory: tests of the axioms; field studies of how individuals make decisions under conditions of uncertainty in the real world; individuals' ability to process information to facilitate 'rational' decision making under uncertainty; and the importance of context in decision making. Schoemaker concluded that, in certain specific situations, expected utility theory may predict behaviour well; for example, large corporations may use it as a matter of policy in some contexts. However, in general, the theory fails in three ways:

> 'First, people do not structure problems as holistically and comprehensively as EU (expected utility) theory suggests. Second they

do not process information, especially probabilities, according to the EU rule. Finally, EU theory, as an "as if" model, poorly predicts choice behaviour in laboratory situations. Hence, it is doubtful that the EU theory should or could serve as a general descriptive model' (p552).

Even if the axioms of von Neumann-Morgenstern expected utility theory have a poor predictive record in terms of individual behaviour under conditions of uncertainty, it may have a normative role: a theory about how rational decisions *should* be made. Richardson [1994] argued that, at this level too, expected utility theory is flawed. He argued that 'if an outcome is sufficiently unpleasant, it is not irrational to adopt a rule that avoids the outcomes or, perhaps, to adopt a rule that maximises the value of the worst possible outcome' (p12).

It can be argued that the limitations of von Neumann-Morgenstern expected utility theory result in the SG having no stronger a theoretical basis than the TTO. In which case, the practical problems of using the SG discussed in Chapter 5 may support the use of the TTO as the major valuation instrument for CUA.

7.2.3 Alternative approaches to CUA II: the risk-adjusted QALY

A major movement away from the standard QALY is represented by the risk-adjusted (RA) QALY, a concept introduced by Pliskin *et al* [1980] as part of a theoretical framework for the QALY. The rationale for the RA QALY is to overcome the problem with the standard QALY, as discussed in Sections 7.2.1 and 7.2.2, that it inadequately takes account of individuals' risk attitude. In the case of QALYs based on TTO health status values, risk attitude is captured in neither the health status nor the life-years elements of the calculation. In this case, the RA QALY is shown in Equation 7.2:

$$\text{RA QALY (TTO)} = [H(Q) \cdot T]^r \qquad\qquad (7.2)$$

In Equation 7.2 r is known as the risk aversion parameter. If the individual is risk neutral then r = 1 and equation 2 collapses to the standard QALY shown in

Equation 5.1 in Chapter 5. If $r < 1$ or $r > 1$, the individual is risk averse or risk seeking, respectively.

As health state values elicited using the SG are, theoretically at least, taken as reflecting individuals' attitude to risk, the risk aversion parameter applies only to life years, as shown in Equation 7.3:

$$RA\ QALY\ (SG)\ =\ H(Q) \cdot T^r \qquad\qquad (7.3)$$

The major assumption required to link the RA QALY with individual preferences is that individuals need to exhibit constant proportional risk posture over life years for all health states.

Empirically, r can be estimated using the method of certainty equivalence (McNeil *et al*, 1981), where individuals are asked to state the number of years of future life at which they would be indifferent between those years with certainty and a gamble involving a risk of immediate death and a counter risk of full survival. The empirical work hitherto undertaken on estimating the value of r to risk-adjust QALYs has been largely developmental. Pliskin *et al* [1980] applied their model to the evaluation of coronary artery bypass grafting. Using small numbers, they concluded that 'on the basis of these preliminary findings it appears that the mathematical form used to represent utilities gives internally consistent results and is suitable for further work' (p219).

Miyamoto and Eraker (1985) further developed Pliskin *et al*'s RA QALY model and, on the basis of certainty equivalent exercises with 46 individuals with coronary artery disease, the authors concluded '....that the model deserves consideration as a medical utility model, despite some preliminary evidence that assumptions of the model are descriptively false, because it provides a simple representation of the utility of survival duration and health quality' (p191).

One of the implications of the RA QALY is that TTO health state values can be adjusted, using the risk aversion parameter, to become equivalent to SG values; that is, $SG = TTO^r$. Shiell *et al* [1995] tested this particular aspect of the model on the basis of 119 interviews with women with early stage breast cancer. They found a correlation coefficient between the risk-adjusted TTO and the SG or 0.65-0.72, and concluded that it was not possible to reject the null hypothesis of equivalence between these two measures.

The RA QALY only addresses one limitation of the standard QALY model; that is, the failure to incorporate risk attitude into the life-years element (for SG-based QALYs) or into either element (for TTO-based QALYs). Hence the other major assumptions needed for QALYs to reflect individual preferences are still needed - in particular, constant proportional trade-off and independence of health state values with respect to time and sequence. Furthermore, the RA-QALY is rooted firmly in the axioms of von Neumann-Morgenstern expected utility theory which, as discussed in Section 7.2.2, has some major limitations.

7.2.4 Alternative approaches to CUA III: healthy-years equivalents based on the SG

A major recent development in CUA, at a theoretical level, has been the proposal of an alternative outcome measure: the healthy-years equivalent (HYE). The HYE was introduced by Mehrez and Gafni [1989] as a way of tying the outcome measure used in CUA more firmly to individual preferences. It sought to do this by avoiding the strongest assumptions of the standard QALY model; in particular, that of risk neutrality (*per se* with the TTO and with respect to future life-years with the SG), constant proportional trade-off and the independence of health state values from duration and from the sequence of states. The HYE retains the QALY's purpose of incorporating the impact of a technology on both the quantity and quality of life, and maintains the QALY's intuitive appeal for decision makers but, its originators claim, it is more firmly tied to utility theory [Mehrez and Gafni, 1989 and 1991].

The HYE can be defined as in Equation 7.4:

$$U(Q \cdot T) = U(Q^*, HYEs) \tag{7.4}$$

where U(·) is a utility function, T is a patient's future years, Q represents a level of health status (or HRQL) considered less desirable than full health and Q* is perfect health status. Hence the HYE can be defined as 'the hypothetical combination of years in a state of full health, which is equal, in terms of the individual preferences (utility), to the patient's current projected lifetime health profile' [Mehrez and Gafni, 1989, p143].

Mehrez and Gafni suggest that the HYE should be measured using a two-stage lottery based on the SG. They offer an algorithm for the measurement of a chronic health state, and for the case of many possible lifetime health profiles [Mehrez and Gafni, 1991]. In brief, the first part of the two-stage lottery is similar to a conventional SG and asks the respondent to select the probability (p*) that makes them indifferent between a gamble involving perfect health for the rest of their life and immediate death, versus the certainty of a less than perfect health state over a period of time. In the second stage of the lottery, the gamble is similar to that in the first stage, but p* is taken as a fixed probability of perfect health for a lifetime ((1-p*) being the risk of immediate death). The respondent is asked to indicate the number of years in perfect health with certainty with which they would be indifferent to the gamble.

Mehrez and Gafni argue that the HYE offers a means of avoiding the restrictive assumptions of the QALY. Furthermore, they argue that, because HYEs are elicited under conditions of uncertainty using the two-stage lottery process, HYEs will reflect individuals' attitudes to risk. An important further theoretical advantage of the HYE is that it makes unnecessary the rather arbitrary process of discounting benefits in economic evaluation using a small constant discount rate. Discounting of benefits is well established in economic evaluation despite evidence that the conventional exponential discount model does not adequately

describe individuals' behaviour [Loewenstein and Prelec, 1993; Redelmeier and Heller, 1993; Dolan and Gudex, 1995]. Indeed, it has been argued that it is simply not possible to measure empirically individuals' pure time preference, because it is not possible to divorce it from other preferences, such as that over sequences of events [Gafni, 1995].

In eliciting HYE responses, individuals are presented with time profile information, so their intertemporal preferences are incorporated into their values. Hence, there is no need separately to adjust benefit measures using a discount rate. Not only does this avoid the need to estimate pure time preference rates by isolating this form of preference from others, but variation between individuals in their intertemporal preferences can be registered directly in the economic evaluation.

The HYE offers, at the theoretical level at least, an extra degree of flexibility over the standard QALY. However, the HYE has been associated with considerable controversy, focusing in particular on the two-stage lottery measurement technique. Several commentators have argued that, in using the two-stage lottery, the HYE is effectively no different to the TTO [Buckingham, 1993; Johannesson *et al*, 1993; Johannesson, 1994; Culyer and Wagstaff, 1993 and 1995; Loomes, 1995]. In brief, this argument is based on the fact that the gamble elements of both stages of the lottery are identical and cancel each other out. What is left is the result of a TTO question: indifference between a period of time in less than perfect health and a shorter period of time in full health.

Related to this point, it has also been suggested that the HYE does not reflect individuals' attitude to risk. Johannesson *et al* [1993] argued that the equal and opposite effects in the two-stage lottery result in risk attitude being eliminated from HYE values. They argue: 'the net result, combining the two stages, would be the same for both the risk-averse and risk-neutral individual, because the final comparison is made under certainty in the form of a time trade-off' (p284). Mehrez and Gafni [1993] refute the claim that the HYE elicited using the two-

stage lottery is merely a TTO. They argue that the SG generates utilities and the TTO produces values and that these need not necessarily be equal. Shiell *et al* [1995] set out to assess empirically whether HYEs are equal to the values implied by the results of TTO exercises. On the basis of interviews with 119 women, the authors concluded that the two measures did not produce equivalent results.

Although it can be argued that the HYE offers a theoretical advance over the standard QALY, it does so at the cost of increasing the burden of the valuation tasks appreciably. The increased use of CUA as a tool of economic evaluation has been facilitated partly by the development of health state valuation systems. The advantage of this has been that patients' lifetime prognoses can be divided into a series of health states each of which is, as closely as possible, related to a set health state in a valuation system. A standard QALY is, therefore, calculated by weighting a patients' time in each health state by the appropriate value, and aggregating across their lifetime (or the relevant time horizon of the evaluation). This process is ideal for CUA based on decision analytic models like the decision tree or the Markov model, of which the 'segmentation' of prognoses into separate health states is a fundamental part [Weinstein *et al*, 1980; Sonnenberg and Beck, 1993].

Of course, this approach to CUA produces the standard QALY that the HYE is trying to improve upon, and rests crucially on the assumptions of constant proportional trade-off and the independence of value from duration and sequence. By avoiding these assumptions, the HYE looses the flexibility of the standard QALY, particularly in decision analytic models. In order to use the HYE within a CUA of a technology which involves a large number of possible lifetime pathways (or profiles), in principle, each pathway needs to be translated into a HYE using the two-stage lottery. How the HYE approach could be incorporated into a CUA Markov model is not clear, as the model itself determines the lifetime profiles.
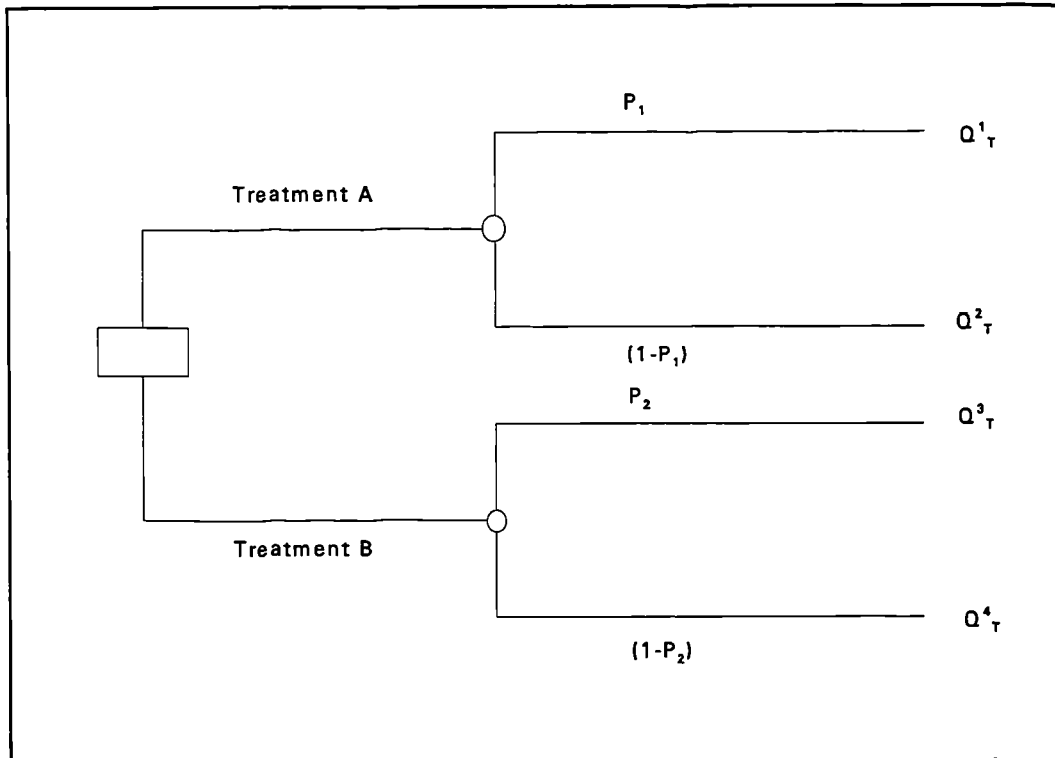
Figure 7.1   A hypothetical decision tree illustrating the choice between two
treatments (source: Gafni *et al*, [1995])

Gafni *et al* (1995) have responded to the suggested valuation burden and

inflexibility of the HYE, and to the argument that it fails to reflect individuals' risk

attitude. They have developed a two-stage lottery for use with decision trees,

and extended the HYE concept by developing, what Johannesson [1995] has

termed, the *ex ante* HYE. Figure 7.1 shows a hypothetical decision tree

illustrating a choice between two treatments, A and B. Each pathway has a

probability (P_j) and a lifetime health profile Q^i_T. The *ex ante* HYE is estimated as

follows. Firstly, a conventional SG is used to find the utility $[U(Q^i_T)]$ of each

pathway in the tree. Secondly, the expected utility of the two treatments is

calculated by, for each treatment, multiplying the utility of each arm with its

probability and summing the two arms. The expected utility of each treatment

will be a number between 0 and 1. Finally, the expected utility (EU) is used as a

probability in a lottery. Taking treatment A as an example, the respondent is

asked to compare a gamble involving a chance of a lifetime in perfect health with

a probability of EU_A and a chance of immediate death with probability $(1-EU_A)$.

239

The individual is asked to select a period of time in perfect health with certainty that would make them indifferent between that and the gamble. The outcome of this is an *ex ante* HYE for the two treatments.

As well as indicating how HYEs could be used with a decision tree model, Gafni *et al* [1995] were responding to Johannesson *et al*'s [1993] charge that HYEs fail to incorporate individuals' attitude to risk. This is because the final part of the above estimation procedure involves asking respondents to select a certainty equivalent period of time they consider equivalent to a risky prospect. Of course, the link between *ex ante* HYEs and individuals' risk attitude depends crucially on the validity of the axioms of expected utility theory.

In order to use HYEs with decision trees, however, the valuation burden is a major consideration. It remains the case that each feasible outcome profile needs to be valued separately. Gafni *et al* [1995] argue that the use of HYEs reduces the number of valuation tasks required because, by using profiles rather than a series of separate health states, fewer values are required. The disadvantage of this, though, is that the health profile descriptions are likely to be quite complex. Furthermore, HYE values have to be estimated for each study undertaken, and there is no apparent role for standardised and previously-valued health states as provided by the valuation systems for estimating standard QALYs.

7.2.5 Alternative approaches to CUA IV: healthy years equivalents based on the TTO

The concept of the HYE is clearly very close to that of the TTO; that is, a period of time in perfect health considered equivalent to a longer period of time in a dysfunctional health state. Although the SG has been the focus for discussion about the measurement of HYEs, the TTO can be used directly for this purpose.
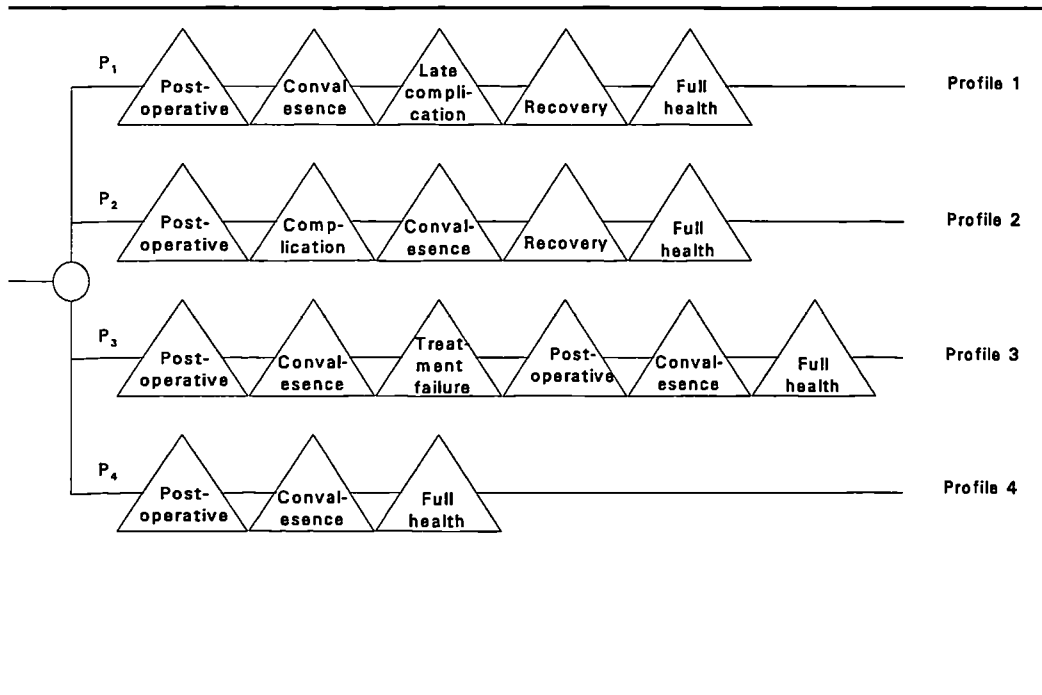
'igure 7.2   A hypothetical probability tree showing four health state profiles
              involving a series of transitory health states

The conventional way to use the TTO as part of CUA is, as described in the
context of AH and TCRE in Chapter 5, to estimate a value on a 0-1 scale for
each relevant health state, and then to use those health states in the standard
QALY model (ie. multiplying the value of a health state by a patient's duration in
that state). However, the TTO can be used to value health state *profiles* in
terms of HYEs. As for SG-based HYEs, TTO-based HYEs can be developed in an
ex post or *ex ante* form.

To describe the *ex post* HYE, consider Figure 7.2. The figure shows a
hypothetical probability tree relating to some form of surgery. The patient can
pass through one of four pathways (or profiles) with a given probability, where
each profile is made up of a number of transitory health states relating to such
things as complications and treatment failure. To represent the probability tree
in terms of the standard QALY model, each of the transitory health states would
be valued so that, for example, the post-operative health state received the same
value regardless of its duration or its position in the sequence. Each profile

would then be given a QALY score by multiplying the value of a health state by its duration, and then summing across the profile. With *ex post* HYEs, however, raters would be presented with four descriptive scenarios, one for each profile. These scenarios would detail each transitory health state and their respective durations, and the rater would be asked to state a period in full health that they would consider equivalent for each of the four profiles (ie. they would provide four HYE estimates).

The only published CUA based on the TTO-based *ex post* HYE that has been identified in the literature was undertaken by Hall *et al* [1992]. As part of an economic evaluation of mammography screening, 104 women were presented with a range of breast cancer-related scenarios. The scenarios described health state profiles which differed according to type of surgery, physical health and mental health. *Ex post* HYEs were derived for each profile using the TTO. As the 95% confidence intervals around the mean values overlapped for some of the profiles, the six profiles were divided into two broadly similar groups which were termed 'good health' and 'poor health', and mean values were allocated to each. These two general profiles were then used in a CUA model. The expected HYEs resulting from screening and no screening were calculated by multiplying the HYEs for each profile by the probability of a given woman following that profile. Expected HYEs were then related to expected costs in the form of a cost-utility ratio.

The *ex post* HYE based on the TTO exercise has the strength that it avoids the strong assumptions of constant proportional trade-off and of values being independent of time and sequence, which are necessary with the standard QALY. However, the HYE values will not incorporate the risk attitude of raters. The TTO exercise follows the usual approach of being undertaken under conditions of certainty. The incorporation of risk, in terms of an expected HYE, is undertaken outside the valuation exercise and, unless individuals are risk neutral with respect to life-years, the *ex post* HYE is unlikely to equate with the HYE chosen if the raters were aware of the risks involved.

242

To overcome this problem, the TTO can be used to elicit *ex ante* HYEs which reflect individuals' attitude to risk. This involves presenting raters with descriptions, not only of all the possible lifetime profiles associated with an intervention, but also of the probabilities associated with those profiles. With reference to Figure 7.2, the TTO-based *ex post* HYE would require four separate TTO exercises, one for each profile, with information on probabilities not provided. The TTO-based *ex ante* HYE, however, would require each of the four profiles to be detailed in one scenario, together with their respective probabilities. This process is similar to the conventional certainty equivalent exercise used in earlier studies [Pliskin *et al*, 1980; McNeil *et al*, 1981], except the range of possible outcomes presented to the rater could, in principle, be very large.

The advantage of the TTO-based *ex ante* HYE is that it requires few of the assumptions associated with the standard QALY and the other alternative outcome measures. Like other forms of the HYE, it avoids assumptions about constant proportional trade-off, and sequence and duration independence. Because risks are incorporated directly into the descriptive scenarios, raters' responses should reflect their attitude to risk and, unlike the SG, in a way that relates directly to the intervention in question. Furthermore, the TTO-based *ex ante* HYE does not require that the axioms of expected utility theory are theoretically, descriptively or normatively valid, as do HYEs based on the SG. Another advantage of the TTO-based *ex ante* HYE is that is asks directly the question which is at the heart of the QALY and of the HYE: what period in good health is considered equivalent to a longer period in dysfunctional health. Hence it is possible to get at this measure directly without a series of intervening gambles. The evidence reviewed in Chapter 5, that suggests raters find the TTO an easier instrument to use than the SG, is another advantage of the TTO-based *ex ante* HYE.

The major disadvantage of the HYEs is that the descriptive scenarios used to generate them have to include a large amount of information on alternative levels

of health status over time. The trade-off that exists between the number of
scenarios (and hence valuation exercises) and the detail in those scenarios is at
its most stark with TTO-based *ex ante* HYEs which also need to include a range
of probability data. As Johannesson [1995] and Gafni *et al* [1995] comment, it
is an empirical issue whether individuals can process the sort of information
contained in these scenarios. It is worth noting, however, that the descriptions
necessary for TTO-based *ex ante* HYEs would have a high level of informational
content for patients and are what many patients would expect their clinician to
provide them with anyway - that is, a reasonably detailed, but comprehensible,
list of the risks, benefits and long-term prognosis of a given intervention. Indeed,
in undertaking the TTO exercise, it may be possible to draw on existing
information sources for patients such as leaflets and possibly videos. This
characteristic of TTO-based *ex ante* HYE scenarios is not shared by any of the
other benefit measures suitable for CUA, which divide up a patients possible
prognosis in some way. Furthermore, the information content of TTO-based *ex
ante* HYE scenarios could mean that they make more sense to raters, especially
if they are patients who would have thought about many of the risks and
benefits prior to the exercise [Wakker, 1996]. It may be the case, therefore,
that this type of *ex ante* HYE may actually be easier to elicit.

The closest thing to a CUA that has used a TTO-based *ex ante* HYE is that
undertaken by Cook *et al* [1993, 1994]. The context was an economic
evaluation of three alternative treatments for gallstone disease: open and
laparoscopic cholecystectomy and extracorporeal shockwave lithotripsy (ESWL).
Using information from interviews with patients and patient questionnaires, a
series of health state scenarios was developed. These scenarios took a 'partial
ex ante' (p158) perspective in that the treatment-related scenarios included
information on the risk of operative mortality. On the basis of interviews with
96 members of the general public, TTO exercises were undertaken to value each
scenario. The authors then compared the loss of QALYs per 100 patients
associated with the alternative treatments and their aftermaths using an *ex post*
and partial *ex ante* approach. The *ex post* approach involved adding the QALY

Benefit measure	Restrictive assumptions					Theoretical underpinning with VN-M	Valuation burden
	CPT	DI	SI	RNLY	D		
TTO-based standard QALY	✓	✓	✓	✓	✓	No	*
SG-based standard QALY	✓	✓	✓	✓	✓	Partly	**
Risk-adjusted QALY	✓	✓	✓	X	✓	Yes	***
SG-based ex post HYE	X	X	X	✓	X	Partly	****
SG-based ex ante HYE	X	X	X	X	X	Yes	****
TTO-based ex post HYE	X	X	X	✓	X	No	****
TTO-based ex ante HYE	X	X	X	X	X	No	****

CPT = Constant proportional trade-off
DI = Health state values independent of duration
SI = Health state values independent of sequence
RNLY = Risk neutral with respect to future life-years
D = Time preference represented by the exponential discounting function
VN-M = Axioms of von Neumann-Morgenstern expected utility theory

Valuation burden: * = modest
                   **** = major

loss associated with the procedure, the risk of complication, the risk of death and HRQL over 18 months. The partial *ex ante* approach involved adding each of these except the QALY loss associated with operative death, which was incorporated into the scenario. The authors found large differences between the two perspectives with the two treatments. The authors justified their partial *ex ante* perspective rather than a complete one because 'the inability of an individual to process large amounts of information in a reliable and valid way makes such an analysis (full *ex ante*) difficult' (p158).

7.2.6 Summarising alternative approaches to CUA

A range of alternative benefit measures has, therefore, been developed in recent years for use in CUA. Table 7.1 summaries each alternative measure of benefit, indicating the restrictive assumptions necessary to link it to individual preferences, whether the measure is based on the von Neumann-Morgenstern axioms and the valuation burden imposed. The table clearly shows that a trade-off appears to present itself in selecting outcome measures for CUA. The standard QALY, whether based on SG or TTO values, requires some strong assumptions if individual preferences are to be reflected in the analysis. As alternative approaches have developed in an attempt to avoid some or all of these assumptions, a progressively greater measurement and valuation burden has apparently been imposed.

7.3   Methods

7.3.1 Introduction

The availability of a range of alternative benefit measures for use in CUA introduces another area of analytical uncertainty into studies: which measure is the most appropriate in a given context. This uncertainty is particularly pronounced in the assessment of MAS interventions, where patients' preferences about outcome trade-offs are likely to be considered important in decisions about resource allocation.

The remainder of this chapter considers the importance of this analytical uncertainty in the context of the economic evaluation of AH versus TCRE, by focusing on one of the alternative benefit measures reviewed above, the *ex ante* HYE. The starting point of this empirical work is that, in principle based on the foregoing review, the *ex ante* HYE based on the TTO is most likely to be able to reflect women's attitudes to the risks and trade-offs associated with the process and outcomes of surgical treatment for menorrhagia. This is based on the argument that this approach to the measurement of HYEs is more direct and intuitive than the two-stage lottery, benefits from the practical advantages of the TTO over the SG discussed in Chapter 5 and can reflect patients' attitudes to risks without requiring the axioms of von Neumann-Morgenstern to be valid.

The following specific questions are addressed: Is it practical to use *ex ante* HYEs to estimate the benefits of the two surgical treatments? How consistent are *ex ante* HYEs with individuals' descriptive preferences? Are the conclusions of the base-case analysis in Chapter 5 robust to the use of *ex ante* HYEs as the measure of benefit?

7.3.2 Valuation exercise

Given the importance of not overburdening women with too many valuation tasks, it was not considered appropriate to use the same sample of women to obtain *ex ante* HYEs as was used to elicit health state values for the standard QALY. Therefore, a further sample of women was identified in a second centre. In order to identify a sample of women with very similar characteristics to the Bristol sample used for the standard QALY valuation exercise, the same process was used to recruit women into the study. All women who had recently been referred by their GP to the Princess Margaret Hospital in Swindon, for apparently uncomplicated menorrhagia, were the population from which the HYE valuation sample was drawn.

Potentially eligible women were identified from GP referral letters received by the hospital between August 1994 and March 1995. These women were sent a letter explaining the study and asking if they were willing to be interviewed by a trained female interviewer prior to their visit to the out-patient clinic. Women were also asked to complete the same questionnaire as the Bristol sample, focusing on their health status and preferences for treatments and the characteristics of treatment, full details of which are provided in Chapter 4. If women responded positively to the invitation, they were contacted by telephone to arrange a convenient date and time to be interviewed, which would take place in their homes unless they preferred to come to the hospital. Women were excluded from the valuation exercise if, on the information they provided in the questionnaire, they had significant concomitant illness; if they lived too great a distance from the Princess Margaret Hospital to make an interview practicable; or if interview prior to their hospital appointment was not feasible.

As for the Bristol valuation exercise, a target sample of 60 women was established. Interviews were undertaken by a trained female researcher. The interview schedule used in the valuation exercise consisted of two elements.

Introduction. During the introduction section of the interview, the researcher introduced herself and the nature of the exercise; she also asked for permission to tape the interview. Finally, a short additional questionnaire was given to the woman to complete, asking a series of socio-demographic questions.

Valuation. The second stage of the interview involved eliciting from women values for health profiles in terms of *ex ante* HYEs. Two *ex ante* profiles were presented to women, and included a clear time dimension running from initial surgery until the menopause, and then until death. These profiles included estimates of the risks associated with therapy: the risk of operative death (for both treatments) based on estimates for AH which were assumed to apply to

both treatments [Dicker *et al*, 1982], and of re-treatment (for TCRE) based on the results of the Bristol trial (see Chapter 3). The two scenarios were based partly on the synthesis of the health state descriptions used in the Bristol valuation exercise, and partly on additional information from the same sources used to develop the original descriptions. Appendix 7.1 reproduces the two profiles used to elicit *ex ante* HYEs.

As the profiles had a time dimension lasting until death, they had to be 'customised' for each woman's life expectancy. As for the Bristol valuation exercise, life expectancy was assumed to be 60 years (for women aged between 20 and 29 years); 50 years (for those aged between 30 and 39 years); 40 years for those aged between 40 and 49 years; and 30 years (for those aged between 50 and 59 years). The profiles assumed the menopause would occur with approximately 35 years of life remaining.

On being presented with the profiles, the women were asked to rank them. The *ex ante* HYEs for the two profiles were then elicited using the TTO. As in the Bristol study, the 'converging ping-pong' method was used to avoid anchoring bias [Mohide *et al*, 1988]. The process of eliciting *ex ante* HYEs using the TTO is the same as that used to elicit health state values to estimate QALYs, the only difference being that the period of time in imperfect health considered comparable to a lifetime profile is itself the HYE estimate, and is not then transformed to a value on a 0 to 1 scale.

7.3.3 Assessing the consistency of ex ante HYEs with women's stated preferences

One of the characteristics of the *ex ante* HYE is that it can be associated with a specific treatment without first having to be combined with other health state values and probability data. In other words, the *ex ante* HYE is a single and all-embracing treatment-related preference measure. As a result of this, *ex ante*

HYEs can be compared with other indications of treatment preference provided by individuals (in this case patients), as an assessment of consistency.

As described in Chapter 4, the first part of Section 4 of the questionnaire completed by women in the interview samples listed a series of 10 characteristics of surgical treatment for menorrhagia, and women were asked to rate the importance of these characteristics on a 4-point scale from 'very important' to 'not important'. The second part of Section 4 described two options for treatment, one representing AH and one TCRE. Women were asked to indicate which, if any, they would prefer, and to rate each of them on a visual analogue scale. In Section 5 of the questionnaire women were asked to indicate whether they had strong preferences for or against treatments and, if so, to name them.

Women's responses to Sections 4 and 5 of the questionnaire have been used to assess the consistency of the *ex ante* HYE estimates with the more descriptive treatment preferences, at both the level of the individual and of the group. The following analyses have been undertaken.

     Ex ante HYEs and characteristic groups. Based on women's responses to the questions about the importance of the various characteristics of treatment for menorrhagia, two 'characteristic groups' are defined. If women considered the characteristics typical of TCRE as important, they are placed in the TCRE characteristic group; if they felt the characteristics typical of hysterectomy were important, they are put into the hysterectomy characteristic group.

Women are put into the TCRE group if they indicated that all of the following characteristics of treatment were 'very important' or 'of some importance': treatment that will *not* remove the womb; treatment causing the least pain and discomfort during convalescence; treatment that will reduce periods but not stop them for good; treatment that will result in getting back to usual activities as

soon as possible after the operation. Women are placed in the hysterectomy characteristic group if they felt that both of the following were 'very important' or 'of some importance': treatment that *will* remove the womb; and treatment that will stop periods for good. It would be reasonable to expect that women's *ex ante* HYEs would be greater for a treatment if they are allocated to its characteristic group.

Ex ante HYEs and choices about treatment options. Women's responses to the choice between the two treatment options described in the questionnaire are compared to their *ex ante* HYEs. Consistency would require that *ex ante* HYEs for a given treatment would be higher for women who indicated that they would select the option describing that treatment in the questionnaire.

Ex ante HYEs and visual analogue scores for treatment options. The visual analogue scores women provided for the treatment options described in the questionnaire are compared to their *ex ante* HYEs. It would be expected that the higher the visual analogue score for a treatment, the higher would be the *ex ante* HYE elicited in the interview.

Ex ante HYEs and stated treatment preferences. Women's responses to questions about positive and negative preferences for actual treatments are also compared to *ex ante* HYE values. Due to small numbers, both positive and negative preferences are grouped as being for hysterectomy or for other treatments. Although much depends on women's prior information about the characteristics of treatments, it would be reasonable to expect *ex ante* HYEs to be higher (lower) when a woman stated a strong positive (negative) preference for that treatment.

7.3.4 CUA of AH versus TCRE using ex ante HYEs

In order to assess whether a CUA using *ex ante* HYEs produces different conclusions to that using the standard QALY, much of the original CUA detailed

in Chapter 5 is retained here. In particular, the cost side of the original CUA is unaltered, so the focus is on substituting the estimates of *ex ante* HYEs for TCRE and AH for the QALYs in the original model. The *ex ante* HYE descriptive profiles which were presented to the women were made consistent with the parameters used in the original CUA. For example, the probabilities of death and of re-treatment were the same in the profiles as in the standard CUA model.

The fact that *ex ante* HYEs relate to lifetime profiles with risks has two implications for CUA. Firstly, the base-case results of the standard QALY-based CUA in Chapter 5 are taken over a time horizon of two years and hence are not the appropriate ones against which to compare the lifetime *ex ante* HYE-based CUA. The sensitivity analysis reported in Chapter 5 extrapolated the base-case results over women's lifetimes using some assumptions, and it is these lifetime results against which the *ex ante* HYE-based CUA results are compared.

The second implication of the fact that *ex ante* HYEs relate to lifetime health profiles is that, because the time dimension is fixed, the *ex ante* HYEs will reflect women's rate of time preference. Therefore, no allowance for time preference, using the conventional discounting formula, is necessary with this benefit measure.

7.4   Results

7.4.1 The sample of women interviewed

As part of the valuation exercise, 202 women were identified from referral letters sent to the Princess Margaret Hospital in Swindon, of whom 121 returned completed interview consent forms. Of these women, 107 agreed to be interviewed. A total of 63 women were eventually interviewed, thus just exceeding the target sample of 60 women. In reaching this number, 44 women were excluded for reasons detailed in Table 7.2. The fact that 27 (62%) exclusions were because either the woman's out-patient appointment was too

Table 7.2   Details of reasons for exclusions from the HYE valuation exercise

Number	%	Reason for exclusion
14	32	Unable to find acceptable date and time for interview
13	30	Date of out-patient appointment too close or passed
6	14	Concomitant illness
5	11	Unable to make contact with woman
3	7	Failed to attend interview
1	2	Interviewee unwell on day of interview, unable to arrange alternative date
1	2	Previous uterine surgery
1	2	Inter-menstrual bleeding

close or had passed, or a convenient time and date could not be found before that appointment, was largely due to the fact that a waiting list initiative was underway at the Princess Margaret during this period. This resulted in a shorter than usual period between referral and the out-patient appointment for most women.

Table 7.3 presents the socio-demographic and clinical characteristics of the sample of women interviewed. The table compares the characteristics of the Swindon sample with those of the women in the Bristol valuation study which provided the health state values for the standard QALY analysis. The table shows that, for most characteristics, the two interview groups were very similar, with no statistically significant differences between them. The one exception to this is the duration of menorrhagia. As detailed in Chapter 4 in relation to the larger sample of women with menorrhagia who returned questionnaires in the two centres, women in the Bristol sample reported that they had suffered their menstrual problems for a longer duration than women in the Swindon sample (median 24 versus 12 months, p = 0.03).

Table 7.3   Socio-demographic and clinical details of women in the Swindon and Bristol valuation
                exercises

Characteristic	Swindon (n = 63)	Bristol (n = 60)	P value
Socio-demographic			
Mean (SE) age (years)	41.00 (0.92)	41.09 (0.77)	0.94*
Number (%) who have experienced serious illness:			
- themselves	15 (24)	18 (33)	0.28†
- in their family	41 (65)	33 (63)	0.86†
- in caring for others	20 (32)	14 (30)	0.83†
Numbers (%) smoking:			
- currently	17 (27)	19 (32)	0.47†
- formally	15 (24)	18 (30)	
- never	31 (49)	23 (38)	
Number (%) who have worked in health or social services	17 (27)	22 (37)	0.22†
Number (%) in employment	41 (65)	33 (55)	0.25†
Number (%) leaving school at minimum leaving age	30 (48)	21 (35)	0.16†
Number (%) with degree or equivalent professional qualification	8 (13)	10 (17)	0.53†
Clinical			
Median (range) duration of menorrhagia (months)	12 (2-420)	24 (3-360)	0.03*
Median (range) days per month bleeding	8 (3-31)	8 (4-20)	0.22*
Median (range) days per month with heavy flow	4 (1-25)	4 (2-14)	0.49*
Number (%) passing clots	54 (92)	51 (86)	0.38†
Number (%) with flooding episodes	53 (91)	58 (97)	0.27ø
Maximum number of pads/tampons on heaviest day of period (numbers (%)):			
- 1-9	17 (29)	18 (31)	0.84†
- More than 9	41 (71)	40 (69)	
Median (range) days lost from work due to menstrual problems over last year for those in work††	4 (0-36)	2 (0-48)	0.87*

*	Wilcoxon rank-sum test
†	Chi square test
ø	Fisher's exact test
††	n = 35 (Swindon) and n = 37 (Bristol)

Table 7.4   Perceived health status of women in the Swindon and Bristol valuation studies on the basis of
                   the EuroQol instrument

	Swindon (n = 63)		Bristol (n = 60)	
	Today	'Heaviest'*	Today	'Heaviest'*
EuroQol classification (Numbers (%))				
Group 1: mobility				
No problems in walking about	53 (93)	14 (25)	47 (87)	10 (19)
Some problems in walking about	2 (4)	40 (73)	7 (13)	42 (78)
Confined to bed	2 (4)	1 (2)	0 (0)	2 (4)
Group 2: Self care				
No problems with self care	54 (98)	47 (90)	54 (96)	39 (81)
Some problems washing or dressing	1 (2)	5 (10)	2 (4)	9 (19)
Unable to wash or dress	0 (0)	0 (0)	0 (0)	0 (0)
Group 3: Usual activities				
No problems with usual activities	52 (93)	7 (13)	46 (84)	6 (11)
Some problems with usual activities	3 (5)	40 (74)	7 (13)	38 (69)
Unable to perform usual activities	1 (2)	7 (13)	2 (4)	11 (20)
Group 4: Pain				
No pain or discomfort	39 (74)	3 (5)	41 (73)	2 (3)
Some pain or discomfort	14 (26)	30 (54)	14 (25)	25 (43)
Extreme pain or discomfort	0 (0)	23 (41)	1 (2)	31 (53)
Group 5: Emotional				
Not anxious or depressed	34 (65)	3 (5)	40 (73)	6 (11)
Moderately anxious or depressed	17 (33)	32 (56)	15 (17)	31 (54)
Extremely anxious or depressed	1 (2)	22 (39)	0 (0)	20 (35)
EuroQol Visual analogue scores (mean (SE))	82.63 (1.87)	42.51 (3.22)	78.19 (2.71)	42.71 (3.04)

Note: * Day during woman's period when bleeding is heaviest

Table 7.4 presents the subjective health status of the two groups as measured
by the EuroQol instrument. In answering the EuroQol questions, the women
were asked to think about their health both on the day they were completing the
questionnaire and on the day when their period was heaviest. The table shows
in the interview samples what was found in the larger sample of women
completing the questionnaire (see Chapter 4); namely, the significant impact of
women's periods on their subjective health status, with a marked deterioration,
on the 'heaviest' day compared to the day the questionnaire was completed,
throughout the classification and in the visual analogue scores. The responses of
the women were very similar for all EuroQol questions when the two samples are

Table 7.5   *Ex ante* HYEs elicited from women in the Swindon study*

Treatment	Mean (SE)	Median (Range)
AH	34.84 (1.44)	37.5 (0-60)
TCRE	32.62 (1.55)	(34.0 (0-60)

*   Mean years of future life used in TTO exercise was 44.76

compared. If the classification data are assumed to be ordinal, differences between the two samples in each of the groups, on both the day the questionnaire was completed (Mann-Whitney U test, p = 0.14-0.92) and the 'heaviest' day (p = 0.18-0.49), all failed to reach conventional levels of statistical significance. Differences between the two samples in the visual analogue scores too are not statistically significant (Wilcoxon rank sum test, p = 0.59 for day of questionnaire; p = 0.95 for 'heaviest' day).

7.4.2 The valuation exercise

The age of the Swindon sample of interviewees detailed in Table 7.3 translates into the following assumptions about life expectancy for the TTO exercise: six (9.5%) women were in the 20 to 29 years age group (assumed life expectancy 60 years); 22 (34.9%) were in the 30 to 39 years age group (assumed life expectancy 50 years); 31 (49.2%) women were in the 40 to 49 years age group (assumed life expectancy 40 years); and 4 (6.3%) women were in the 50 to 59 years age group (assumed life expectancy 30 years).

Women were asked if they were happy for the interview to be taped, and all but one (1.6%) agreed to this. Interviews lasted for a mean duration of 52.5 minutes (SE 1.2 minutes). Table 7.5 details the *ex ante* HYE values elicited from women in the study.
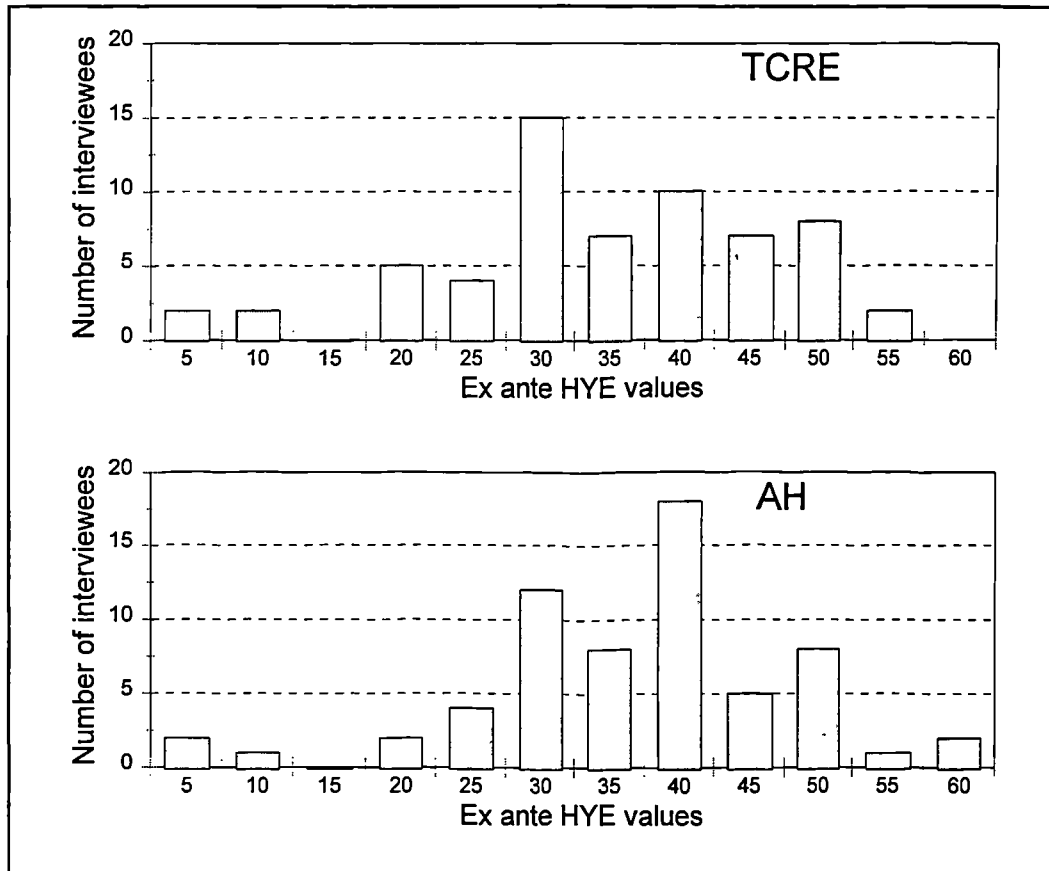
Figure 7.3   Distribution of ex ante HYE values for TCRE and AH among women in the Swindon study

The *ex ante* HYEs show that women considered a mean of 32·62 years (median 34) and 34·84 years (median 37.5) in perfect health to be equivalent to their full life expectancy (on average 44 years) following TCRE and AH, respectively. This difference did not reach statistical significance (mean difference 2.22 (95% CI - 1.98 to 6.42)), probably due to the relatively small sample size, but the results indicate that, on average, women valued the health profile following AH more highly than that following TCRE. The distribution of *ex ante* HYEs amongst women in the sample is shown in Figure 7.3. At the level of the individual woman interviewed, 27 (43%) women valued AH more highly than TCRE, 21 (33%) valued the two interventions equally and 15 (24%) valued TCRE more highly than AH.
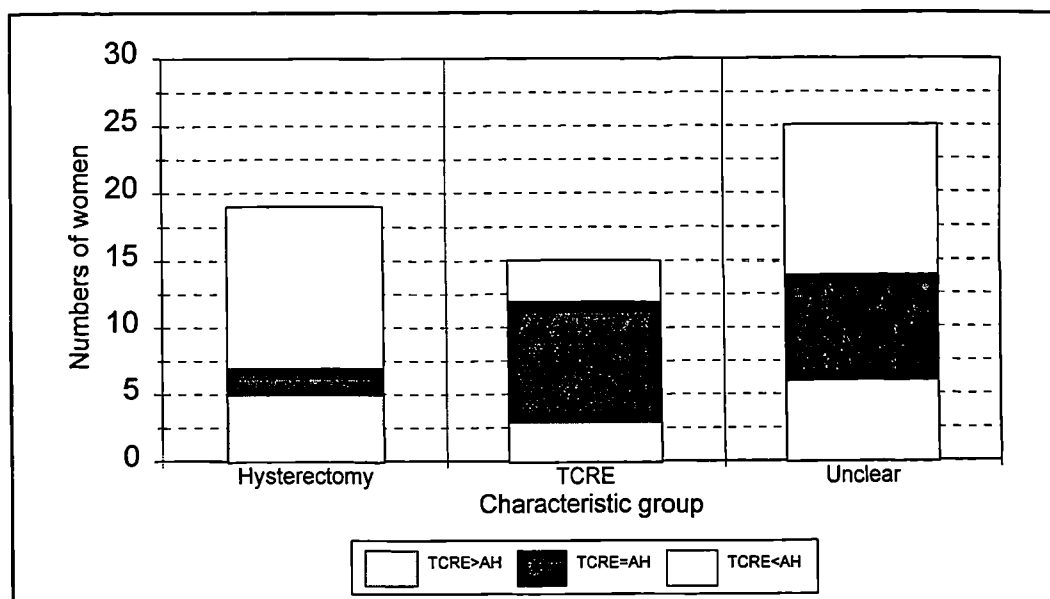
Figure 7.4   Analysis of consistency at the individual level. Comparison of women's relative *ex ante* HYE values for AH and TCRE and their characteristic group implied by their repsonses to the questionnaire

7.4.3 The consistency of ex ante HYEs

Four comparisons are undertaken between women's *ex ante* HYEs for AH and TCRE, and their stated descriptive preferences regarding treatment and the characteristics of treatment provided in the questionnaire.

Ex ante HYEs and characteristic groups. The first of these compares *ex ante* HYEs elicited from women allocated to the two characteristic groups. The consistency of responses at the level of the individual is illustrated in Figure 7.4. The majority of women allocated to the hysterectomy characteristic group (12/19) had a higher *ex ante* HYE for AH than for TCRE. Amongst those women who were allocated to the TCRE characteristic group, the majority (9/15) valued the two treatments the same, with equal numbers valuing AH higher than TCRE and *vice versa*. Amongst those women who could not be clearly allocated to either characteristic group, 11/25 valued AH more highly than TCRE. Therefore, although 8 of the 59 women who completed a questionnaire (and 8 out of 34

Table 7.6   Ex ante HYEs for TCRE and AH according to women's characteristic group*

Characteristic group	N	Mean (SE)		Mean difference (95% CI)	Median (range)	
		AH	TCRE		AH	TCRE
Hysterectomy	19	38.53 (2.06)	34.87 (2.56)	3.66 (-3.00 to 10.3)	37.50 (27.5-60)	34.00 (18-60)
TCRE	15	32.07 (3.24)	33.87 (2.72)	-1.8 (-6.87 to 10.5)	37.50 (6-47.5)	37.50 (18-47.5)
Other	25	33.82 (2.49)	30.40 (2.79)	3.42 (-4.11 to 10.9)	37.50 (0-51)	32.50 (0-51)

* Four patients excluded because their responses placed them in both AH and TCRE characteristic groups. Women placed in AH characteristic group if they considered the following as 'very important' or 'of some importance': a treatment that will remove the womb; and a treatment that will stop periods for good. Women were placed in the TCRE characteristic group if they considered the following as 'very important' or 'of some importance': a treatment that will not remove the womb; a treatment that will cause the least pain and discomfort during convalescence; a treatment that will reduce periods but not stop them for good; and a treatment that will result in getting back to usual activities as soon as possible.

259
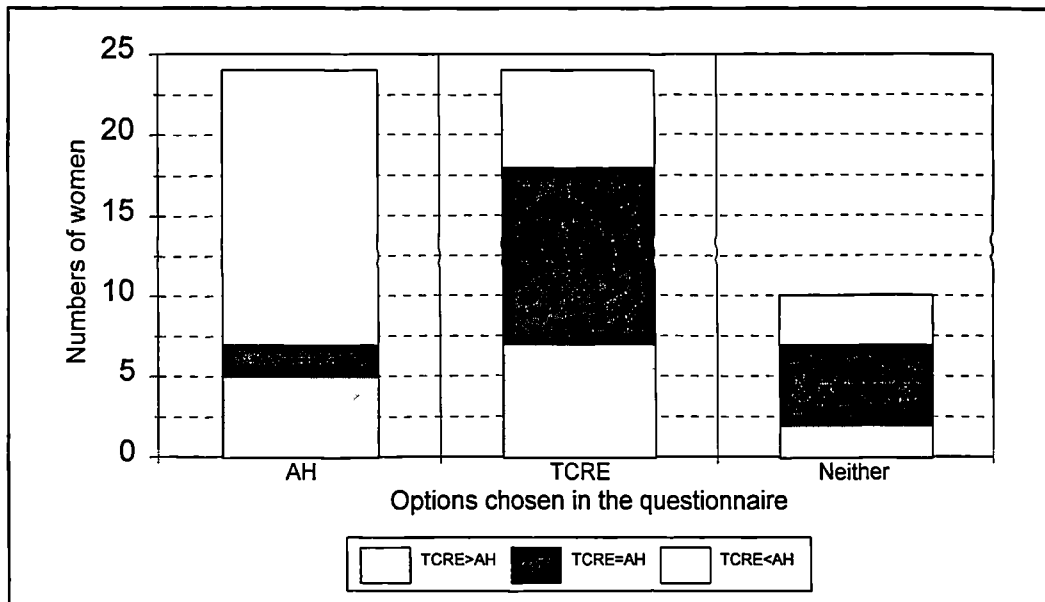
Figure 7.5   Analysis of consistency at the individual level. Comparison of women's relative *ex ante* values for AH and TCRE and the treatment option they chose in the questionnaire

who could be allocated to a characteristic group) gave values which were clearly inconsistent with the characteristic group to which they have been allocated, the majority of women provided *ex ante* HYEs which were not inconsistent with their group.

Some measure of consistency is also shown at the level of the group, as detailed in Table 7.6. The mean *ex ante* HYE for AH is higher than that for TCRE amongst women allocated to the hysterectomy characteristic group, and it is higher for TCRE than for AH amongst women in the TCRE characteristic group. However, although the median *ex ante* HYEs amongst women in the hysterectomy characteristic group are higher for AH than TCRE, they are equal amongst women in the TCRE characteristic group. Although there is no clear indication of inconsistency at the group level, it should be emphasised that the differences in mean *ex ante* HYEs are not significant (ie. the 95% confidence intervals around the mean differences cross 0).

Table 7.7 *Ex ante* HYEs for TCRE and AH according to women's preferences for the treatment options described in the questionnaire

Choice of treatment option in questionnaire	N	Mean (SE)		Mean difference (95% CI)	Median (range)	
		AH	TCRE		AH	TCRE
Option 1 (AH)	24	38.27 (1.87)	32.15 (2.76)	6.12(-0.584 to 12.8)	38.00 (21-60)	31.50 (0-60)
Option 2 (TCRE)	24	29.85 (2.82)	30.69 (2.56)	-0.84 (-6.84 to 8.52)	32.00 (0-51)	33.25 (0-51)
Could not choose or would choose neither	10	37.85 (2.44)	37.25 (2.88)	0.60 (-7.33 to 8.53)	37.75 (26-47.5)	38.00 (25.5-47.5)

261

Figure 7.6   **Correlation plot between the visual analogue scores women gave to the description of AH in the questionnaire and the *ex ante* HYEs for AH that were elicited from them at interview**

Ex ante HYEs and choices about treatment options. The second test of consistency involves comparing women's descriptive preferences for the two treatment options (AH and TCRE) described in the questionnaire with their *ex ante* HYEs for those treatments as elicited in the interview. The results of this analysis at the individual level are shown in Figure 7.5. For this analysis, slightly more women provided unequivocally inconsistent *ex ante* HYE values given their treatment choice in the questionnaire: 11 out of 58 women who completed the questionnaire, and 11 out of 48 who made a treatment choice. However, the majority of women gave *ex ante* HYEs which were not inconsistent with their treatment choice.

The analysis of consistency at the group level between *ex ante* HYEs and treatment choices is shown in Table 7.7. Both the mean and median *ex ante* HYEs elicited from those women who preferred AH as described in the questionnaire are higher for AH than for TCRE. Similarly, although the differences are smaller, the mean and median values for TCRE are higher than

Figure 7.7   **Correlation plot between the visual analogue scores women gave to the description of TCRE in the questionnaire and the *ex ante* HYEs for TCRE that were elicited from them at interview**

those for AH amongst women who preferred TCRE in the questionnaire. Again, however, these difference do not reach statistical significance.

Ex ante HYEs and visual analogue scores for treatment options. The third test of consistency examines the correlation between the visual analogue scores women provided for the TCRE and AH treatment options, as described in the questionnaire, and the *ex ante* HYEs elicited from them in the interview. Although the valuation instrument was different (the TTO instead of the visual analogue scale) and the descriptions were in a different format and included slightly different information, some degree of correlation would be expected. The results, however, do not support this expectation. Figure 7.6 plots the *ex ante* HYEs and visual analogue scores for AH, and Figure 7.7 does the same for TCRE. No clear correlation can be discerned from these plots, and this is confirmed by the statistics: Spearman's rank correlation between the VAS score for AH in the questionnaire and the *ex ante* HYE for AH was 0·149, and was not statistically significantly different from 0 (p = 0.32); Spearman's rank correlation between the VAS score for TCRE in the questionnaire and the *ex ante* HYE for
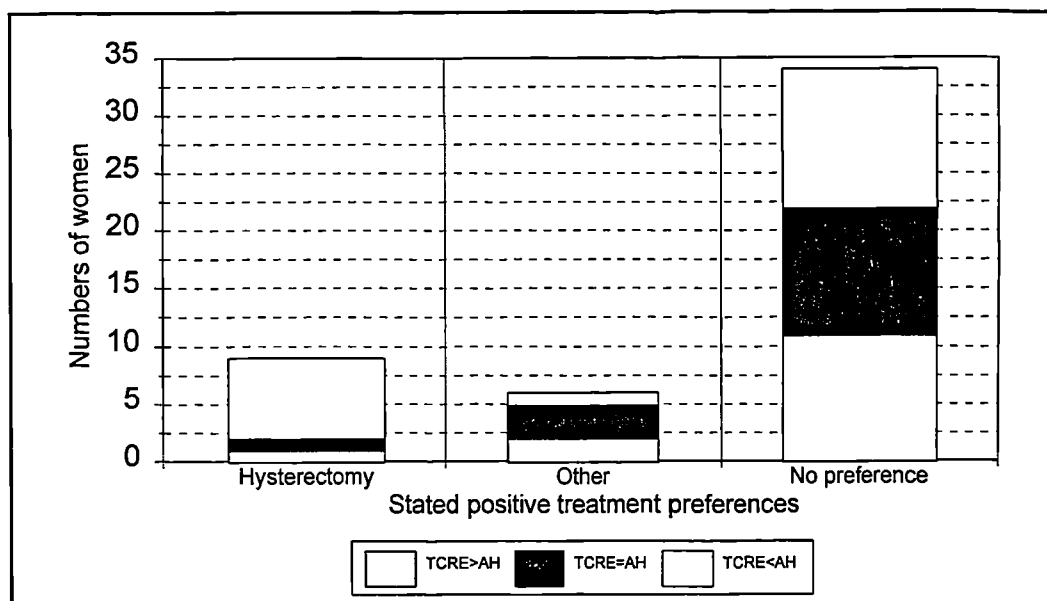
Figure 7.8   Analysis of consistency at the individual level. Comparison of women's relative *ex ante* HYE values for AH and TCRE and their stated positive treatment preferences in the questionnaire

TCRE was -0.273, which was also not statistically significantly different from 0 (p = 0.07). Although the absence of statistical significance is partly related to the relatively small sample size, the size of the coefficients and, in the case of TCRE, the sign, are surprising.

Ex ante HYEs and stated treatment preferences. The final consistency test focused on stated (positive and negative) preferences for actual treatments. Given the small numbers, only two 'treatment groups' are defined: for hysterectomy and for other treatments. The *ex ante* HYEs provided by those women who stated that they wanted a hysterectomy are compared with the *ex ante* HYEs elicited from women who said they wanted some other form of treatment and with those who had no positive preference. Similarly, the *ex ante* HYEs of women who said they did *not* want a hysterectomy are compared to those of women who said they did *not* want some other treatment and to those of women who said they had no negative preference.
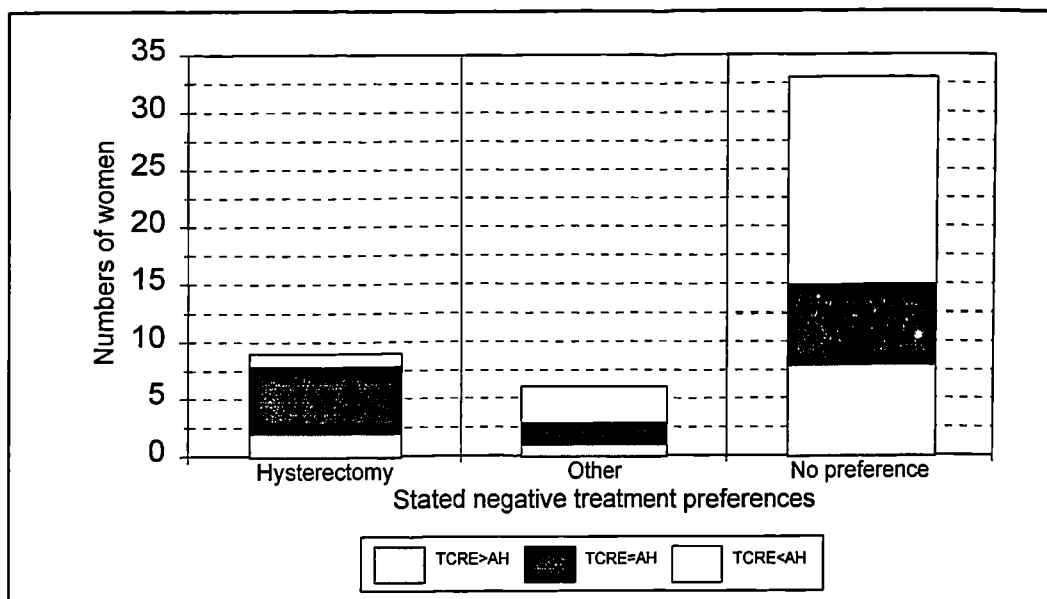
Figure 7.9   Analysis of consistency at the individual level. Comparison of women's relative *ex ante* HYE values for AH and TCRE compared to their stated negative treatment preferences in the questionnaire

The results of the analysis at the level of the individual are shown in Figures 7.8 and 7.9. Figure 7.8 shows that, of the nine women indicating a strong positive preference for hysterectomy, seven (78%) provided higher *ex ante* HYE values for AH than for TCRE. Of the six women who indicated a strong preference for another form of treatment, only one (17%) gave a higher *ex ante* HYE to AH than to TCRE, and two (34%) gave a higher value to TCRE than to AH. Amongst the 34 women who indicated no strong positive preference, roughly equal proportions had higher *ex ante* HYE values for TCRE than AH, higher values for AH than TCRE and equal values

Figure 7.9 shows that, of the nine women who indicated a strong negative preference for hysterectomy, only one (11%) gave a higher *ex ante* HYE to AH than to TCRE, two (22%) gave a higher value to TCRE than AH and the remainder valued the two equally in terms of *ex ante* HYEs. Of the six women who had a strong negative preference for some other treatment, three (50%) had a higher value for AH than TCRE. The majority of the 33 women who had no strong negative preference valued AH more highly than TCRE.

Table 7.8 *Ex ante* HYEs for TCRE and AH according to women's stated positive and negative preferences for actual treatments

Stated preference	N	Mean (SE)		Mean difference (95% CI)	Median (range)	
		AH	TCRE		AH	TCRE
Positive stated preference						
Hysterectomy	9	35.89 (2.32)	27.06 (4.06)	8.83 (-1.09 to 18.70)	37.50 (25.5-50)	27.50 (0-42.5)
Other treatments	6	32.33 (4.67)	33.67 (5.16)	-1.34 (-14.20 to 16.80)	32.00 (18-47.5)	38.00 (18-47.5)
No preference	34	33.51 (2.26)	33.99 (2.09)	-0.48 (-5.66 to 6.62)	37.50 (0-60)	34.00 (0-60)
Negative stated preference						
Hysterectomy	9	22.67 (4.60)	23.33 (4.49)	-0.66 (-13.00 to 14.30)	25.50 (0-38)	25.50 (0-38)
Other treatments	6	37.42 (4.29)	32.17 (4.19)	5.25 (-8.11 to 18.60)	37.50 (22-50)	27.50 (22-47.5)
No preference	33	37.39 (1.38)	33.50 (2.11)	3.89 (-1.15 to 8.93)	37.50 (21-57)	34.00 (0-51)

Table 7.9   Results of the alternative CUA using *ex ante* HYEs, compared to the
               lifetime results from the standard QALY model (1994 prices)

	AH	TCRE	Incremental cost per additional unit of benefit (£)
Standard QALY model [*]			
Expected lifetime cost (£)[†]	1162	816	442
Expected lifetime benefit (QALYs)[†]	15.195	14.413	
Alternative *ex ante* HYE model			
Expected lifetime cost (£)[*†]	1162	816	156
Mean lifetime benefit (HYEs)	34.84	32.62	

[*]   See Chapter 5 for full details
[†]   Discounted at 6% per annum

The results of this analysis of consistency at the group level are shown in Table
7.8. Although the numbers are small, and hence differences do not reach
statistical significance, the mean and median *ex ante* HYEs are consistent with
expectations. The mean and median *ex ante* HYEs for AH are higher than those
for TCRE for women who stated in the questionnaire that they had a positive
preference for hysterectomy, and lower for women who stated that they had a
positive preference for some other treatment. For those women who expressed
no positive treatment preference, the mean *ex ante* HYEs are very similar and the
median values are higher for AH. As regards negative preferences, the mean *ex
ante* HYEs for AH are lower than those for TCRE for women who had a negative
preference for AH and the median values are the same. The median and mean
values are higher for AH than for TCRE amongst women who had a strong
negative preference for some other treatment.

7.4.4 An alternative CUA of AH versus TCRE using ex ante HYEs

The results of the alternative *ex ante* HYE-based CUA are shown in Table 7.9,
alongside the lifetime results from the standard QALY-based CUA. On the basis
of both mean health state values (for the lifetime QALY analysis) and mean *ex
ante* HYEs, AH would be considered the more effective treatment. Furthermore,

when expected lifetime costs are combined with the alternative benefit measure, the incremental cost of an additional unit of benefit appears modest (£442 per additional QALY and £156 per additional *ex ante* HYE). This suggests that the base-case conclusion of Chapter 5 - that, under most circumstances other than extreme simultaneous variation of all parameters and assuming that the suggested illustrative threshold cost per QALY ratios are accepted, AH would be considered a more cost-effective treatment than TCRE - is robust to the use of an alternative means of expressing benefit.

However, the variability around the *ex ante* HYE values is, in terms of the overall results, more important than that around the health state values used to construct QALYs. Indeed, the fact that the mean *ex ante* HYEs are not statistically significantly different (Table 7.5), indicates that conclusions about the relative cost-effectiveness of the two treatments on the basis of *ex ante* HYEs have to be tentative.

An important point illustrated in the comparison of QALYs and HYEs in this context is that the *absolute* value of these benefit measures, for both treatments, over women's full life expectancy is quite different, although this has little effect on the overall *relative* value for money of the two treatments. The difference is due predominantly to the effect of discounting at a constant positive rate in the standard QALY-based CUA. The long time horizon of the analysis (44 years average life expectancy) means that, in the QALY-based CUA, a QALY occurring four years from death is, in present value terms, worth only 0.10 of a QALY on the basis of a 6% discount rate. The *ex ante* HYEs contain within them women's time preference rates, so no subsequent adjustment using the constant exponential discount rate approach is necessary.

The *ex ante* HYEs elicited from women in this study would seem to indicate that discounting benefits at a constant exponential rate of 6% may not reflect individual preferences. If a 0% discount rate is substituted into the standard QALY model, the absolute estimates of QALYs become much closer to that of

HYEs, for both treatments. This result is consistent with growing evidence to indicate that individuals do not exhibit a positive time preference [Loewenstein and Prelec, 1993; Redelmeier and Heller, 1993; Dolan and Gudex, 1995]. In certain contexts, this apparent failure to reflect accurately individuals' time preferences using discounting would cause more significant differences between the results and conclusions of a QALY-based and a *ex ante* HYE-based CUA, than appear to exist in relation to TCRE versus AH. For example, if the time distributions of benefits differ markedly between the treatment options under consideration, results could be very sensitive to the choice of discount rate [Petrou *et al*, 1993].

7.5   Discussion

7.5.1 Theory of benefit measurement in CUA

The QALY has become a widely used tool in the economic evaluation of health care programmes and interventions [Gerard, 1992]. However, the feeling that it can facilitate the incorporation of patients' preferences into the calculus of economic evaluation is not supported by strong evidence. For the economic evaluation of MAS technologies, where patients preferences may be considered important to factor into the analysis, this may be a major shortcoming of QALYs.

Two questions must, then, be asked about QALYs in the economic evaluation of MAS. The first question is whether the use of the QALY as a benefit measure within CUA is preferable to reverting to simple CEA or moving to cost-benefit analysis (CBA). When considering the role of CEA, MAS invariably generates multi-dimensional outcomes, as emphasised in Chapter 4 in relation to the AH versus TCRE comparison. These outcomes might be mortality and HRQL; more frequently there is a series of effects on different dimensions of HRQL. Furthermore, MAS often has an important effect on the *process* of care over which individuals are likely to have preferences; for example, location of treatment. Multi-dimensional outcomes and the importance of process result in

269

CEA having limited value in helping decision makers reach conclusions about relative cost-effectiveness. Invariably, the use of CEA shifts the task of synthesising the multi-dimensional effects of MAS, and the burden of making judgements about the value of the outcomes of those technologies relative to those generated by other interventions in different programmes and disease areas, on to decision makers. This process is usually hidden from scrutiny and is often implicit.

The strength of CUA is that the value judgements necessary to synthesise multi-dimensional measures of outcome into a single benefit measure - namely the health state values and the assumptions underlying the construction of QALYs - are (or should be) made explicit. When presented with the results of a QALY-based economic evaluation, a decision maker can accept or reject its conclusions; but in rejecting them, alternative value judgements will have to be discussed and presented. The QALY, therefore, can serve a valuable role within resource allocation: as one tool in the decision maker's armamentarium for purposes of resource allocation; as a means of making judgements about the synthesis of multi-dimensional outcomes explicit; as a broad-brush means of comparing outcomes, as well as costs, across programmes and disease areas; and as a way of initiating a consideration of the economic characteristics of health care technologies. However, this 'decision making perspective' on the value of QALYs and CUA is quite different to the view that QALYs are a means of incorporating individuals' preferences into resource allocation.

In order for CBA to be used as a framework within which to assess the relative value for money of MAS interventions, some way of valuing the outcomes and process of health care in monetary terms is required. In recent years the methods of willingness to pay have been used more widely in economic evaluation of health care [O'Brien and Viramontes, 1994; O'Brien *et al*, 1995; Donaldson *et al*, 1995; Chestnut *et al*, 1996]. An advantage of CBA is its focus on allocative efficiency, supporting decisions about the most appropriate level of funding for the health service as a whole, as well as about allocation within it.

However, the focus of this thesis is the most appropriate means of measuring benefits from interventions with a range of process characteristics and outcomes, to inform decisions about resource allocation within the health service. In that context, willingness to pay methods would seem to offer no methodological advantages over other forms of valuation, and may introduce some practical difficulties, especially in the UK where individuals may find it unacceptable to consider payment for health-related outcomes even at the hypothetical level [Propper, 1988; Donaldson *et al*, 1995].

Indeed, as Johannesson [1995] comments, there are similarities between a TTO exercise to generate *ex ante* HYEs and a conventional willingness to pay exercise where individuals are presented with descriptions of outcomes in terms of uncertainty and asked how much they would be willing to pay (accept) to avoid (experience) those risks [Gafni, 1991]. With the TTO exercise, though, the numeraire is years of life rather than money. If economic evaluation is to move away from the flexible standard QALY towards a measure of value that may be able to reflect individuals' utility functions more adequately, it is an important question for further research whether willingness to pay or *ex ante* HYE methods are the preferred advancement.

The second question that should be asked about the use of QALYs to evaluate MAS is whether there are any ways of strengthening the theoretical basis of the benefit measure used in CUA, to adhere more firmly to individuals' preferences. Section 7.2.2 argues that the *ex ante* HYE requires fewer assumptions to link it to individuals' preferences and, in this sense, may be considered theoretically stronger than the QALY. However, HYEs may impose a greater measurement burden on the analyst in the form of a need for more detailed descriptive scenarios. Furthermore, HYEs are inherently less flexible than QALYs. The decision analytic model is a popular framework for QALY-based CUA, where parameters such as probabilities and durations in health states can be varied, and the implications for cost and benefit results assessed, without altering the health state values. Because scenarios for HYEs include more information, the HYE

estimates have analytical parameters locked within them, and an assessment of the robustness of an analysis to changes in these parameters is impossible unless revised HYEs are elicited.

A clear trade-off, therefore, exists in selecting benefit measures for CUA of MAS. QALYs are relatively easy to estimate, can be based on 'off the shelf' health state values from valuation systems like EuroQol and are flexible when incorporated into decision analytic models. However, there are clear doubts about the consistency of QALYs with individuals' preferences. *Ex ante* HYEs require fewer assumptions to link them with individuals' preferences, but are less flexible for use in economic evaluation and probably impose a greater valuation burden.

Quite what direction economic evaluation should take, given this trade-off, is as yet unclear. A number of issues need to be considered. Firstly, further research may be able to make the trade-off less pronounced. For example, it may be possible to identify valid and efficient means of eliciting *ex ante* HYEs from large numbers of valuers, using computer and video technology, such that the valuation burden is reduced. Alternatively, new QALY models could be developed which link the benefit measure more closely to preferences.

The second point to note is that the choice between QALYs and HYEs will depend on whether they lead to radically different conclusions when they are used in practice; if they do not, then the choice is less crucial. To date, experience with HYEs in applied studies has been limited. It is important for further research to be undertaken, in the context of applied evaluations, to compare the two benefit measures. It is likely that the choice of measure will be crucial in specific contexts: for example, where there is a range of different risks associated with the technologies under comparison, where the process and likely outcomes of care differ markedly between the comparators or where the timing and/or sequence of outcomes differ between comparators. Further research is needed to identify such contexts.

A third point to note about the choice between HYEs and QALYs is that, if HYEs were to be used widely in economic evaluation of health care interventions, it would require a rather different approach to the timing of benefit valuation work within the overall study. If a RCT is the major source of outcome information for an economic assessment, the valuation data necessary to facilitate a QALY-based CUA would usually be collected within the trial using an instrument such as the EuroQol. If HYEs are to be estimated, the valuation exercise would have to await the results of the trial, so that descriptive profiles could be developed to include all relevant information. If effectiveness data are being generated using modelling techniques rather than as part of a trial, the estimation of QALYs would require the valuation data to be incorporated into the model, with QALYs being a major outcome of the exercise. With HYEs, however, the model would have to be used to generate the information to go into the descriptive scenarios, which would then be used to elicit HYEs.

7.5.2 *The consistency of HYEs and women's preferences*

The validation of health state values is notoriously difficult, as there is no gold standard for health state preferences apart from actual behaviour, which is difficult to observe. Hence the rigour in assessing validity and reliability, which is such an important part of developing *descriptive* HRQL instruments [McDowell and Newell, 1987], cannot easily be replicated with valuation exercises. Within the Swindon HYE exercise of AH and TCRE, women's completion of a questionnaire exploring their attitudes to the treatment of menorrhagia afforded some opportunity to explore the consistency between *ex ante* HYEs and stated preferences about treatment and the characteristics of treatment. The relatively small sample size limited the statistical power of this exercise and, as with any analysis of this kind, judgements have to be made about what is sufficient evidence to indicate consistency, but it is possible tentatively to identify some indication of consistency.

Evidence of consistency seems stronger at the level of the group than that of the individual. This observation has also been made in relation to the consistency of

the ranking of health state values with a 'logical' order [Measurement and Valuation of Health Group, 1994]. At the individual level, the majority of women did not unequivocally contradict with their *ex ante* HYE values their various descriptive preferences in the questionnaire. However, between 7% and 19% did provide inconsistent *ex ante* HYEs and only between 25% and 47% gave values which were unequivocally consistent with their descriptive preferences. Despite not reaching statistical significance, at the group level, the mean and median *ex ante* HYEs for AH were always higher than for TCRE amongst women whose descriptive preference was for AH, and lower or the same as TCRE amongst women whose descriptive preference was for TCRE. If one makes the assumption that women's responses to a series of short questions in a postal questionnaire are likely to be a reliable descriptive gauge of their preferences, then consistency between these responses and *ex ante* HYEs, elicited at interview using a choice-based valuation instrument, is important to demonstrate.

7.5.3 The alternative CUA of AH versus TCRE

As an element of the overall economic evaluation of TCRE versus AH, the alternative CUA using *ex ante* HYEs is a form of sensitivity analysis focusing on uncertainty in analytical method [Briggs *et al*, 1994]. In other words, given the controversy that surrounds the relationship between QALYs and individuals' preferences and the likely importance of women's preferences in the management of menorrhagia, the alternative CUA assesses how robust the conclusions of the CUA model described in Chapter 5 are to an alternative approach to benefit measurement.

The characteristics of these two surgical treatments for menorrhagia are such that the assumptions of standard QALYs might be considered overly simplistic. In particular, the trade-offs that face women and clinicians in selecting between the treatments are unlikely to be adequately represented, in terms of preferences, by a standard QALY model. The *ex ante* HYE approach presents women with a simplified picture of the risks and benefits of TCRE and AH, with

a clear time dimension extending until the menopause and then until death. The descriptive scenarios used to elicit the *ex ante* HYEs are similar to *the* information a woman might expect her clinician to provide in order to help her to choose an appropriate therapy. Using these scenarios as a basis to elicit *ex ante* HYEs provides an all-embracing treatment-related measure of benefit for use in CUA which, in principle, reflects women's attitudes to the risks, processes and outcomes involved.

Given that HYEs are relatively underdeveloped methodologically and have been used in very few empirical studies, it is reasonable that the HYE-based CUA is the 'alternative analysis' to the 'base-case' of the standard QALY model. The results suggest that the broad conclusions of the QALY model - that the additional benefits generated by AH compared to TCRE come at a relatively modest incremental cost- hold true for the HYE-based analysis.

A number of caveats should be born in mind in reaching this conclusion, however. The first of these is that the conclusions of both CUAs are sensitive to the variability in the benefit data elicited from the sample of women valuers. In the QALY-based CUA, the key health state values are those for the pre-menopausal (post-convalescence) periods after TCRE and AH. The variability around the *ex ante* HYE estimates was even greater than for the health state values to the extent that there was no statistically significant difference between TCRE and AH. If interpreted strictly, therefore, the 95% confidence intervals around the mean difference in *ex ante* HYE values could imply, at one extreme, that TCRE dominates AH (less costly and more effective) or, at the other extreme, that AH is probably better value for money with a very low incremental cost per additional unit of benefit.

The second caveat regarding the comparative results of QALY- and HYE-based CUA relates to the scope of the benefits considered. Given that the follow-up period of the firm evidential basis of the CUA - the Bristol trial - is only two years, a number of the trade-offs that are likely to exist in the longer-term

between TCRE and AH are not explicitly considered. Hence, although the time horizon of the comparison between the two CUAs was the average life expectancy of a woman with menorrhagia, some of the risks from one treatment relative to the other after the menopause are not reflected in the model. For example, AH provides prophylaxis against endometrial and, usually, cervical cancer, which gives it a differential risk profile relative to TCRE. However, some evidence links AH to premature ovarian failure and, hence, possible early menopausal symptoms, for women not taking hormone replacement therapy. The standard QALY analysis in Chapter 5 does not build in these differential risks into the model because, given that TCRE has been used for a relatively short period of time, no estimates of their magnitude exist; furthermore, the long-term effects of hysterectomy, as well as TCRE, are not well understood. Similarly, the descriptive scenarios upon which the *ex ante* HYEs were estimated contained no information about these risks, although it is possible that women in the valuation exercise augmented the information in the scenarios with their own from other sources. As discussed in Chapter 5 in the context of the QALY-based CUA, the effect of including these longer-term differential risks in the CUA would almost certainty confirm the base-case conclusions of the analysis, that AH offers additional benefits over TCRE at modest incremental cost, because the risks on balance favour AH.

A third caveat concerning the results of the two CUAs is that both analyses are based on the premise that one of the two therapies should be preferred, in terms of relative value for money. An implication of this is that purchasers will select one therapy, based on relative cost-effectiveness, for those women for whom either is *clinically* feasible. However, given the fact that relative cost-effectiveness is so sensitive to women's preferences concerning outcomes, this policy would be based on the view that mean benefits are of primary importance. Based on the results described in this chapter and in Chapter 5 and assuming the illustrative cost per QALY thresholds are acceptable, this 'all or nothing' approach would probably select AH as the more cost-effective treatment option for all women, despite the fact that a sub-group of women have preferences

which would suggest that TCRE is the better treatment for them. Contrary to
the way most economic evaluations are structured and undertaken, it makes
sense to explore the costs and benefits of determining a woman's treatment
based on her preferences. A full economic evaluation of this preference-based
strategy would rest heavily on the type of valuation data used in the CUAs
reported in Chapters 5 and 7.

7.6   Conclusions

Given the absence of evidence confirming a clear link between QALYs and
patients' preferences, this chapter has considered how an alternative benefit
measure to the TTO-based standard QALY might be used in a CUA of MAS. The
use of the TTO-based *ex ante* HYE has been shown to be feasible in the context
of the evaluation of AH and TCRE, and the chapter has described some results
showing that women's *ex ante* HYEs are consistent with their descriptive
treatment preferences. The alternative CUA using *ex ante* HYEs generates an
incremental cost per additional unit of benefit which is relatively modest
suggesting, as did the base-case analysis, that AH is likely to be the more cost-
effective treatment. However, the variability around the *ex ante* HYE estimates
generates substantial uncertainty about this conclusion. Although AH may be
considered the more cost-effective of the two surgical therapies, the *ex ante*
HYE values suggest that TCRE would be the preferred therapy from the
viewpoint of a proportion of women. Chapter 8 considers the costs and benefits
of allocating treatment on the basis of women's preferences.

Appendix 7.1   Descriptive scenarios used in the Swindon *ex ante* HYE valuation study

TCRE

She suffers from heavy and painful periods in the same way as the person in the first description that you read, and is about to have surgery for the condition. This will involve:

- a small risk of death of 1 in 1000;

- a stay in hospital of 1 day;

- an interval of about 1 week before she resumes her daily activities, during which she will experience some discomfort and sometimes feel tired;

- an interval of about 2 weeks before she can return to work;

- an interval of about 3 weeks before she can resume her sex life;

- the operation does not leave a scar.

Once she has recovered from the operation, she experiences the following results from surgery after about 4 months:

- she still has periods but they are much lighter since her operation;

- she still has some pain with her periods;

- she has no limitation on her social activities or daily activities such as work;

- she occasionally feels moody, irritable or depressed;

- she still has her womb, but it is unlikely that she will become pregnant.

- she is able to enjoy her sex life.

Two years after surgery she is likely to be happy with the results of her treatment and be in good health. However, during this time she would have faced the following risks:

- a 12% chance that she would have had the same operation again because of her menstrual problems returning;

- a 16% chance that she would have had another type of surgery because of her menstrual problems returning, involving 6 days in hospital and 4 weeks away from her usual activities.

Within 5 years after surgery any bleeding she still has ceases due to the start of her menopause.

She lives in good health for the remaining 35 years of her life.

AH

She suffers from heavy and painful periods in the same way as the person in the first description that you read, and is about to have surgery for the condition. This will involve:

- a small risk of death of 1 in 1000;

- a stay in hospital of about 6 days;

- an interval of about 4 weeks before she resumes her daily activities during which she will feel tired, need occasional pain killers and be unable to lift objects or walk very far;

- an interval of about 11 weeks before she can return to work;

- an interval of about 6 weeks before she can resume her sex life;

- she no longer has a womb, so she is unable to bear children;

- she will be left with a faint scar on her abdomen.

Once she has recovered from the operation, she experiences the following results from surgery after about 4 months:

- she no longer has periods or experiences pain;

- she has no limitation on her social activities or daily activities such as work;

- she occasionally feels moody, irritable or depressed;

- she is able to enjoy her sex life.

Two years after surgery she is happy with the results of her treatment and in good health.

Her menopause starts within 5 years after surgery and she lives in good health for the remaining 35 years of her life.

Chapter 8

Incorporating Women's Preferences into the Economic Evaluation of Surgical Treatments for Menorrhagia

8.1   Introduction

The conventional approach to economic evaluation in health care is to compare the costs and benefits of two or more interventions with the aim of establishing which one is the more cost-effective. This is invariably conducted by representing the distributions of costs and benefits of the options under comparison in terms of their means, undertaking incremental analysis using these measures of central tendency and employing sensitivity analysis to explore the importance of variability. This can be termed an 'all or nothing' approach to economic evaluation, because it is usually expected that purchasers would allocate resources towards the single economically superior option. There are circumstances where it is accepted that the all or nothing approach may be inappropriate; these relate to situations when there are clear and important sources of heterogeneity between patients in terms of their clinical characteristics. Here, it may be the case that an intervention is cost-effective for one sub-group of patients but not for others.

It may also be inappropriate to use the all or nothing approach when there is
marked variation in patients' preference characteristics. However, it is very rare
in economic (and clinical) evaluation for sub-groups to be defined in terms of
patients' preferences, and for the costs and benefits of preference-based
management strategies to be assessed. This remains the case despite research
showing that patients have preferences about treatments and outcomes [McNeil
et al, 1978], and that many want to participate in decision making [Strull *et al*,
1984].

MAS interventions, in particular, are characterised by a range of treatment
processes and outcomes over which patients are likely to have preferences.
Using the case-study of TCRE and AH, it is the aim of this chapter to consider
ways in which the economic evaluation of MAS can be extended to assess
management strategies which allocate patients to treatment on the basis of their
preferences. In Section 8.2, alternatives to the all or nothing approach to
economic evaluation are considered in more detail. In Section 8.3, various ways
of modelling the costs and benefits of preference-based management strategies
are discussed. The potential cost-effectiveness of preference-based
management strategies are then compared with standard management (TCRE
only or AH only). Three preference-based management strategies are modelled
and evaluated: treatment allocation on the basis of patient choice (Section 8.4);
treatment allocation on the basis of patient values (Section 8.5); and treatment
allocation on the basis of patient-specific cost-effectiveness analysis (Section
8.6). Section 8.7 discusses the analysis and Section 8.8 offers some
conclusions.

8.2   Sub-group analysis in economic evaluation

8.2.1 Clinical sub-groups

The focus of economic evaluation in health care is the comparison of alternative
interventions to identify which single option is the more cost-effective based on
mean costs and benefits. It would then be hoped that purchasers would adopt

this single economically superior intervention or programme. This all or nothing approach has weaknesses at a number of levels [Asch and Hershey, 1995]. Reflecting one of these weaknesses, the all or nothing focus may be considered inappropriate if sub-groups of patients can be defined in terms of *clinical* and demographic characteristics which are felt to influence outcome. In this situation, an assessment can be made of whether a given form of management might be cost-effective for one sub-group, whilst not being so for others or for the population of patients as a whole.

For example, in an economic evaluation of antihyperlipemic therapy in the prevention of heart disease, it was found that the cost-effectiveness of therapy varied considerably by sub-group of patient, with incremental costs per life year saved ranging from $36,000 to $1 million [Oster and Epstein, 1987]. The authors reported that therapy was more likely to be considered cost-effective for younger patients with multiple coronary risk factors and severe elevation of cholesterol levels. This form of sub-group analysis has the potential to improve the cost-effectiveness of health care delivery markedly, but its use will always be constrained by the limited data on baseline clinical characteristics which influence outcomes. For example, the cost-effectiveness of TCRE could be increased relative to AH if more was known about the clinical factors likely to determine which women fail on the treatment.

8.2.2 Preference sub-groups

If efficient health care delivery is concerned with how patients value the process and outcomes of interventions, in some situations the cost-effectiveness of alternative interventions may be sensitive to patients' preferences. A major implication of an all or nothing approach in such a context is that an optimal treatment for a population of patients, based on *mean* costs and benefits, may be sub-optimal from the perspective of a single patient. This point is illustrated in Figure 8.1 which relates to a population of women requiring surgical treatment for menorrhagia. The figure shows a hypothetical probability distribution of net QALYs of AH relative to TCRE; that is, the QALYs that each woman would

282
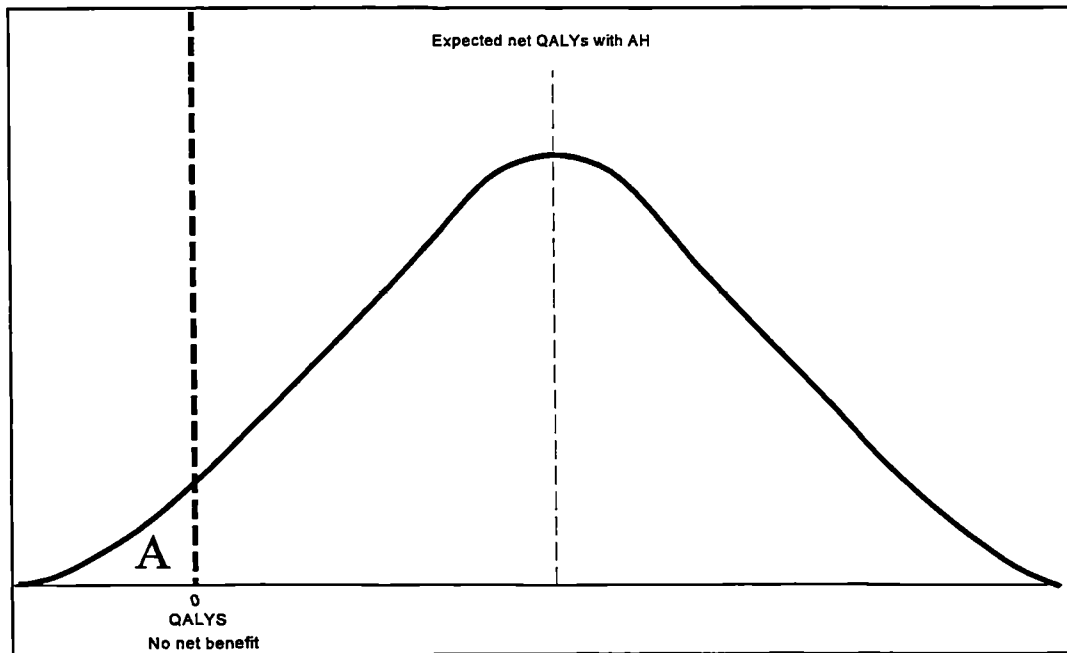
Figure 8.1   Hypothetical probability distribution of net QALYs associated with AH
relative to TCRE. Adapted from Heald and Owens [1994].

expect if she were to undergo AH, minus the QALYs each woman would expect
if she were to have a TCRE. The example assumes that the differences between
women in terms of their individual net QALYs are driven solely by differences in
preferences (ie. they are clinically homogenous but heterogenous in terms of
values they would attach to health states).

The mid-point of the distribution shows the mean (expected) net QALYs with AH
which, together with expected incremental costs, is the basis of all or nothing
economic evaluation. The majority of women in this notional population would
experience positive net QALYs with AH. So, confining the comparison to these
two treatments alone, an all or nothing decision to purchase AHs for the entire
group would be individually optimal for most women. However, because of the
nature of their values, the all or nothing policy would be individually sub-optimal

for a minority of women in this population, because they would experience
negative net QALYs from AH (equivalent to the area marked 'A' in Figure 8.1)[1]

In the context illustrated in Figure 8.1, it may be appropriate to undertake sub-
group analysis where the sub-groups would be defined in terms of patients'
preferences. Hence, if treatment were allocated to patients on the basis of their
preferences, this may prove more cost-effective than reliance on one intervention
alone. Furthermore, it is possible that patients value the process of choice in
itself [Ryan and Shackley, 1995], which may further increase the potential cost-
effectiveness of this strategy[2].

The term 'patient preference' is often used in relation to two distinct concepts.
The first of these is *patient choice*, where patients select a treatment that they
consider better from their perspective using any information they have access to.
The second concept is *patient values,* where patients show the strength of their
attitude to a health state or a prognosis following treatment on a cardinal scale.
Both of these concepts could be used to define patient sub-groups and treatment
allocation strategies.

Patient choice. In some disease areas, many patients have clear ideas
about the characteristics of treatment that are important to them and are keen to
be able to choose the treatment they undergo based on relevant information
[McNeil *et al*, 1978]. The surgical treatment of menorrhagia appears to be one
such area. Chapter 4 of this thesis indicated that, when information on AH and

---

[1]The conflict between the group and the individual optimum might be even more
extreme than shown in Figure 8.1. In principle, net QALYs might be positive with AH
on a group basis, but the treatment might result in *negative* net QALYs for the *majority*
of women. This could happen if the positive net QALYs for the minority of women for
whom AH is the personal optimal choice are very large, but the negative net QALYs
associated with AH in the majority of women for whom TCRE is the personal optimal
choice are small

[2]Evidence does exist, moreover, to indicate that allowing patients' preferences to
determine choice of therapy can have a direct influence on outcomes [Greenfield *et al*,
1985; Brody *et al*, 1989].

TCRE was presented to women with menorrhagia, over 80% were willing to choose one of these therapies. The chapter also showed that different characteristics of treatment were more important to some women than to others, and this may explain the fairly even split between women who would choose TCRE, based on the information provided, and those who would choose AH.

Patient values. In many economic evaluations, the cost-effectiveness of an intervention will be sensitive to the values attached to health states. This situation has been shown clearly to exist in relation to the comparison of AH and TCRE in the cost-utility analyses reported in Chapters 5 and 7 of this thesis. In this context, patients' values could be elicited prior to their knowing the outcome of treatment for them (ie. *ex ante* values), which could then be used to predict which therapy would be optimal from their individual viewpoint.

Studies have explored how values could be used as a basis of individual decision making. For example, a recent study of how women value the information generated by antenatal screening explored the concept of individual decision making based on the individual values elicited from a sample of women [Cairns *et al*, 1996]. The study found that expected benefit would be maximised for the majority of women if screening was used, but the type of screening strategy that would generate greatest benefit was finely balanced, using analysis at the level of both the individual and the group. Few studies have looked at the cost-effectiveness of management strategies using treatment allocation based on patients' individual values.

To an extent, patients' treatment choices would be expected to reflect their health state values. However, there may be inconsistency between these two concepts. As reviewed in Chapter 7, there is evidence to indicate that when health state values are used to estimate QALYs, this measure of benefit may not be consistent with individual choices. Given a possible inconsistency between choices and values, management strategies could be designed which allocate patients to treatments using either concept.

8.3   Building patients' preferences into economic evaluation

An all or nothing approach to the economic evaluation of TCRE and AH would almost certainly result in some women receiving a treatment which is sub-optimal from their individual perspective when their individual preferences are considered. It is important to assess, therefore, whether a different approach to economic evaluation is feasible; one which explicitly considers the costs and benefits of allowing patients' preferences (either choice or values) to determine treatment allocation. This is quite distinct from the general methods of CUA which only provide a role for patients' values within the evaluation calculus at the group level, which is where decision rules about treatment allocation are conventionally set.

This extension to the methods of economic evaluation involves adding one or more comparators to the interventions under consideration, where these comparators are management strategies which allocate an intervention to a patient on the basis of their choice or values. This new comparison may not, however, require an entirely new study design. Using sub-groups, the costs and benefits of the new comparators can be modelled if data are available on the values and/or treatment choices of a cohort of patients. Below, three specific approaches are discussed, each in the context of the economic evaluation of AH versus TCRE. These are treatment allocation on the basis of patient choice, on the basis of patient values and on the basis of patient-specific cost-effectiveness analysis (CEA).

8.3.1 Treatment allocation on the basis of patients' choices

The first approach is to undertake an economic evaluation comparing three forms of management for menorrhagia: TCRE only, AH only and a strategy where the woman chooses which of these two treatments she would like after being given relevant information. A modelling approach to the evaluation of a choice-based management strategy would require some specific data collection, but could take the following form, in relation to the treatment of menorrhagia.

A cost-utility model would be developed to compare the costs and benefits of TCRE and AH, where benefits would be expressed in terms of expected QALYs. A group of women with menorrhagia would then be interviewed, given appropriate information about TCRE and AH and asked which treatment they would prefer, if any. The cohort would then be asked to value a series of health states associated with the outcomes of TCRE and AH, which would be the building blocks of the cost-utility model. The model would then be used to assess the aggregate expected costs and benefits of a TCRE only and an AH only treatment policy, where benefits would be calculated using expected QALYs based on the mean values the group attach to each health state.

To assess the costs and benefits of the choice-based approach to management, each woman in the sample would be considered separately: the expected cost and benefit would be calculated based on the cost of the treatment each woman chooses; and the expected benefit would take into consideration the QALYs for each woman's choice of treatment based on her own health state values relating to the chosen treatment. For example, if the first woman in the sample chose TCRE, her cost would be the expected cost of TCRE calculated for the TCRE only option; the QALYs associated with her management would be calculated by 'plugging' into the model *her own* health state values related to TCRE and identifying the expected QALYs for TCRE. If the second woman in the sample preferred AH, the cost of her management would be the expected cost of AH based on the AH only option, and her QALYs would be calculated using her own health state values relating to AH, incorporating them into the model and using the expected QALY estimate for AH.

Decision rules would be needed regarding the treatment allocation for patients who are indifferent between therapies, or who want the doctor to make the decision for them. In the case of indifference, allocation to the cheapest treatment option may be justified. The flexibility of the modelling approach to evaluating this strategy would be an advantage in this respect, as alternative rules could be explored to assess the robustness of the results. It is important to

note that this approach to the evaluation of choice-based management would be from an *ex ante* perspective: women's choices and values would be elicited prior to their actually undergoing any treatment.

Therefore, each woman in the sample would have a cost and QALY estimate allocated to her on the basis of her choice of treatment. The mean cost and QALYs of the sample would then be used as an estimate of the cost and benefit of the choice-based strategy, and compared with TCRE only and AH only options. The decision rule to select between the three options would be the same as normal in CUA at the group level. An economic evaluation of choice-based management would obviously need to consider the resource cost of providing women with sufficient information to make a treatment choice and of eliciting that choice.

As discussed in Chapter 7, the calculation of QALYs is based on strong assumptions about how individuals make choices, and the chapter referred to evidence that would suggest that treatment-specific QALY estimates, based on an individual's own health state values, may not accurately predict the treatment which the individual would choose to have. If the assumptions that link QALYs with individuals' preferences are unsound, it may be unwise to use health state values and patient-specific QALYs as a way of evaluating choice-based forms of patient management. An alternative way of measuring values would be in terms of *ex ante* HYEs which, as discussed in Chapter 7, require fewer assumptions in principle to link them to actual treatment preferences.

8.3.2 Treatment allocation on the basis of patients' values

Section 8.2 makes the distinction between the *choices* patients make about treatment options, and the values they have for particular health states or prognoses. An alternative to a management strategy where patients would be directly asked to choose which treatment they would wish to have, is one which focuses solely on their values. This strategy would have the advantage of not obliging the patient to make a direct choice: in the case of a QALY-based

approach, patients could indicate a preference for a series of health states rather than come to an overall decision about a treatment. On the other hand, the choice-based approach is less constraining for the patient as they can take into consideration what they wish when selecting a preferred therapy.

For the patient values-based management strategy, a decision analytic model would be developed into which each woman's health state values would be incorporated individually to identify the treatment with the highest patient-specific expected benefit, to which the woman would then be allocated as her personal optimal treatment. As before, the expected QALYs and cost of each woman in the sample would be estimated, based on her receiving her individually optimum treatment, which would be compared with the expected costs and benefits of TCRE only and AH only, and the normal decision rule for group-level CUA would apply.

An economic evaluation of a management strategy which allocates treatment on the basis of women's values would need to consider the resource cost of eliciting values as part of routine clinical practice. Unlike the form of management where women are allocated to a treatment based on their choices, eliciting values would probably require each woman to be interviewed, incurring opportunity costs in terms of the time of health service staff and the women themselves.

In order for values (and hence treatment allocation) to reflect patients' underlying utility functions more accurately, it would be possible to express individual values in terms of *ex ante* HYEs rather than health state values and expected QALYs. The difference between the choice-based and the values-based allocation of treatment is clear when the values relate to health states. For a woman's direct choice between AH and TCRE to be identical to that implied by incorporating her health state values into a decision analytic model, not only would she have to make decisions by maximising her expected utility, also the 'model' she forms in her mind when selecting a treatment would have to be

identical, in terms of health states, probabilities etc., to the formal model. Clearly, this is very unlikely to occur in practice. However, if *ex ante* HYEs were the means of measuring values, it would be expected that a woman's choice of actual treatment would be reflected in her HYE values, because she would be valuing a prognosis directly related to a specific treatment, rather than a series of health states which only subsequently are related to a treatment using an external model.

8.3.3 Treatment allocation based on individual cost-effectiveness

The two management strategies described above would allocate women to their individual optimal treatment on the basis of either their direct choice or their values. As a basis for allocating limited resources at a group or societal level this strategy would be inappropriate, because, at the level of the individual, it ignores the cost implications of treatment selection. The logic applied at a group level of asking whether the expected incremental benefits generated by an intervention justify its additional costs can also be applied at the level of the individual patient and incorporated into management strategies based on patients' choice or values.

A strategy of allocating treatments on the basis of patient-specific CEA was discussed by Nease and Owens [1994]. The authors related the concept to the development of clinical guidelines, and distinguished a 'preference-fixed' approach, where treatment allocation is determined solely by a patient's clinical characteristics, from a 'preference-flexible' approach, where their values are used to determine whether the benefit they would derive from a treatment is sufficient to justify its cost. As noted above, the preference-flexible management strategy would have to be compared with a preference-fixed form of management (TCRE only or AH only are the terms used here) in terms of costs and benefits.

Nease and Owens explained their methods using a four-step process. Their illustrative clinical area was the management of mild hypertension, but the steps can also be explained using the example of surgical management of menorrhagia. Step 1 involves the identification of a cohort of menorrhagia patients, and the elicitation of their health state values. Step 2 is the calculation of the expected QALYs and costs of the alternative preference-fixed treatment allocations (AH only and TCRE only). These preference-fixed options are assessed by incorporating the health state values of all individuals into a decision analytic model for each therapy and aggregating the total QALYs and cost assuming everyone in the cohort receives that therapy[1].

Step 3 is the estimation of the expected costs and benefits of the preference-flexible management strategy. This involves calculating the expected QALYs of the treatment options using the health state values of each patient in the cohort. Each patient would be allocated to a given treatment if one of the following applied: (a) their expected QALYs for that therapy, implied by their health state values, are higher than for the other comparators, and the cost of that treatment is no higher than that of the others under consideration; (b) if the cost of a patient having the intervention from which they would enjoy the highest expected benefit is higher than for one of the comparators, but their individualised cost-utility ratio is less than $50,000 per additional QALY.

The expected costs and QALYs associated with the preference-flexible approach are then calculated by adding up the expected costs and QALYs of each patient in the cohort according to the treatment to which they were allocated. In Nease and Owen's work, this calculation included the cost of eliciting the health state values from patients, which was assumed to be $100 per patient. Finally, in Step 4, the expected costs and QALYs of the preference-fixed (AH only or TCRE

---

[1]This is an unusual approach to estimating the costs and benefits of an all or nothing strategy for two or more treatments using a model. Usually the mean values provided by the sample of patients for each health state would be incorporated into the model.

only) and preference-flexible strategies are compared to establish relative cost-effectiveness

This approach to incorporating patients' preferences into an economic evaluation would also need to decide upon the incremental cost per QALY threshold, for a given intervention, above which the individual patient would not be allocated to a therapy, but instead would undergo the cheaper alternative which their health state values suggest will generate fewer benefits. Also, as for the other two approaches, the individual cost-effectiveness approach could employ *ex ante* HYEs as a measure of value. This would again have the advantage of requiring fewer assumptions to link the values to individuals' treatment-related choices.

This section has outlined how three possible preference-driven management strategies might be evaluated using modelling techniques: treatment allocation based on choice, patients' values and patient-specific CEA. The next three sections of the chapter apply these three alternative methods to the assessment of AH versus TCRE. Using the data presented in Chapters 5 and 7, these sections consider a series of alternative additional comparators to TCRE and AH alone, where treatment allocation is decided at the level of the individual patient based on their preferences. The objective of this empirical analysis is to assess the potential cost-effectiveness of moving towards ways of managing menorrhagia that explicitly take account of women's preferences.

8.4   Surgical management of menorrhagia: treatment allocation by patient choice

This first empirical section compares the costs and benefits of TCRE only, AH only and a management strategy where women choose which therapy they would like.

8.4.1 Methods

The methods employed are similar to those described in general terms in Section 8.3.1. Two alternative measures of benefit are used: expected QALYs and *ex*

292

ante HYEs. The expected costs are based on those reported in Chapter 5, and
expected benefits of TCRE only and of AH only are taken directly from Chapter
5 (for the QALY analysis) and from Chapter 7 (for the *ex ante* HYE analysis), and
are based on extrapolation over a woman's expected lifetime. The expected
cost and benefit of the strategy of treatment allocation by choice are modelled
based on the patient-specific valuation data and treatment choice information
collected from the cohorts of women with menorrhagia recruited in Bristol (for
health state values) and Swindon (for *ex ante* HYEs), and which were also
detailed in Chapters 5 and 7, respectively. The various steps used in this
analysis are detailed below.

(a)   As detailed in Chapter 4, each woman in both the Bristol and the Swindon
        cohorts was asked to complete a questionnaire which sought information
        on their attitudes to treatments and to treatment characteristics. In
        Section 4 of the questionnaire, women were presented with two
        descriptions, one of TCRE and the other of AH. These details included
        information on some aspects of the process of care as well as the
        outcomes, and included available prognostic information throughout the
        woman's remaining lifetime. Women were asked to indicate which of
        these two treatments they would choose to undergo, but were given the
        opportunity to respond that they were unsure which treatment to choose
        or that they would not choose either of the treatments. For the current
        analysis, a woman is assumed to have chosen between TCRE and AH on
        the basis on her answer to this question. Women who answered that
        they would not undergo either therapy are excluded from the analysis.
        Two alternative analyses are undertaken to deal with women who said
        they were unsure which treatment to choose: in one they are assumed to
        have a TCRE and in the other they are assumed to undergo AH.

(b)   Depending on her treatment allocation, the cost of a woman's therapy is
        taken as the expected lifetime cost of TCRE or of AH *as detailed in*
        Chapter 5.

(c)   For the QALY analysis, the expected lifetime benefit following the therapy

        is calculated by incorporating the health state values of each woman in

        the Bristol cohort into the cost-utility model described in Chapter 5.

        Depending on the treatment she chooses, the benefit for each woman is

        taken as the expected QALYs associated with TCRE or AH, based on her

        own health state values.

(d)   For the *ex ante* HYE analysis, the estimate of lifetime benefit is based on

        the *ex ante* HYEs elicited from women in the Swindon cohort. Depending

        on which treatment a woman chooses, her *ex ante* HYE is that of AH or

        TCRE, again based on her own values.

(e)   The cost of the treatment allocation by choice strategy is taken as the

        average of the cost of treating the women according to the her chosen

        therapy. In the base-case analysis, it is assumed that the process of

        informing women and of eliciting their choice of treatment imposes no

        additional cost.

(f)   Similarly, the expected benefit of the choice-based strategy is estimated

        by averaging the individual expected QALYs for the Bristol cohort and the

        individual *ex ante* HYEs for the Swindon cohort, according to each

        woman's treatment choice.

8.4.2 Results

        Consistency between patient-specific benefits and their treatment choices

in the questionnaire. Of the 60 women interviewed in Bristol to elicit health

state values, 59 responded to the item in the questionnaire which asked them to

choose between AH and TCRE on the basis of the information provided. Of

these, 29 (49%) chose AH, 21 (36%) chose TCRE, 2 (3%) indicated that they

would choose neither and 7 (12%) said they were unsure which one to choose.
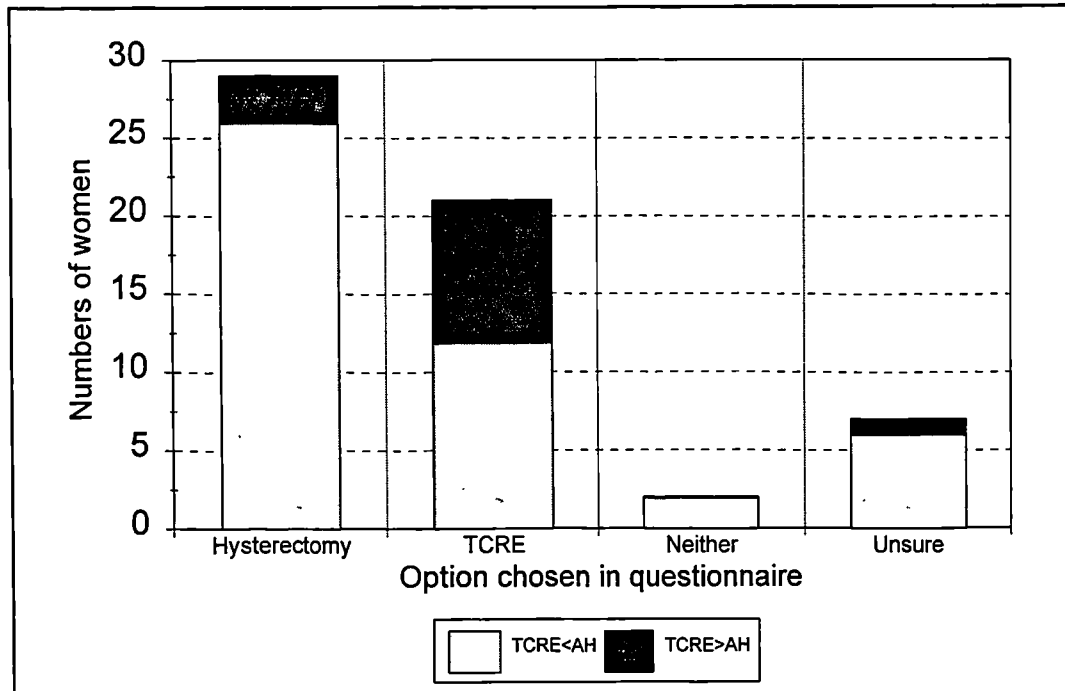
**Figure 8.2   The consistency between women's treatment choices in the
questionnaire and the treatment with the higher expected QALYs based
on incorporating women's individual health state values into the cost-
utility model**

When the health state values elicited from these women are incorporated into

the model, it is possible to assess the consistency between women's choice of

treatment in the questionnaire and the therapy with the highest expected QALYs

on the basis of their particular health state values. Figure 8.2 shows this

assessment of consistency. The majority of women (26/29) who chose AH in

the questionnaire also have higher patient-specific QALYs for AH than for TCRE.

However, this level of consistency is not maintained for women who chose TCRE

in the questionnaire, as 12/19 of these women had higher patient-specific

QALYs for AH than for TCRE. This finding is probably due to the fact that, as

reviewed in Chapter 7, QALYs do not predict accurately individual choice; in

other words, women do not make decisions using the same 'model' as is being

employed for the CUA. For example, if a women chose TCRE in the

questionnaire because the duration and severity of convalescence were very

important to her, this is unlikely to be reflected in her QALYs because

convalescence gets 'drowned out' in the QALY model as it is only a very small

part of the 44 year time horizon. Overall, there was clear inconsistency between women's choice in the questionnaire and the treatment with the higher patient-specific QALYs in 15/59 women.

Of the cohort of 63 women from whom *ex ante* HYEs were elicited in Swindon, 58 responded to the item in the questionnaire about choosing between AH and TCRE on the basis of the descriptions provided. Of these, 24 (41%) chose AH, 24 (41%) chose TCRE, 2 (4%) said that they would choose neither and 8 (14%) said they were unable to choose. Figure 8.3 shows the consistency between these treatment choices and the treatment with the highest individual *ex ante* HYEs. Of those women choosing AH in the questionnaire, 17/24 provided a higher *ex ante* HYE for that therapy than for TCRE. Of the women choosing TCRE, however, 11/24 provided *ex ante* HYEs that were equal for TCRE and AH, and 7/24 gave higher *ex ante* HYEs for AH than for TCRE.

The equality of HYEs in women choosing TCRE could be explained by that fact that, although women might choose TCRE before AH, their strength of preference is not sufficient to register in terms of HYEs. However, the 29% of women who 'chose' TCRE but registered a higher HYE value for AH is a source of inconsistency that is less easily explained. One possibility is that the descriptors used to elicit HYEs were slightly different to those used in the questionnaire; another explanation could be that women's views altered between completing the questionnaire and being interviewed. Overall, there was clear inconsistency between women's choice in the questionnaire and the treatment with the higher patient-specific *ex ante* HYEs in 12/58 women, a slightly lower rate than for the QALY-based analysis.

The cost and benefit of choice-based treatment allocation. Table 8.1 shows the costs and benefits of the management strategy of treatment allocation by choice, compared to TCRE only and AH only. Analyses using expected QALYs and *ex ante* HYEs are presented and, in both cases, the implications are considered of alternative assumptions about the treatment
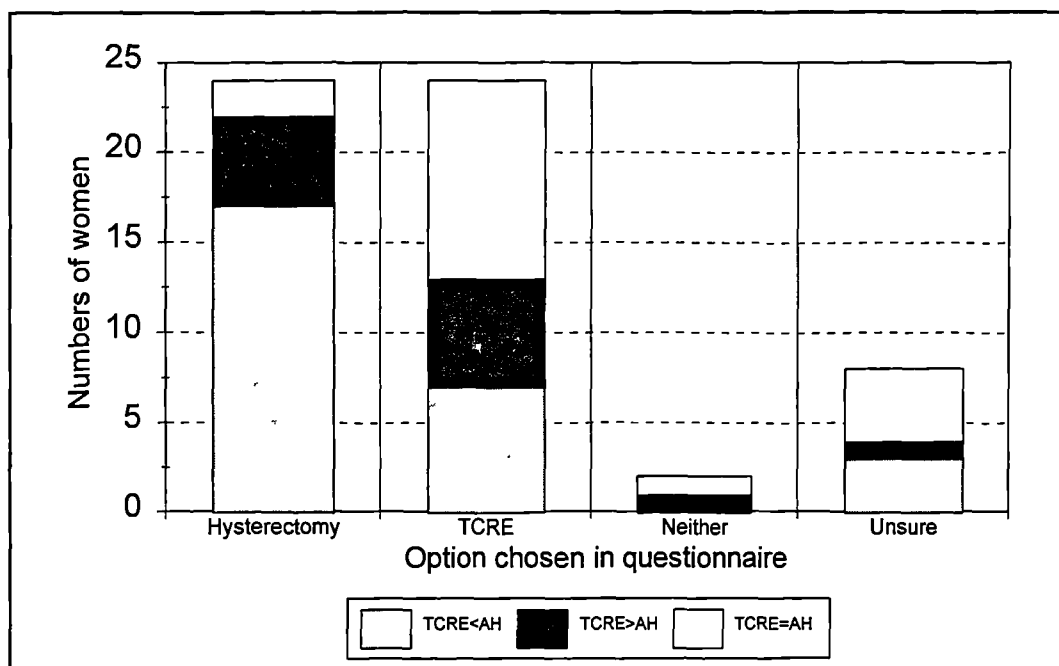
Figure 8.3   The consistency between women's treatment choices in the questionnaire and the treatment with the higher benefit based on their individual *ex ante* HYE values

allocated to women who are unable to choose between AH and TCRE. The choice-based strategy obviously has a mean expected cost per patient lying somewhere between the expected cost of TCRE only and that of AH only, because there will be a mix of the two treatments chosen.

When benefits are measured in terms of expected QALYs using the cohort of women from Bristol, the mean benefit per patient is less than that of AH only, whatever the treatment allocation for women who are unable to choose between the two therapies. In Chapter 5 two illustrative incremental cost per QALY thresholds were suggested to guide decision making using CUA. If these threshold ratios are considered acceptable, the incremental costs per QALY in Table 8.1 suggest that a patient choice strategy, with women who are unable to choose being allocated to AH, would probably be considered better value for money than TCRE only or patient choice with TCRE given to women who are unable to choose. However, with an incremental cost per additional QALY of

Table 8.1   **Lifetime expected costs and benefits of TCRE only, AH only and a management strategy where women are allocated to treatment on the basis of their choice**

Management Strategy	Expected cost per patient (£)	Benefit per patient	Incremental cost pe additional unit of benefit (£)
Using expected (standard) QALYs			
TCRE only	816	14.413	-
Allocation by choice:			
- Given TCRE if unable to choose	992	15.038	282
- Given AH if unable to choose	1035	15.142	413
AH only	1162	15.195	2396
Using *ex ante* HYEs			
TCRE only	816	32.62	-
Allocation by choice:			
- Given TCRE if unable to choose	964	35.06	146
- Given AH if unable to choose	1014	35.24	278
AH only	1162	34.84	Dominated

£2396 over patient choice, AH only is likely to be viewed as the most cost-effective of the comparators if the illustrative thresholds are acceptable.

This conclusion is not replicated when benefits are considered in terms of *ex ante* HYEs, as both choice-based strategies have a higher mean benefit than AH only, and a lower expected cost; in other words, the choice-based strategies dominate AH only. On the basis of the incremental costs per additional HYE shown in Table 8.2, the use of choice, with AH given to those women unable to

choose, would probably be considered the most cost-effective form of management.

This analysis has not included the cost of presenting women with information and eliciting their choice of therapy. If the information element of this process took the form of a video and/or booklet, the resource cost would mainly be fixed and, once spread out across thousands of women, would be quite modest. If a Grade F nurse were used to undertake a semi-structured interview lasting about 15 minutes, the cost would be about £2.50 per patient. Even allowing for hospital overheads, the cost of information and choice elicitation is unlikely to alter the conclusions reached above.

8.5   Surgical management of menorrhagia: treatment allocation by patients' values

This second empirical section compares the costs and benefits TCRE only, AH only and a management strategy where women are allocated to one of these two treatments according to which has the higher patient-specific benefit, where benefit is measured in terms of expected QALYs or *ex ante* HYEs. This differs from the first analysis in that women's treatment-related *choices* play no part in treatment selection, which is based only their values. The premise of this management strategy is that women's values will be a reasonable, albeit imperfect, predictor of their preferred treatment.

8.5.1 Methods

The methods used for this analysis are very similar to those used in the choice analysis, and these are detailed below.

(a)   Again, two analyses are presented: one using expected lifetime QALYs as the means of treatment allocation and of benefit measurement, and the other using lifetime *ex ante* HYEs.

(b)   For the values based strategy, women are allocated to AH or TCRE using the following decision rules. For the expected QALYs analysis, the health state values of each woman in the Bristol interview cohort are independently entered into the cost-utility model. Treatment allocation for a given woman depends on which treatment provides the higher number of expected QALYs. For the *ex ante* HYE analysis, the same principle applies: a given woman in the Swindon interview cohort is allocated between AH and TCRE on the basis of which treatment she valued more highly in terms of *ex ante* HYEs. Women who valued AH and TCRE equally in terms of *ex ante* HYEs are assumed to be allocated to TCRE which would dominate AH for these women (ie. TCRE is valued equally with AH, but AH is more costly).

(c)   The estimated benefit of the values-based treatment allocation strategy is the mean expected lifetime QALYs or *ex ante* HYEs for the group as a whole, where each woman has been allocated to AH or TCRE as described in (b).

(d)   The expected lifetime cost of the values-based strategy is calculated in the same way as for the choice-based allocation strategy. In the base-case analysis, the process of eliciting values is assumed to impose no additional cost.

8.5.2 Results

Figure 8.4 shows the distributions of expected lifetime QALYs for AH and TCRE resulting from incorporating the health state values of each woman in the Bristol cohort separately. The figure shows the greater concentration of expected QALYs at the top of the range for AH compared to TCRE which, of course, explains the higher mean expected QALYs with AH detailed in Chapter 5. Also reflecting this, if women are allocated to one of the two treatments depending on which one generates the higher number of QALYs using women's individual health state values, 77% would be allocated to AH and 23% to TCRE.
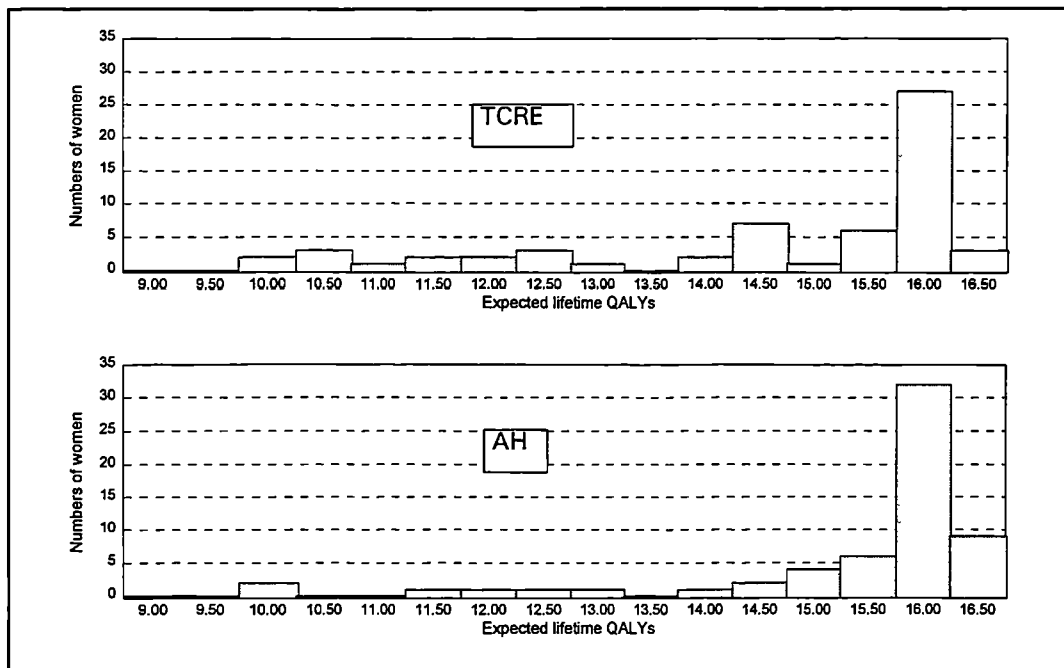
Figure 8.4   Distribution of expected lifetime QALYs based on the individual health state values of women in the Bristol interview cohort

Figure 7.3 in the previous chapter showed the distribution of *ex ante* HYEs elicited from women in the Swindon interview cohort. If treatment allocation is on the basis of the higher *ex ante* HYE value, only 43% would be allocated to AH and 57% to TCRE. However, these figures are partly a product of the assumption that the 29% of women who valued AH and TCRE equally in terms of *ex ante* HYEs are allocated to TCRE because it is the cheaper option.

Table 8.2 shows that, if the values elicited from the Bristol and Swindon cohorts are considered representative of women with menorrhagia generally, a management strategy of treatment allocation on the basis of women's values would almost certainly be considered cost-effective relative to AH only or to TCRE only. On the basis of both expected QALYs and *ex ante* HYEs, the values-based management strategy would dominate AH only, generating greater benefit and costing less per patient. Compared to TCRE, treatment allocation on the basis of values would cost more per patient but generate more benefit, with each extra unit of benefit costing a relatively modest amount - if illustrative cost per

Table 8.2   **Expected costs and benefits of TCRE only, AH only and a management strategy where women are allocated to a treatment on the basis of their health state values or *ex ante* HYEs**

Management Strategy	Expected cost per patient (£)	Benefit per patient	Incremental cost per additional unit of benefit (£)
Using expected (standard) QALYs			
TCRE only	816	14.413	-
Allocation by values	1,081	15.275	307
AH only	1162	15.195	Dominated
Using *ex ante* HYEs			
TCRE only	816	32.62	-
Allocation by values	964	36.63	37
AH only	1162	34.84	Dominated

QALY thresholds suggested in Chapter 5 are accepted, a values-based treatment allocation strategy would be considered more cost-effective than TCRE only.

This analysis does not include the cost of eliciting women's values. If a Grade F nurse were used to undertake a semi-structured interview lasting about 45 minutes, the cost would be about £7.50 per patient - more than for the choice-based strategy as values are likely to take longer to elicit that choices. Even allowing for hospital overheads, the cost of eliciting values is unlikely to alter the conclusions of the analysis.

8.6   Surgical management of menorrhagia: treatment allocation using patient-specific CEA

This third empirical section compares the costs and benefits TCRE only, AH only and a management strategy where women are allocated to one of these

treatments according to patient-specific CEA. This analysis differs from the last one due to the fact that cost as well as benefit are considered when allocating an individual woman to a treatment. Hence, it is not enough for a woman to benefit (on the basis of her health state values or *ex ante* HYEs) more with one therapy than the other; the additional benefits have to be sufficient to justify any additional cost at an individual level.

8.6.1 Methods

The methods used for this analysis are similar to the last, with two analyses being undertaken, one using expected lifetime QALYs and the other lifetime *ex ante* HYEs. The only difference is the way in which women are allocated to AH or TCRE. Taking the expected QALYs approach first, as before, the health state values of each woman in the Bristol cohort are incorporated individually into the cost-utility model. The expected QALYs generated for AH and TCRE are compared with the average cost of delivering those two therapies, and a woman is allocated to one of them based on the following decision rules.

(a)   If a woman's expected lifetime QALYs (based on her specific health state values) are greater for TCRE than for AH, she is allocated to TCRE because, in terms of her personal analysis, TCRE dominates AH, being less costly and generating more benefit.

(b)   If a woman's expected lifetime QALYs are higher for AH than for TCRE, she will be allocated to AH as long as her personal cost-utility ratio (additional expected cost of AH divided by the additional personal expected QALYs with AH) is lower than the threshold ratio. Alternative analyses are undertaken using the upper (£33,000 per additional QALY) and lower (£6,500 per additional QALY) illustrative thresholds adopted in Chapter 5.

Table 8.3   **Proportions of women allocated to TCRE and AH at different critical ratio threshold ratios using patient-specific cost-effectiveness analysis**

Threshold ratio (£)	% of women allocated to:	
	AH	TCRE
Using expected (standard) QALYs		
£6,500	58	42
£33,000	75	25
Using *ex ante* HYEs		
100	36	64
80	25	75
60	22	78
40	16	84
20	6	94

The methods for the *ex ante* HYE analysis are similar. Each woman's personal *ex ante* HYEs for AH and TCRE are considered, and she is allocated to AH or TCRE using the following decision rules.

(a)   If a woman's *ex ante* HYEs for TCRE are greater than *or equal* to those for AH, she is allocated to TCRE because, in terms of her personal analysis, TCRE dominates AH.

(b)   If a woman's *ex ante* HYEs are higher for AH than for TCRE, she will be allocated to AH as long as her personal cost-utility ratio is lower than a threshold ratio. Given that *ex ante* HYEs have yet to be used widely in applied economic evaluation, it is not clear what values would be attached to cost per unit of benefit thresholds above which treatment would not be considered cost-effective. However, for the purposes of this analysis, alternative analyses are reported using five illustrative thresholds.

8.6.2 Results

Table 8.3 shows the proportions of women, under alternative threshold ratios, who would be allocated to TCRE and AH using a management strategy based on patient-specific CEA. Clearly, as the threshold ratio increases, more women would be allocated to AH.

Table 8.4 shows the expected costs and benefits of a management strategy based on patient-specific CEA, in comparison with TCRE only and AH only. As with the management strategy based on women's health state values, a policy of AH only is dominated on the basis of expected QALYs and of *ex ante* HYEs. Furthermore, using expected QALYs, patient-specific CEA would be considered better value for money than TCRE only if the lower illustrative threshold ratio of £6,500 is accepted, with an incremental cost per additional QALY of only £236. If the upper threshold of £33,000 were accepted as the threshold above which women would be allocated to TCRE, this would be more costly and generate more expected QALYs, which each additional QALY costing an additional £9,667 over patient-specific CEA with the lower threshold. A similar position prevails when *ex ante* HYEs are used as the benefit measure. Although it is difficult to suggest what might be a generally acceptable cost per HYE threshold, the incremental cost per additional HYE of moving from a TCRE only to a patient-specific CEA (with a £20 threshold ratio) would only be £13. As with the other strategies analysed above, the cost of eliciting the values from women is unlikely to alter the position shown in Table 8.4 markedly.

8.7   Discussion

8.7.1 The concept of preference-based treatment allocation

To date, economic evaluation has concentrated on identifying a single cost-effective intervention from amongst those under comparison. In those clinical areas where there are clear trade-offs between aspects of the process and outcomes of treatment, strict adherence to an all or nothing approach by

Table 8.4   **Lifetime expected costs and benefits of TCRE only, AH only and a management strategy where women are allocated to treatment on the basis of patient-specific cost-effectiveness analysis**

Management Strategy	Expected cost per patient (£)	Benefit per patient	Incremental cost per additional unit of benefit (£)
Using expected (standard) QALYs			
TCRE only	816	14.413	-
Allocation by patient-specific CEA with threshold ratios of:			
£6,500	1018	15.269	236
£33,000	1076	15.275	9,667
AH only	1162	15.195	Dominated
Using *ex ante* HYEs			
TCRE only	816	32.62	-
Allocation by patient-specific CEA with threshold ratios of:			
£20	838	34.28	13
£40	871	35.39	30
£60	893	35.87	46
£80	904	36.02	73
£100	942	36.47	84
AH only	1162	34.84	Dominated

purchasers will lead to a proportion of patients receiving interventions that are inconsistent with their preferences. Just as it is an established part of clinical (and economic evaluation) to explore management strategies which allocate patients to therapies on the basis of clinical characteristics which may influence outcomes, there is surely a need to consider ways in which the impact of patients' preferences on the benefits they derive from treatment can be reflected in treatment allocation. Whatever the form of preference-based treatment

allocation, the principal aim is to maximise benefits just like any other economic evaluation. However, this approach to patient management explicitly accepts that, given appropriate information, patients may be the best judges, on an *ex ante* basis, of the benefits they are likely to experience from treatment. This judgement could be direct, through choice-based allocation, or indirect, through a values-based approach.

MAS is a prime example of a group of technologies where there is likely to be a trade-off in the process and outcomes of care relative to conventional surgery, where patients are likely to have clear attitudes to these trade-offs if made aware of them and, therefore, where the role of preferences in patient management should be carefully assessed. Surgical treatment for menorrhagia is a good, but by no means the only, example of where these methods could be used. For example, the trade-offs between the process and outcomes of percutaneous transluminal coronary angioplasty (PTCA) and coronary artery bypass surgery (CABG) are perhaps even more stark, and suggest a possible role for preference-based treatment allocation in that clinical area [Hlatky 1995].

In the specific context of surgical treatment for menorrhagia, the foregoing analysis has indicated that a greater use of patients' preferences in making treatment decisions may prove a cost-effective alternative to reliance on one therapy for all patients. If such strategies are to be used in practice, decisions will be required about which of the three approaches considered here is likely to be the most cost-effective. Table 8.5 compares the costs and benefits of all three preference-based forms of management and of AH only and TCRE only. Given that the pattern of relative costs and benefits is broadly similar whether expected QALYs or *ex ante* HYEs are used as the benefit measure, the table presents data relating to expected QALYs only, where it is easier to judge what incremental costs might be considered worth paying to generate additional benefits.

Table 8.5   **Lifetime expected costs and QALYs of TCRE only, AH only and the three alternative preference-based management strategies**

Management Strategy	Expected costper patient (£)	Benefit per patient	Incremental cost additional unit of benefit (£)*
TCRE only	816	14.413	-
Preference-based allocation by:			
- choice (given TCRE if unable to choose)	992	15.038	ED
- personal CEA (threshold ratio = £6,500)	1018	15.269	236
- choice (given AH if unable to choose)	1035	15.142	D
- personal CEA (threshold ratio = £33,000)	1076	15.275	9667
- Values	1081	15.275	D
AH only	1162	15.195	D

D   =   Subject to domination (more costly, no more effective than next least costly option)

ED   =   Subject to extended dominance (higher incremental ratio than that relating to a more costly and more effective option).

*   =   Relative to the next least costly option not subject to dominance or extended dominance

The table ranks all options under consideration in ascending order of expected cost. An incremental cost per additional QALY is then calculated relative to the next least costly non-dominated option. Coming down the table, options that are more costly and no more effective than earlier ones are subject to dominance (they are unequivocally inferior in economic terms). If an option has a higher incremental ratio than a more costly and more effective option, it is subject to extended dominance and can also be rejected [Cantor, 1994]. On that basis, patient-specific CEA would seem to be the most cost-effective preference-based

strategy: it dominates the choice-based and values-based strategies and AH only, and has only a modest incremental cost per additional QALY over TCRE only. This conclusion is not surprising as the strategy goes furthest towards 'fine-tuning' treatment allocation to patient benefit but with a consideration of cost at the level of the individual.

There are some important caveats, however, associated with this conclusion. Firstly, Chapters 5 and 7 of this report detailed the uncertainty which exists in key parameters of the CUA of TCRE versus AH. In particular, the health state values elicited from the Bristol cohort of women and the *ex ante* HYEs provided by the women interviewed in Swindon showed considerable variability. Although the conclusions of the QALY-based analysis in Chapter 5 were found to be robust to individual variation in these parameters, the variability in the *ex ante* HYE values limited the firmness of the conclusions of the alternative CUA in Chapter 7. Indeed, the sensitivity of the relative cost-effectiveness of the two surgical options to women's values was a major reason for assessing the costs and benefits of preference-based management strategies. A key source of uncertainty associated with cost and benefit estimates presented in this chapter is the relatively small size of sample of women from whom values and choices were elicited. For this reason, the absolute value of the incremental cost and benefit estimates related to preference-based management should be considered tentative. However, the general finding that expected benefits can be increased by tailoring treatments to women's preferences, and that this has the potential also to be cost-effective, is important and, although in need of confirmation with further research, is likely to be robust to the various sources of uncertainty.

A second and related caveat is that the generalisability of the values and treatment choices elicited from the women in Bristol and Swindon needs to be considered. The fact that the cohorts were drawn from just two centres would suggest that these samples may not be representative of women with menorrhagia in the country as a whole. Again, although the broad conclusion

remains valid, definitive estimates of cost-effectiveness should await further research.

A third caveat associated with the conclusions of the analysis is that women's choices in the study were based on restricted information. A management strategy which allows women to select their therapy would only be feasible if sufficient detailed and comprehensible information on the alternative therapies were provided to women in an accessible form. In the analysis presented here, women's choices were based on their responses to an item in the questionnaire sent to them shortly after referral which provided limited information on the two forms of surgery.

A fourth caveat concerns the scope of the benefits considered in the model. A preference-based approach to the management of menorrhagia could include two important benefits which are not typical outcomes from care: information provision and choice. These can be viewed as process-related benefits [Ryan and Shackley, 1995], or as specific forms of outcome [Dowie, 1993]. Either way, although the use of *ex ante* HYEs in Chapter 7 allowed patients' values in relation to the process of treatment to be reflected in the measure of benefit, the combination of a modelling approach and the type of descriptive scenarios chosen for the valuation exercise has effectively excluded patients' values regarding information and choice. Indeed, it is not clear whether these consequences of preference-based management would be benefits or disbenefits: evidence indicates that, whilst most patients value information, many prefer not be to involved with decision making and may derive disbenefits from being obliged to make choices [Cassileth *et al*, 1980; Beisecker *et al*, 1990].

A final caveat relates to the cost of preference-based treatment allocation. If a health care provider has to be able to provide two or more treatments when, on the basis of all or nothing economic evaluation, they would have offered only one, there may be cost implications not allowed for in the analysis here. For example, if the procedures involve the use of expensive equipment, the use of

preference-based strategies would probably result in fewer patients undergoing any given treatment, resulting in higher equipment costs per patient as these fixed costs are spread over fewer procedures.

8.7.2 The practicalities of preference-based management

If the realities of using the three preference-based strategies in routine practice are considered, the patient-specific CEA may be considered the least practical of the strategies, although the analysis presented here suggests it is likely to be the most cost-effective. There may be logistical difficulties in eliciting values for each woman presenting for surgical treatment for menorrhagia (a problem shared by the treatment allocation by values strategy), but the development of valuation tools such as interactive videos and computers could address some of these difficulties [Nease *et al*, 1996]. It is possible that patients will understand that the higher the values they give for a specific treatment or health state, the more likely they are to receive the treatment they desire, so a problem of 'gaming' may exist. Perhaps a more fundamental problem for some would be the equity implications of patient-specific CEA, in that some patients would not receive an intervention despite being *clinically* identical to other patients who would undergo the treatment: in this context the health care system would be saying that a woman would derive clinical improvement from a procedure but they would not value that improvement sufficiently to justify the cost. However, if it is accepted that patients' benefits from a given intervention are related to their values as well as to expected clinical outcomes, and that resource allocation should consider patients' capacity to benefit, it follows that values and clinical characteristics should be assessed at the individual level, as well as the level of the group as with standard CUA.

Given that the use of choice-based management avoids the need to elicit values from patients and is probably less controversial from the viewpoint of its equity implications, it might be considered more feasible for routine practice. Although the analysis here suggests it may not be as cost-effective as AH alone in terms of QALYs, it is also found to dominate AH in terms of *ex ante* HYEs.

Furthermore, this balance might change once 'process-related' benefits are incorporated, such as the value patients attach to information and choice *per se*. The major challenge for all forms of preference-based treatment allocation is the need to provide accurate, comprehensive and accessible information to women. Whereas a strategy based on women's values to estimate individualised QALYs relates information to a limited number of health states and allows the model to indicate the treatment with the highest expected benefit, choice-based management requires that the patient is able to come to a decision themselves about their preferred therapy on the basis of any information they care to use. For this sort of approach to work, there needs to be more research into the use of novel information tools such as interactive videos and computers [Barry *et al*, 1995; Shepperd *et al*, 1995].

It is important to consider the potential conflicts which may arise between clinical judgements about specific patients and preference-based treatment allocation. The assumption that underlies the analysis in this chapter is that the women in the Bristol and Swindon cohorts are clinically homogenous and differ solely in terms of their preferences. Clearly, however, clinicians will form opinions about the clinical appropriateness of treatments for specific women which may or may not be based on good evidence. The use of preference-based treatment allocation is likely to work best when the clinician is satisfied that AH and TCRE are both clinically feasible and acceptable. Clearly, it will not be possible to use the strategy when a woman has an *absolute* clinical contra-indication for one of the treatments. However, conflicts may arise when a woman is considered to have a *relative* clinical contra-indication for a particular therapy. Generally, a clinician would probably rule out the therapy in those circumstances. However, under preference-based management strategies, the woman should ideally be made aware of the contra-indications of therapy and their implications for her prognosis, and be allowed to make her choices or provide her values accordingly.

8.8 Conclusions

This chapter has modelled the costs and benefits of three preference-based forms of management, and has shown, in the context of AH versus TCRE, that each of them has the potential to be more cost-effective than TCRE alone, and that values-based and patient-specific CEA-based approaches could be more cost-effective than both TCRE and AH alone. Although the empirical estimates presented in this chapter should be interpreted with caution, the conclusion that preference-based management has the potential to be more cost-effective than the conventional all or nothing strategy is an important one for the economic analysis of MAS. Given the trade-offs in the outcomes between many forms of MAS and open surgery, it may be the case that expecting all patients to undergo the same therapy, whatever their choices or values, is unrealistic. It is likely, therefore, that management based on patients' preferences will play a greater role in these contexts.

Chapter 9

Conclusions

9.1 Introduction

The purpose of this thesis has been to explore some of the methodological and empirical issues related to the economic evaluation of minimal access surgery MAS using the particular example of the comparison of abdominal hysterectomy (AH) and transcervical endometrial resection (TCRE) for the treatment of menorrhagia. The economic characteristics of MAS and the major increase in health service expenditure on this group of interventions which projections suggest will take place, emphasise the importance of further development in the methods used in economic evaluation. However, these methods have relevance to the economic evaluation of all health care technologies.

The starting point of this analysis was a consideration of trial-based economic evaluation of MAS. Like all economic evaluations, this form of analysis introduces a range of different uncertainties relating to data inputs, generalisability, extrapolation and analysis [Briggs *et al*, 1994]. The subsequent chapters of the thesis focused on issues of method associated with illuminating particular areas of uncertainty related to benefit measurement, generalisability and incorporation of patients' preferences into treatment allocation.

The purpose of this final chapter is to pull together and summarise the main methodological and empirical findings of the six main chapters, to draw some broad conclusions about the economic evaluation of MAS and to indicate the contribution of the thesis to the methods of economic evaluation.

9.2   Trial based evidence

9.2.1 Methodological analysis

The economic evaluation of many applications of MAS will begin at a much earlier stage than the randomised controlled trial (RCT). Ideally, analysis should begin when new MAS technologies are to be used in a few centres by enthusiasts, using available data and modelling techniques - Stage I analysis in the language of iterative economic analysis [Sculpher *et al*, forthcoming]. In the case of the evaluation of TCRE relative to AH, however, no economic analysis had been undertaken prior to the start of the Bristol trial which, therefore, became a valuable source of data with which to begin the economic evaluation.

Chapter 3 presented the economic analysis of AH and TCRE undertaken alongside the Bristol RCT. The analysis represents the foundation on which the subsequent analyses in the thesis is built. The main purpose of this element of work was to identify the limitations of trial-based economic evaluation in the particular context of the evaluation of MAS, and hence to act as a rationale for the methodological developments described in later chapters. Many of the methodological issues highlighted by the analysis in Chapter 3, therefore, have

been identified in earlier studies. However, it remains an important contribution of the thesis to have indicated which shortcomings of economic analysis alongside trials are particularly problematic for the evaluation of MAS applications, and these are detailed below.

■   A trial set up to evaluate a new MAS application from a clinical perspective is likely to exhibit weaknesses in terms of its usefulness in facilitating economic conclusions. These may include a failure to collect all the items of data which are important for economic analysis; and too small a sample size to generate reliable estimates of those parameters central to the economic comparison, a problem often compounded by the difficulty in recruiting patients into trials involving MAS techniques. However, the value of RCT data for economic analysis is its high level of internal validity. This can provide a valuable empirical platform upon which to assess relative cost-effectiveness, but this form of analysis cannot be seen as the final word in the economic evaluation of a new form of MAS.

■   A major limitation of data from an RCT is the limited external validity they often exhibit. Most of the RCTs undertaken to compare MAS with conventional surgery have been undertaken within a small number of specialist centres using clinical practice which is atypical of that routinely used in the majority of centres. The trial used as a starting point for the economic analysis of AH versus TCRE was an extreme example of this, taking place in a single medical school, with both forms of surgery undertaken by a single surgeon using specific versions of the operations. The external validity of trials can be increased by incorporating a greater mix of centres, clinicians and techniques, but limitations to generalisability will remain. Furthermore, trials of MAS tend to take many months, often years, to report their results, during which time the technologies have moved on in the form of minor or substantive modifications. The uncertainty in terms of generalisability that this generates needs to be addressed using other data sources.

316

■   Trial-based economic analysis of MAS is often limited by the outcome
    data collected. Many trials are set up principally to answer clinical
    questions, and they tend to collect a range of outcome data, most of
    which are not patient-based. It is frequently the case that MAS
    applications have a range of different outcomes which, in comparison
    with those of conventional surgery, tend to go in different directions.
    This situation is made more complex by differences in process
    characteristics. Even when trials have been set up with an explicit aim of
    informing economic analysis of MAS, it is usually difficult to identify a
    single, all-embracing and uni-dimensional measure of outcome suitable for
    cost-effectiveness analysis (CEA) of this group of technologies. The
    patient-based measures of outcome have usually been descriptive health-
    related quality of life (HRQL) profiles which do not provide a single score.
    A number of RCTs of MAS procedures have been powered on patient
    satisfaction rates. In relation to AH and TCRE, Chapter 3 showed that
    dichotomising satisfaction offers a means of undertaking CEA in this area.
    However, differences in measuring satisfaction between studies and
    uncertainties about measurement techniques make this an imperfect
    measure of outcome upon which to base CEA.

■   Even if 'natural' measures of outcome suitable for CEA can be identified
    in RCTs, at best they will be able to inform resource allocation within
    programmes and disease areas. Furthermore, CEA does not provide a
    clear means of factoring individuals' (patients' or others') preferences into
    the economic analysis. If, from the outset, a trial has been set up to
    explore economic issues, there is now scope to collect data necessary to
    develop a cost-utility analysis (CUA) and to inform system-wide resource
    allocation. Valuation systems such as the EuroQol instrument allow
    patients to be 'described' over time using a generic classification linked to
    a tariff of health state values. In the economic analysis of AH versus
    TCRE, the trial was planned from the perspective of clinical evaluation,
    and no such data were collected. In certain contexts, the usefulness of

the valuation system approach may be limited by too few data collection time points being feasible within a trial and by the limited sensitivity of the instruments. It will, therefore, frequently be necessary to collect data for CUA subsequent to a RCT.

9.2.2 Empirical analysis

From this trial-based empirical work, it is possible to reach a number of conclusions about the economic comparison of AH and TCRE as an example of MAS compared to conventional surgery.

- On the basis of the outcome data collected in the trial, it is not possible to select either therapy as unequivocally more effective. On the side of hysterectomy is the once-and-for-all end to heavy blood loss, a greater improvement in other menstrual symptoms, some evidence of a greater improvement in HRQL and higher rates of satisfaction with treatment on the part of women. In favour of TCRE is the shorter convalescence and the fact that menstrual symptoms and HRQL do improve, albeit not to the extent experienced by women undergoing hysterectomy. The failure rate of TCRE (between 23% and 25%), manifesting itself in terms of repeat surgery, is a major shortcoming of the procedure.

- If it is assumed that women weigh up their perception of the various aspects of the process and outcomes of the treatment they received to come to an overall assessment in terms of their stated satisfaction with treatment, then this measure perhaps come nearest to an all-embracing measure of the relative effectiveness of AH and TCRE from a RCT. If this is the case, AH would probably be considered more effective than TCRE: 96% of women were 'very satisfied' or 'quite satisfied' with AH compared to 79% with TCRE in the Bristol trial at a mean follow-up of 2.2 years.

- The implications of the two forms of treatment for NHS costs is rather clearer than their relative effectiveness. The initial cost of AH is

318

significantly higher than that of TCRE, a cost difference which is driven largely by the higher length of stay in hospital. The key question has been whether this cost differential will be substantially eroded, or even removed, by the re-treatment necessary for some women after TCRE? On the basis of the data reported in Chapter 3, the answer to this question appears to be that it will not. The longer the follow-up, the more women will reach the menopause or be prepared to put up with any failure of TCRE until the menopause, so the failure curve shown in Figure 3.1 is likely to flatten out. This, together with the effect of discounting of future costs, would suggest that TCRE will retain its cost advantage.

■   A key issue, therefore, is whether the improved effectiveness of AH, in terms of women's satisfaction with treatment, is sufficient to justify its additional cost. The tools of CEA express this problem for purchasers in terms of an incremental cost effectiveness ratio. On the basis of the evidence from the two-year follow-up in the Bristol RCT, this ratio was estimated to be £1882 per additional women satisfied with treatment. It remains for decision makers to decide whether this is a reasonable cost to pay.

The RCT is seen as the gold standard of clinical evidence, and is becoming an important vehicle for the economic analysis of MAS. However, both in general terms and in the specific context of the comparison of AH and TCRE, RCTs need to be augmented by data from other sources to assist decisions about resource allocation. However, methods for moving beyond standard trial-based analysis are not refined. An important goal of this thesis has been to develop evaluative methods in three areas: generic measures of benefit for use in CUA which adequately reflect patients' preferences for the process and outcomes of care; methods to assess the generalisability of RCT-based economic evaluation when the resource and non-resource consequences of routine clinical practice are considered; and an analysis of how choices about treatment allocation might be shaped by the preferences of individual patients.

9.3   Cost-utility analysis

Chapters 5 and 7 of the thesis considered in detail the methods of CUA and their
implications for the economic comparison of AH and TCRE. The purpose of
undertaking such an analysis in this clinical context is threefold:

(a)   to assess cost-effectiveness using a measure of benefit which draws
        together the multi-dimensional outcomes of the two treatments more
        meaningfully than the satisfaction rates collected in the trial;

(b)   given the fact shown in Chapter 4 that women do have preferences about
        the characteristics of treatment, to undertake this combination of
        outcomes into an overall measure of benefit in a manner which reflects
        patients' preferences about the relative value of those outcomes; and

(c)   to provide a generic measure of benefit which has meaning outside the
        area of menorrhagia and can, therefore, provide an input into decision
        making across specialties and treatment areas.

9.3.1 Methodological analysis

Although CUA is now frequently used in the economic evaluation of health care
programmes [Gerard, 1992], much uncertainty exists about the appropriate
methods for such analysis. The methodological issues relating to CUA
considered in this Chapters 5 and 7 are detailed below.

■   The methods adopted in the base-case of the CUA reported in Chapter 5
      focused on standard quality-adjusted life years (QALYs). The process
      involved the development of descriptive scenarios to represent key health
      states experienced after the procedures; the valuation of those health
      states, using the time trade-off instrument, by a sample of 60 women
      recently referred to hospital with menorrhagia; and the estimation of
      QALYs using this group's values. Other methods have been used in
      economic evaluation and could have been adopted here. For example,
      women's prognosis after surgery could have been broken up into a
      different series of health states; the scenarios could have been framed

differently; an alternative valuation instrument could have been employed; and health state values could have been elicited from a different sample of individuals. The evidence on framing effects is conflicting and a recent large-scale study in the UK indicates that the time trade-off is probably the strongest of the choice-based valuation instruments, at least in practical terms. However, the range of methods available for CUA inevitably emphasises the degree of analytical uncertainty which exists in this and all other CUAs.

■ The use of CUA to reach conclusions about the cost-effectiveness of MAS compared to conventional surgery, where one intervention is more costly and generates more benefits in terms of QALYs, is limited by the absence of widely accepted cost per QALY thresholds. Such thresholds are necessary, if CUA is to be used to inform cross-programme resource allocation, as a means of systematically exploring the robustness of conclusions to the myriad sources of uncertainty that exist within analyses. However, these thresholds are likely to vary by location and over time. Tentative suggestions for the value of these thresholds have been been put forward in Canada. These have no firm methodological or empirical basis and, although the broad range of thresholds used throughout this thesis is based on the suggested Canadian thresholds, they are illustrative only - local decision makers will have to decide the extent to which they reflect policies towards resource allocation in their areas. Based on these illustrative thresholds and following on from earlier work undertaken by the author and colleagues [Briggs *et al*, 1994; Briggs and Sculpher, 1995], the organisation and presentation of the sensitivity analyses in Chapter 5 represent one contribution of this thesis.

■ The standard QALY is characterised by splitting possible prognoses into a series of health states which are valued and aggregated over time. Chapter 7 shows that there is little evidence to link this approach to benefit measurement in CUA to individual preferences. The value of the standard QALY approach to CUA is as a management tool, as one

element of the decision maker's resource allocation armamentarium. If the benefit measure used in the economic evaluation of MAS is to reflect how patients trade-off the various process and short- and long-term outcomes of these interventions relative to open surgery, something less crude than the standard QALY is required. A range of adaptions and alternatives to the standard QALY have been suggested, but rarely employed in applied studies.

■ Chapter 7 used one of these measures - the *ex ante* healthy years equivalent (HYE) - to assess the robustness of the standard CUA in Chapter 5 to an alternative approach to benefit estimation. Instead of splitting patients' possible prognoses following surgery into discrete health states which are valued separately and in isolation, the *ex ante* HYE values prognoses taken as a whole which include their inherent risks and time-related factors. *Ex ante* HYEs were estimated from a sample of 63 women with menorrhagia, presenting each with a scenario describing the lifetime profile of health-related factors that may follow the two forms of surgery. The time trade-off was used to obtain a direct measure of the periods of time in good health each woman considered equivalent to these lifetime profiles. A notable outcome of the work presented in Chapter 7 showed that it is feasible to elicit *ex ante* HYEs from a patient population.

■ Women's mean and median *ex ante* HYEs were found to be higher for AH than TCRE, which is consistent with estimated benefits in terms of expected QALYs. However, the sampling variation around these estimates is such that the two HYE estimates are not statistically significantly different. As an all-embracing measure of benefit related to particular technologies, *ex ante* HYEs can be directly related to other indications of individuals' treatment-related preferences, both stated and revealed by behaviour. Chapter 7 described a comparison of the descriptive preferences provided by women in relation to AH and TCRE and their *ex ante* HYEs. Although the sample size was small, some indications of consistency between these two measures of preference can

322

be discerned, particularly at the level of the group, and this is an important contribution of the thesis.

■   The work reported in Chapters 5 and 7 emphasises that the choice between the HYE and the standard QALY for applied economic evaluation presents a trade-off to analysts. The QALY has value as a management tool, but has no clear link with individual preferences; and it imposes a reasonable measurement task, particularly when valuation systems are used. The HYE has a clearer link to individual preferences, but it is likely to impose a greater measurement burden; it also lacks the flexibility of QALYs for use in decision analytic modelling.

■   Recently, a parallel area of methodological research in economic evaluation in health, to the development of benefit measures for CUA, has been the refinement of stated preference methods to facilitate cost-benefit analysis (CBA). The major difficulty of valuing health-related outcomes in monetary terms is the principal reason why CBA has not been widely used in the field of health. However, recent work on willingness to pay (WTP) methods has opened the possibility that CBA can more frequently be applied in the area. Indeed, by trying to tie the measurement of benefit used in economic evaluation more firmly to the principles of welfare economic theory, the development of the *ex ante* HYE mirrors that of WTP: typically, both methods present responders with descriptions of interventions including information on risk, but the basis of valuation with HYEs is healthy years whereas that for WTP is money. It remains an important research question whether economic evaluation in health should move towards HYEs or WTP, or neither.

9.3.2 Empirical analysis

The CUA of AH and TCRE, using both standard QALYs and *ex ante* HYEs, generated some important empirical conclusions.

■   Using the mean health state values elicited from women interviewed in
    Bristol and a simple decision tree to model the clinical pathways along
    which women pass, AH was found to generate more expected QALYs
    than TCRE. Expected QALYs were sensitive to the values women
    attached to the post-convalescence/pre-menopause health states relating
    to the two forms of surgery. However, the value elicited for the post-AH
    health state was statistically significantly higher than that for the post-
    TCRE state, indicating that this conclusion was robust at least to the
    sampling variation in the two health state values.

■   When the estimates of expected QALYs were synthesised with cost
    estimates based on resource use data collected within the Bristol RCT,
    the base-case estimate of the incremental cost per additional QALY with
    AH was £1,500. As with the incremental cost-effectiveness ratio in
    terms of additional women satisfied with treatment, decision makers will
    have to decide whether this is a reasonable additional cost to pay for the
    extra benefits offered by AH for the typical women with menorrhagia.
    Here, two illustrative cost per QALY thresholds were used to guide
    decision makers. If these thresholds are widely accepted, AH is likely to
    be considered more cost-effective than TCRE on the basis of the base-
    case estimate of costs and benefits.

■   A series of sensitivity analyses has found that this conclusion is robust to
    plausible one-way variation in data inputs. Although the estimate of
    incremental cost per QALY was sensitive to such parameters as the cost
    of a day on a ward and the value attached to the post-surgery health
    states, plausible individual variation in these parameters was not
    sufficient to take the ratio above the lower illustrative ratio of £6,500.
    However, the conclusion that AH is more cost-effective than TCRE was
    not robust to simultaneous variation in all uncertain data inputs. When all
    the parameters were varied together to generate a pessimistic scenario
    from the viewpoint of AH, the resulting ratio was as high as £255,000.
    Although the circumstances necessary to generate such a high ratio in

practice are unlikely to occur, the size of this theoretical ratio demands some caution in interpreting the base-case cost per QALY estimate.

■ When costs are related to the HYE-based measure of benefit, the incremental cost per additional HYE was £156, on the basis of mean benefits. Although it is difficult to judge what would be considered an acceptable ratio because very few CUAs using HYEs have been undertaken, this would probably be seen as a modest additional cost per unit of benefit. However, the fact that the HYE estimates for AH and TCRE were not statistically significantly different generates a major source of uncertainty in this conclusion.

9.4 Generalisability

Although trial-based evidence can provide high levels of internal validity for a clinical and economic evaluation of MAS, it often fails to reflect the resource and non-resource consequences of interventions as they are used in routine clinical practice. Furthermore, the reliance only on trial data within an economic evaluation may overlook the fact that variation in costs and outcomes within trials may not be an accurate reflection of variation within routine practice; and that new technologies tend to develop more quickly than trials can evaluate them. These considerations generate a further source of uncertainty for estimates of cost-effectiveness. Chapter 6 of this report, therefore, looked at methods that can be used to assess the generalisability of an economic study.

9.4.1 Methodological analysis

The development of a framework within which to assess the generalisability of economic evaluations is an important contribution of this thesis. The key methodological issues are discussed below.

■ The location and context of an economic evaluation is likely to influence markedly the four elements of an analysis: resource use, outcomes, unit

costs and health state valuation data. Despite this, formal assessment of generalisability is rarely undertaken in economic evaluation, with the limited published work focusing on comparing the results of studies across national boundaries. The review of economic evaluations of MAS in Chapter 2 indicated that a number of studies used sensitivity analysis to assess the importance of variation in some parameters which may vary by location and context. A few evaluations looked at generalisability more explicitly, using scenario analysis to explore under which local circumstances the conclusions of the analysis may alter.

■   The analysis in Chapter 6 of this thesis went a step further by using sensitivity analysis to incorporate alternative sources of data into an analytical framework, where those sources are likely to be more representative of routine practice. An important characteristic of this sort of analysis is the use of data with high internal validity (usually adequately powered RCTs) to provide the main data source for the base-case analysis, and the use of alternative data sources systematically to explore the robustness of the results. As decision makers begin to require analyses to be more relevant to their local situation, this approach to the assessment of generalisability is likely to be adopted more frequently.

■   The advent of meta-analysis offers a valuable tool to increase the statistical power of clinical evaluations by pooling the results of a number of smaller studies. However, little thought has been given to the appropriate role of meta-analysis within economic evaluation. In combining the results of a number of RCTs, meta-analysis is a useful way of generating more reliable estimates of the resource and non-resource implications of interventions for the *base-case analysis* of economic evaluations of MAS and other technologies (Stage III analysis). However, the danger of meta-analysis is that the heterogeneity that exists when any trials are compared is masked, and there will remain an important role for the incorporation into economic analysis of disaggregated trial data, to

explore the implications for variability between trials as an aspect of the analysis of generalisability.

■ The need for analysis of generalisability highlights the importance of the decision analytic model as a framework for economic evaluation. Even when the primary source of data for a study is a RCT, the use of models can provide a platform to explore the importance of variation in parameters and to incorporate non-trial data (Stage IV analysis).

9.4.2 Empirical analysis

The analysis of generalisability in relation to AH versus TCRE focuses on how robust the results presented in Chapter 5 are to variation in key parameters. The alternative estimates come from alternative RCTs and the a national survey, where the latter represents a description of a large number of TCREs undertaken in England, Wales and Northern Ireland over a 18 month period. The following conclusions can be reached from this element of the study.

■ These alternative sources of data indicate differences in a range of parameters. However, if the illustrative cost per QALY thresholds suggested in Chapter 5 are acceptable, these differences are not important in terms of their impact on the incremental cost per additional QALY of AH relative to TCRE. Again, the variation in parameters is not sufficient to take the estimated ratio above the lower illustrative threshold ratio of £6,500.

■ Part of this analysis of the generalisability of the base-case results focused on resource intensive and resource sparing clinical practice in relation to TCRE using data from the national survey, and the extent to which parameter estimates based on these types of practice influence the cost per QALY ratios. Resource sparing practice increases the estimated cost per QALY of AH (to £3153 from £1,500 in the base-case). Resource intensive practice has only a modest impact on the cost per QALY (reducing it to £1484). Therefore, this analysis of the extremes of

327

clinical practice as regards resource use has little impact on the base-case conclusions. However, the national survey provides data only for TCREs; a detailed picture of routine practice in relation to hysterectomy is currently awaited.

■ The analysis of generalisability also looked at the robustness of the base-case conclusions to alternative unit costs. Three specific hospitals were located which use similar costing software and methods, and these centres were asked to provide estimates of two unit costs to which the total expected cost of the two treatments is sensitive: the cost of a day in hospital and of a minute in theatre. The variation between the three hospitals is limited (theatre costs of between £1.96 and £2.04 per minute; ward costs of between £83 and £104 per day), but there is greater difference compared to the base-case unit costs (theatre £1.08 per minute; ward £120 per day) . Overall, the substitution of these alternative values has very limited impact on the estimate of cost per QALY.

■ The final component of the analysis of generalisability was a broad-brush analysis of the comparative costs and benefits of non-hysterectomy forms of surgery other than TCRE with loop diathermy, and of forms of hysterectomy other than AH. The purpose of this analysis was to explore whether the AH versus TCRE analysis, which is the focus of the thesis, can be generalised to one of non-hysterectomy forms of surgery versus hysterectomy. Although the results have to be interpreted cautiously given the absence of experimental data for each comparison, it is unlikely that any of the alternative non-hysterectomy forms of surgery upon which data were collected in the national survey are so markedly different to TCRE with loop diathermy, in terms of either resource or non-resource consequences, to affect the base-case conclusions significantly.

■ The estimated cost of the various alternative forms of hysterectomy to AH suggest that scope does exist to reduce the cost of treatment. In

particular, if a greater use were made of vaginal hysterectomy or laparoscopic-assisted vaginal hysterectomy (LAVH) with reusable equipment, the cost of in-patient care would probably fall. The analysis indicates, however, that LAVH with disposable consumables may increase the cost of hysterectomy markedly. The exploratory analysis presented here suggests that a change in the form of hysterectomy provided is unlikely to take the cost per QALY ratio, relative to TCRE with loop diathermy, over the lower illustrative threshold of £6,500.

9.5   Preference-based treatment allocation

Most applications of MAS have quite different characteristics, in terms of their process and possible outcomes, to open surgery, and the characteristics of MAS may not always be preferable. In the context of surgical treatment for menorrhagia, the importance of Chapter 4 of this thesis is that it indicates that many women have views about the characteristics of treatment they prefer and, in some cases, about the actual treatment they wish to undergo. Furthermore, Chapter 5 showed that the benefit estimates for TCRE and AH are sensitive to the values women attach to health states associated with the prognoses following these two treatments. These findings suggest that, instead of economic evaluation being used to identify which one of these two treatments is the more cost-effective based on mean costs and benefits, it should be possible to explore the economic characteristics of management strategies where women's preferences play an important role in deciding which treatment they should receive. The methodological and empirical analysis of preference-based forms of management in Chapter 8 represents an important contribution of the thesis.

9.5.1 Methodological analysis

The methodological issues related to the modelling and evaluation of management strategies where patients' preferences determine treatment allocation are under researched. A range of issues was covered in Chapter 8.

■   The standard approach to economic evaluation is the all or nothing analysis, where a mean costs and benefits are used to identify a single economically preferred intervention. It would then be expected that this would be the option to be provided to all patients unless there is a good non-economic reason for not doing so. It is, however, recognised that this approach may not be appropriate if there is heterogeneity between patients in their baseline clinical and demographic characteristics, and that these factors may influence outcomes. In this context, sub-group analysis is well established as a means of identifying groups of patients, defined in clinical terms, for a whom an intervention is cost-effective while not being so for other sub-groups.

■   However, heterogeneity in *patients' preferences* can mean that a decision to provide a single treatment option for all patients, on the basis of mean costs and benefits, may result in a significant proportion of women not receiving the treatment which their preferences indicate is their personal optimum. If the values patients attach to the process and outcomes of care are to be used to determine resource allocation (a key premise of the principles of economic evaluation in health), it is surely valid to assess whether sub-groups can be defined in terms of patients' preferences, which may then used to drive treatment allocation.

■   Of the three forms of preference-based treatment allocation considered here, the one that adheres most closely with the principles of economic evaluation is patient-specific CEA. This approach considers a patient's (health state or profile) values in the context of the incremental cost of the treatment which those values indicate is optimal for that patient. However, from a practical viewpoint, the development of systems to allocate patients to therapies on this basis may prove hard to accomplish.

■   The use of health state values as part of a values-based or patient-specific CEA-based treatment allocation policy again raises the issue of

the relationship between the benefit measure used in economic evaluation and patient preferences. Given the evidence reviewed in Chapter 7 suggesting that the link between preferences and QALYs is tenuous, the use of health state values to allocate treatment may result in patients receiving a treatment that they would not choose. The possible role for alternative benefit measures such as the *ex ante* HYE would, therefore, have to be considered further, if these forms of management were to be employed.

9.5.2 Empirical analysis

The analysis presented in Chapter 8 generates some important conclusions in the context of AH and TCRE.

■   If women are allocated according to their choice, it would be expected that the mean cost of care would lie somewhere between that of AH and TCRE. On the basis of questionnaire responses from women in Bristol and Swindon, the cost of this form of management would be between £964 and £1,035. However, the choice-based strategy would probably be considered more cost-effective than a TCRE only strategy, with an incremental cost per QALY of between £282 and £413. The extent to which this particular preference-based strategy was found to be more cost-effective than AH alone depends on the benefit measure used: in terms of QALYs and using health state values from the women in Bristol, AH was found to have a relatively modest cost per QALY (£2,396); in terms of *ex ante* HYEs, however, the choice-based strategy was found to dominate AH.

■   The values-based approach was found to have an expected cost of between £964 and £1,081. In terms both of expected QALYs and *ex ante* HYEs, it would probably be considered more cost-effective than either AH only or TCRE only, dominating AH and with an incremental cost per QALY over TCRE of £307.

■ Management based on patient-specific CEA has the potential to be the most cost-effective of the preference-based strategies. Compared to TCRE only it was found to offer additional benefits at a modest incremental cost; and the strategy was found to dominate AH. Compared to the other preference-based forms of management, patient-specific CEA would probably be considered the more cost-effective.

■ The systematic use of women's preferences to determine treatment allocation has, then, been shown to have the potential to be a cost-effective form of management in the treatment of menorrhagia. In practical terms, the use of patient-specific CEA is perhaps less feasible than the values-based approach, and the choice-based form of management is probably the most likely to be adopted.

9.6 Further research

The projection that, by the year 2000, 70% to 80% of surgical practice will be based on MAS techniques highlights the importance of detailed evaluation in this area [Cushieri, 1993]. This is likely to require the further development of the methods of clinical and economic evaluation. This thesis has focused specifically on developments in the methods of economic evaluation necessary to cope with the characteristics of MAS. Some specific areas of further research are necessary.

■ If CUA is to have an important role in supporting decisions made by local purchasers, the systematic handling of uncertainty within studies would benefit greatly from some guidelines regarding acceptable cost per QALY thresholds. Research would be valuable to identify whether purchasers have such thresholds in their minds when they consider the results of studies; and whether work can be undertaken with purchasers to translate explicit policy objectives into threshold values.

■   A programme of further research is required to assess the extent to which QALYs and HYEs generate different conclusions in a range of applied evaluations; whether the measurement of HYEs can be made more manageable, for example, by using video and group valuations; and whether a firmer link between QALYs and individual preferences can be established.

■   A modelling approach to the economic evaluation of preference-based management strategies is adopted in Chapter 8, which has the advantage of flexibility. It would be useful to extend this modelling work in a number of directions. These would include the use of larger and more representative samples of women; the provision of improved information on alternative treatments to these women upon which they can make choices; and the development of a more detailed cost-utility model which would facilitate the incorporation of patient-specific clinical information as well as preferences. In addition or as an alternative, it is important to subject the preference-based strategies to prospective evaluation using a randomised trial. This would facilitate an analysis of how this sort of management might work in practice when patients are actually in the position to have to make treatment decisions.

9.7   Concluding comments

The characteristics of MAS highlight a number of weaknesses in the methods of economic evaluation. An important finding of this thesis is that, despite being considered by many to be the gold standard for clinical evaluation, the use of the RCT as a vehicle for economic evaluation is likely, if used in isolation, to leave unanswered a number of important issues relating to the cost-effectiveness of MAS applications. Given the likely significant increase in the use of this group of technologies in the near future, it essential to develop further key areas of evaluative method. The contribution of this thesis has been to begin that process in the important areas of benefit measures for CUA, the analysis of

generalisability and the economic evaluation of preference-based management strategies.

References

Adams ME, McCall NT, Gray ET, *et al.* [1992]. Economic analysis in randomized controlled trials. *Medical Care*, **vol. 30**, pp231- 243.

Advisory Council on Sceince and Technology (ACOST) [1993]. *A Report on Medical Research and Health.* London: HMSO.

Anderson GM, Lomas J [1988]. Monitoring the diffusion of a technology: coronary artery bypass surgery in Ontario. *American Journal of Public Health*, **vol. 78**, pp251-254.

Anderson JB, Wolinski AP, Wells IP, *et al.* [1986]. The impact of percutaneous transluminal angioplasty on the management of peripheral vascular disease. *British Journal of Surgery*, **vol. 73**, pp17-19.

Andersen TF, Bronnum-Hansen H, Sejr T, *et al.* [1990]. Elevated mortality following transurethral resection of the prostate for benign hypertrophy. *Medical Care*, **vol. 28**, pp870-881.

Armitage P, Berry G [1987]. *Statistical Methods in Medical Research.* Oxford: Blackwell Scientific.

Arrow K, Solow R, Portney PR, *et al.* [1993]. Report of the NOAA panel on contingent valuation. *Federal Register*, **vol. 58**, pp4601-14.

Asch DA, Hershey JC [1995]. Why some health policies don't make sense at the bedside. *Annals of Internal Medicine*, **vol. 122**, pp846-850.

Association of the British Pharmaceutical Industry [1994]. *Guidance on Good Practice in the Conduct of Economic Evaluations in Medicine.* London: ABPI.

Audit Commission [1992]. All in a day's work: an audit of day surgery in England and Wales. *NHS Occasional Papers*, London: Audit Commission.

Backhouse ME, Backhouse RJ, Edey SA [1992]. Economic evaluation bibliography. *Health Economics*, **vol. 1 (supplement 1)**, pp1-236.

Bailey KR [1994]. Generalizing the results of randomized clinical trials. *Controlled Clinical Trials*, **vol. 15**, pp15-23.

Banta HD [1991]. The budget is blind: the case of minimally invasive therapy in Europe. *TNO Institute of Ageing and Vascular Research*, The Netherlands: Center for Medical Technology.

Banta HD [1992]. Diffusion of minimally invasive therapy in Europe. *Minimally*

335

Invasive Therapy, vol. **1**, pp189-195.

Banta HD [1993A]. The cost-effectiveness of 10 selected applications in minimally invasive therapy. *Health Policy,* vol. **23**, pp135-151.

Banta HD [1993B]. Minimally invasive surgery: implications for hospitals, health workers, and patients. *British Medical Journal,* vol. **307**, pp1546-1549.

Barkun JS, Barkun AN, Sampalis JS, *et al.* [1992]. Randomised controlled trial of laparoscopic versus mini cholecystectomy. *Lancet,* vol. **340**, pp1116-1119.

Barry MJ, Fowler FJ, Mulley AG, *et al.* [1995]. Patient reactions to a program desgined to facilitate patient participation in treatment decisions for benign prostatic hyperplasia. *Medical Care,* vol. **33**, pp771-782.

Bass EB, Pitt HA, Lillemoe KD [1993]. Cost-effectiveness of laparoscopic cholecystectomy versus open cholecystectomy. *American Journal of Surgery,* vol. **165**, pp466-471.

Baum M [1993]. New approach for recruitment into randomised controlled trials. *Lancet,* vol. **341**, pp812-813.

Baxter JN, O'Dwyer PJ [1992]. Laparoscopic or minilaparotomy cholecystectomy? *British Medial Journal,* vol. **304**, pp559-560.

Bdesha AS, Bunce CJ, Kelleher JP, *et al.* [1993]. Transurethral microwave treatment for benign prostatic hypertrophy: a randomised controlled clinical trial. *British Medical Journal,* vol. **306**, pp1293-1296.

Becker G [1964]. *Human Capital.* New York: Columbia University Press.

Beisecker A, Beisecker T [1990]. Patient information-seeking behaviors when communicating with doctors. *Medical Care,* vol. **28**, pp19-28.

Berget A, Andreasson B, Bock JE, *et al.* [1987]. Outpatient treatment of cervical intra-epithelial neoplasia - the CO_2 laser versus cryotherapy, a randomized trial. *Acta Obstet Gynaecol Scand,* vol. **66**, pp531-536.

Berggren U, Gordh T, Grama D, *et al.* [1994]. Laparoscopic versus open cholecystectomy: hospitalization, sick leave, analgesia and trauma responses [see comments]. *Br J Surg,* vol. **81**, pp1362-5.

Berry DA [1993]. A case for Bayesianism in clinical trials. *Statistics in Medicine,* vol. **12**, pp1377-1393.

Bevan PG [1989]. *The Management and Utilisation of Operating Departments.* London: NHS Management Executive.

Bisonni RS, Lawler FH, Holtgrave DR [1993]. Transurethral prostatectomy versus transurethral dilatation of the prostatic urethra for benign prostatic hyperplasia: a cost-utility analysis. *Family Practitioner Research Journal*, vol. **13**, pp25- 36.

Bleichrodt H [1995]. QALYs and HYEs: under what conditions are they equivalent? *Journal of Health Economics*, vol. **14**, pp17-37.

Borzak S, Ridker PM [1995]. Discordance between meta-analyses and large-scale randomized, controlled trials: examples from the management of acute myocardial infarction. *Annals of Internal Medicine*, vol. **123**, pp873-877.

Boto TCA, Fowler CG [1990]. Surgical alternatives to hysterectomy for intractable menorrhagia. *British Journal of Hospital Medicine*, vol. **44**, pp93-99.

Bouyssou D, Vansnick JC [1988]. A note on the relationship between utility and value functions. In: Munier BR (ed), *Risk, Decision and Rationality*. Dordrecht: Reidel.

Bradlow J, Coulter A, Brooks P [1992]. *Patterns of Referral, Oxford*. Oxford: Health Services Research Unit.

Brazier JE, Harper R, Jones NMB, *et al.* [1992]. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *British Medical Journal*, vol. **305**, pp160-4.

Brazier J, Jones N, Kind P [1993]. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnarie. *Quality of Life Research*, vol. **2**, pp169-180.

Brazier JE, Usherwood TP, Harper R, *et al.* [1994]. Deriving a single index measure for health from the SF-36 - preliminary results. Health Economists' Study Group, Newcastle, July: University of Sheffield.

Brewin CR, Bradley C [1989]. Patient preferences and randomised clinical trials. *British Medical Journal*, vol. **299**, pp313-315.

Briggs AH [1995]. *OHE Briefing. Handling uncertainty in the results of economic evaluation*. London: OHE.

Briggs A, Sculpher M [1995]. Sensitivity analysis in economic evaluation: a review of published studies. *Health Economics*, vol. **4**, pp355-371.

Briggs A, Sculpher M, Buxton M [1994]. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Economics*, vol. **3**, pp95-104.

British National Formulary [1995]. *British National Formulary*. London: British Medical Association and the Royal Pharmaceutical Society of Great Britain.

Brody DS, Miller SM, Lerman CE, *et al.* [1989]. Patient perception of involvement in medical care: relationship to illness attitudes and outcomes. *Journal of General Internal Medicine*, **vol. 4,** pp506-511.

Brooks R [1996]. EuroQol: the current state of play. *Health Policy,* **vol. 37,** pp53-72.

Brown J, Sculpher MJ [1993]. Economics of screening programmes to prevent cervical cancer. *Contemporary Reviews in Obstetrics and Gynaecology,* **vol. 5,** pp221-229.

Brumsted JR, Blackman JA, Badger GJ, *et al.* [1996]. Hysteroscopy versus hysterectomy for the treatment of abnormal uterine bleeding: a comparison of cost. *Fertility and Sterility,* **vol. 65,** pp310-316.

Bryan J [1993]. No large savings yet made from small cuts. *Healthcare Management,* pp50-51.

Buckingham K [1993]. A note on HYE (healthy years equivalent). *Health Economics,* **vol. 11,** pp301-309.

Buxton M, Ashby J, O'Hanlon M [1987]. *Alternative Methods of Valuing Health States*. Uxbridge: HERG, Brunel University.

Buxton M, Sculpher M [1994]. Stopping of clinical trials. *Lancet,* **vol. 343,** pp179.

Cairns J [1992]. Discounting and health benefits: another perspective. *Health Economics,* **vol. 1,** pp76-81.

Cairns J [1996]. Measuring health outcomes. Condition-specific and patient specific measures are of limited use when allocating resources. *British Medical Journal,* **vol. 313,** pp6.

Cairns J, Shackley P, Hundley V [1996]. Decision making with respect to diagnostic testing: a method of valuing the benefits of antenatal screening. *Medical Decision Making,* **vol. 16,** pp161- 168.

Canadian Coordinating Office for Health Technology Assessment [1994]. *Guidelines for Economic Evaluation of Pharmaceuticals: Canada*. Ottawa: CCOHTA.

Cantor SB [1994]. Cost-effectiveness analysis, extended dominance, and ethics: a quantitative assessment. *Medical Decision Making,* **vol. 14,** pp259-265.

Carlsson P, Naucler J, Petterson S, *et al.* [1989]. A cost- effectiveness analysis of

extracorporeal shock wave lithotripsy and percutaneous nephrolithotomy. *Scandanavian Journal of Urology Nephrology Supplement*, vol. **122**, pp44-51

Carlsson P, Kinn AC, Tiselius HG, *et al.* [1992]. Cost- effectiveness of extracorporeal shock wave lithotripsy and percutaneous nephrolithotomy for medium-sized kidney stones. *Scandinavian Journal of Urology & Nephrology*, vol. **26**, pp257- 263.

Cassileth BR, Zupkis RV, Sutton-Smith K, *et al.* [1980]. Information and participation preferences among cancer patients. *Annals of Internal Medicine*, vol. **92**, pp832-836.

Centerwall BS [1981]. Premenopausal hysterectomy and cardiovascular disease. *American Journal of Obstetrics and Gynaecology*, vol. **139**, pp58-61.

Chang RW, Falconer J, Stulberg D, *et al.* [1990]. Prerandomisation: an alternative to classic randomisation. *The Journal of Bone and Joint Surgery*, vol. **72A**, pp1451-1455.

Chartered Institute of Public Finance and Accountancy (CIPFA) [1990]. *Health Service Trends, Volume 2 - Detailed Tables*. London: Health Care Financial Management Association.

Chestnut LG, Keller LR, Lambert WE, *et al.* [1996]. Measuring heart patients' willingness to pay for changes in angina symptoms. *Medical Decision Making*, vol. **16**, pp65-77.

Clarke A, Black N, Rowe P, *et al.* [1995]. Indications for and outcome of total abdominal hysterectomy for benign disease: a prospective cohort study. *British Journal of Obstetrics & Gynaecology*, vol. **102**, pp611-620.

Cleary PD, Greenfield S, McNeil BJ [1991A]. Assessing quality of life after surgery. *Controlled Clinical Trials*, vol. **12**, pp189S- 203S.

Cleary PD, Greenfield S, Mulley AG, *et al.* [1991B]. Variations in length of stay and outcomes for six medical and surgical conditions in Massachusetts and California. *Journal of the American Medical Association*, vol. **266**, pp73-79.

Cohen DJ, Breall JA, Ho KKL, *et al.* [1994]. Evaluating the potential cost-effectiveness of stenting as a treatment for symptomatic single-vessel coronary disease: use of a decision- analytic model. *Circulation*, vol. **89**, pp1859-1874.

Concato J, Horwitz RI, Feinstein AR, *et al.* [1992]. Problems of comorbidity in mortality after prostatectomy. *Journal of the American Medical Association*, vol. **267**, pp1077-1082.

Consumers' Association Ltd [1990]. Drugs for menorrhagia: often disappointing.

Drug and Therapeutics Bulletin, vol. **28,** pp17-20.

Cook J, Richardson J, Street A [1993]. *Cost utility analysis of treatment for gallstone disease: interim results. Working Paper 28.* Melbourne: National Centre for Health Program Evaluation, Monash University.

Cook J, Richardson J, Street A [1994]. A cost utility analysis of treatment options for gallstone disease: methodological issues and results. *Health Economics,* vol. **3,** pp157-168.

Cordray DS [1990]. Strengthening causal interpretations of nonexperimental data: the role of meta-analysis. In: Sechrest L, Perrin E, Bunker J. *Research methodology: Strengthening Causal Interpretation of Nonexperimental Data,* Rockville, MD: AHCPR.

Cotton PB [1992]. Therapeutic gastrointestinal endoscopy: problems in proving efficacy. *New England Journal of Medicine,* vol. **326,** pp1626-1628.

Coulter A [1994]. Trends in gynaecological surgery. *Lancet,* vol. **344,** pp1367.

Coulter A, Bradlow J, Agass M, *et al.* [1991]. Outcomes of referrals to gynaecology outpatient clinics for mentrual problems: an audit of general practice records. *British Journal of Obstetrics and Gynaecology,* vol. **98,** pp789-796.

Coulter A, Klassen A, MacKenzie IZ, *et al.* [1993]. Diagnostic dilatation and curettage: is it used appropriately? *British Medical Journal,* vol. **306,** pp236-239.

Coulter A, Peto V, Jenkinson C [1994A]. Quality of life and patient satisfaction following treatment for menorrhagia. *Family Practitioner,* vol. **11,** pp394-401.

Coulter A, Peto V, Doll H [1994B]. Patients' preferences and general practitioners' decisions in the treatment of menstrual disorders. *Family Practice,* vol. **11,** pp67-74.

Coulter A, Long A, Kelland J, *et al.* [1995A]. Managing menorrhagia. *Quality in Health Care,* vol. **4,** pp218-226.

Coulter A, Kelland J, Peto V, *et al.* [1995B]. Treating menorrhagia in primary care. An overview of drug trials and a survery of prescribing practice. *International Journal of Technology Assessment in Health Care,* vol. **11,** pp456-471.

Cromwell J, Mitchell JB, Stason WB [1990]. Learning by doing in CABG surgery. *Medical Care,* vol. **28,** pp6-18.

Cuckow PM [1994]. Minimally invasive surgery. Increases the cost of treatment. (Letter). *British Medical Journal,* vol. **308,** pp340-341.

Culyer AJ, Wagstaff A [1993]. QALYs versus HYEs. *Journal of Health Economics,* **vol. 11,** pp311-323.

Culyer AJ, Wagstaff A [1995]. QALYs versus HYEs: a reply to Gafni, Birch and Mehrez. *Journal of Health Economics,* **vol. 14,** pp39-45.

Cuschieri A [1991]. Minimal access surgery and the future of interventional laparoscopy. *American Journal of Surgery,* **vol. 161,** pp404-407.

Cuschieri A [1993]. *Minimal Access Surgery: Implications for the NHS (Report from a Working Group).* Edinburgh: HMSO.

Daly E, Gray A, Barlow D, *et al.* [1993]. Measuring the impact of menopausal symptoms on quality of life. *British Medical Journal,* **vol. 307,** pp836-840.

Dasbach EJ, Fryback DG, Newcomb PA, *et al.* [1991]. Cost- effectiveness of strategies for detecting diabetic retinopathy. *Medical Care,* **vol. 29,** pp20-39.

Dauleh MI, Rahman S, Townell NH [1995]. Open versus laparoscopic cholecystectomy: a comparison of postoperative temperature. *J R Coll Surg Edinb,* **vol. 40,** pp116-8.

Davis CE [1994]. Generalizing from clinical trials. *Controlled Clinical Trials,* **vol. 15,** pp11-14.

de Boer MJ, van Hout BA, Liem AL, *et al.* [1995]. A cost-effective analysis of primary coronary angioplasty versus thrombolysis for acute myocardial infarction. *American Journal Cardiology,* **vol. 76,** pp830-3.

Department of Health and Social Security (DHSS) [1987]. *General Medical Practitioners' Workload.* London: DHSS.

Department of Health [1993A]. *Research for Health.* London: Department of Health.

Department of Health [1993B]. *The Government's Response to the ACOST Report on Medical Research and Health.* London: Department of Health.

Department of Health [1995]. *Report of the NHS Health Technology Assessment Programme 1995.* Leeds: Department of Health.

Detsky AS, Naglie IG [1990]. A clinicians guide to cost- effectiveness analysis. *Annals of Internal Medicine,* **vol. 113,** pp147-154.

Dicker RC, Greenspan JR, Strauss LT, *et al.* [1982]. Complications of abdominal and vaginal hysterectomy among women of reproductive age in the United States. *American Journal of Obstetrics and Gynaecology,* **vol. 144,** pp841-848.

Diehl AK [1993]. Laparoscopic cholecystectomy: too much of a good thing? *Journal of the American Medical Association*, vol. **270**, pp1469-1470.

Dolan P [forthcoming]. Modelling valuations for health states: the effect of duration. *Health Policy*,

Dolan P, Gudex C [1995]. Time preference, duration and health state valuations. *Health Economics*, vol. **4**, pp289-299.

Dolan P, Gudex C, Kind P, *et al.* [1996]. Valuing health states: a comparison of methods. *Journal of Health Economics*, vol. **15**, pp209-231.

Doll HA, Black NA, Flood AB, *et al.* [1993]. Criterion validation of the Nottinhgam health profile: patient views of surgery for benign prostatic hypertrophy. *Social Science and Medicine*, vol. **37**, pp115-122.

Donaldson C, Shackley. P., Abdalla M [1995]. Willingness to pay for antenatal carrier screening for cystic fibrosis. *Health Economics*, vol. **4**, pp439-452.

Dowie J [1993]. 'Process utility' can seriously damage your health service evaluation but the generic measure of benefit should include 'service outcomes'. HESG paper, Strathclyde, June.

Drummond M [1994]. *Economic Analysis Alongside Controlled Trials: an Introduction for Clinical Researchers*. Leeds: Department of Health.

Drummond MF, Jefferson TO [1996]. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *British Medical Journal*,

Drummond MF, Stoddart GL [1984]. Economic analysis and clinical trials. *Controlled Clinical Trials*, vol. **5**, pp115-128.

Drummond MF, Bloom BS, Carrin G, *et al.* [1992]. Issues in the cross-national assessment of health technology. *International Journal of Technology Assessment n Health Care*, vol. **8**, pp671- 682.

Drummond MF, Brandt A, Luce B, *et al.* [1993A]. Standardizing methodologies for economic evaluation in health care. *International Journal of Technology Assessment*, vol. **9**, pp26-36.

Drummond MF, O'Brien BJ [1993]. Clinical importance, statistical significance and the assessment of economic and quality-of-life outcomes. *Health Economics*, vol. **2**, pp205-212.

Drummond M, Torrance G, Mason J [1993B]. Cost-effectiveness league tables: more harm than good? *Social Science Medical*, vol. **37**, pp33-40.

Dukes MNG [1986]. The regulation of drugs: worlds of difference. *International Journal of Technology Asessment in Health Care*, **vol. 2**, pp629-636.

Dwyer N, Hutton J, Stirrat GM [1993]. Randomised controlled trial comparing endometrial resection with abdominal hysterectomy for the surgical treatment of menorrhagia. *British Journal Obstetrics Gynaecology*, **vol. 100**, pp237-43.

East M [1994]. Comparative costs of laparoscopy assisted vaginal hysterectomy. *New Zealand Medical Journal*, **vol. 107**, pp371-374.

Eddy DM [1990]. Comment: the role of meta-analysis. In: Sechrest L, Perrin E, Bunker J. *Research methodology: Strengthening Causal Interpretation of Nonexperimental Data*, Rockville, MD: AHCPR.

Effective Health Care [1995]. *The Management of Menorrhagia. Effective Health Care Bulletin No. 9*. Leeds: Nuffield Institute for Health, University of Leeds.

Eisenberg JM [1989]. Clinical economics: a guide to the economic analysis of clinical practice. *Journal of the American Medical Association*, **vol. 262**, pp2879-2886.

Eisenberg JM, Glick H, Koffer H [1989]. Pharmacoeconomics: economic evaluation of Pharmaceuticals. In: Strom BL. *Pharmacoepidemiology*, New York: Churchill Livingstone.

England WL, Roberts SD, Grim CE [1987]. Surgery or angioplasty for cost-effective renal revascularisation. *Medical Decision Making*, **vol. 7**, pp84-91.

Eraker SA, Sox HC [1981]. Assessment of patients' preferences for therapeutic outcomes. *Medical Decision Making*, **vol. 1**, pp29- 39.

EuroQol Group [1990]. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy*, **vol. 16**, pp199-208.

vans RG [1995]. Manufacturing consensus, marketing truth: guidelines for economic evaluation. *Annals of Internal Medicine*, **vol. 123**, pp59-60.

Falk PM, Beart RW, Wexner SD, *et al.* [1993]. Laparoscopic colectomy: a critical appraisal. *Disease of the Colon and Rectum*, **vol. 36**, pp28-34.

Farley DE, Ozminkowski RJ [1992]. Volume-outcome relationships and inhospital mortality: the effect of changes in volume over time. *Medical Care*, **vol. 30**, pp77-94.

Feinleib M, Havlik RJ, Gillum RF, *et al.* [1989]. Coronary heart disease and related procedures: national hospital discharge survey data. *Circulation*, **vol. 79**, ppI-13-I-18.

Fendrick AM, Escarce JJ, McLane C, *et al.* [1994]. Communication: hospital adoption of laparoscopic cholecystectomy. *Medical Care*, **vol. 32**, pp1058-1063.

Ferguson JA, Goldacre MJ, Henderson J, *et al.* [1991]. Audit of workload in gynaecology: analysis of time trends from linked statistics. *British Journal of Obstetrics and Gynaecology*, **vol. 98**, pp772-777.

Fitzpatrick R [1993]. Measurement of patient satisfaction. In: Fitzpatrick R, Hopkins A, *Measurement of Patients' Satisfaction With Their Care*. London: Royal College of Physicians.

Fletcher RH, Fletcher SW [1979]. Clinical research in general medical journals. a 30 year perspective. *New England Journal of Medicine*, **vol. 301**, pp180-183.

Frieman JA, Chalmers TC, Smith H, *et al.* [1978]. The importance of beta, the type II error and sample size in the design and interpretation of the randomised control trial. *The New England Journal of Medicine*, **vol. 299**, pp690-694.

Froberg DG, Kane RL [1989A]. Methodology for measuring health- state preferences - I: measurement strategies. *Journal of Clinical Epidemiology*, **vol. 42**, pp345-354.

Froberg DG, Kane RL [1989B]. Methodology for measuring health- state preferences - II: scaling methods. *Journal of Clinical Epidemiology*, **vol. 42**, pp459-471.

Furlong W, Feeny D, Torrance GW, *et al.* [1990]. Guide to design and development of health-state utility instrumentation. *CHEPA Working paper No.90-9*, Hamilton, Ontario: CHEPA, McMaster University.

Gafni A [1994]. The standard gamble method: what is being measured and how it is interpreted. *Health Services Research*, **vol. 29**, pp207-224.

G fni A [1995]. Time in health: can we measure individuals' "pure time preferences"? *Medical Decision Making*, **vol. 15**, pp31- 37.

Gafni A [1991]. Willingness to pay as a measure of benefits. *Medical Care*, **vol. 29**, pp1246-1252.

Gafni A, Birch S, Mehrez A [1993]. Economics, health and health economics: HYEs versus QALYs. *Journal of Health Economics*, **vol. 11**, pp325-339.

Gafni A, Birch S, O'Brien B [1995]. Healthy years equivalents (HYEs) and decision trees: a simple, two-stage lottery-based algorithm. *CHEPA Working Paper 95-4*, Hamilton, Ontario: CHEPA, McMaster University.

344

Garratt AM, Ruta DA, Abdalla MI, *et al.* [1993]. The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *British Medical Journal*, vol. **306**, pp1440-1444.

Garry R, Reich H, Lui CY [1994]. Laparoscopic hysterectomy: definitions and indications. *Gynaecological Endoscopy*, vol. **3**, pp1-3.

Gath D, Osborn M, Bungay G, *et al.* [1987]. Psychiatric disorder and gynaecological symptoms in middle aged women: a community survey. *British Medical Journal*, vol. **294**, pp213-218.

Gelijns AC, Fendrick AM [1993]. The dynamics of innovation in minimally invasive therapy. *Health Policy*, vol. **23**, pp153-166.

Gerard K [1992]. Cost-utility in practice: a policy maker's guide to the state of the art. *Health Policy*, vol. **21**, pp249- 279.

Gerard K, Dobson M, Hall J [1993]. Framing and labelling effects in health descriptions: quality adjusted life years for treatment of breast cancer. *Journal of Clinical Epidemiology*, vol. **46**, pp77-84.

Gerard K, Mooney G [1993]. QALY league tables: handle with care. *Health Economics*, vol. **2**, pp59-64.

Giles JA, Gafar A [1991]. The treatment of CIN: do we need lasers? *British Journal of Obstetrics and Gynaecology*, vol. **98**, pp3-6.

Glaucoma Laser Trial Research Group [1990]. The Glaucoma laser trial (GLT). 2. Results of argon laser trabeculoplasty versus topical medicines. *Ophthalmology*, vol. **97**, pp1403-1413.

Goldrath MH, Fuller TA, Segal S [1981]. Laser photovaporization of endometrium for the treatment of menorrhagia. *American Journal of Obstetrics and Gynaecology*, vol. **140**, pp14-19.

Goodman C [1992]. *Percutaneous Transluminal Coronary Angioplasty in Coronary Revascularization: Evidence, Assessment and Policy*. Stockholm: Swedish Council on Technology Assessment in Health Care.

Graboys TB, Biegelsen B, Lampert S, *et al.* [1992]. Results of a second-opinion trial among patients recommended for coronary angiography. *Journal of the American Medical Association*, vol. **268**, pp2537-2540.

Grant JM, Hussein IY [1984]. An audit of abdominal hysterectomy over a decade in a district hospital. *British Journal of Obstetrics and Gynaecology*, vol. **91**, pp73-77.

Gray DT, Thorburn J, Lundorff P, *et al.* [1995]. A cost- effectiveness study of a

randomised trial of laparoscopy versus laparotomy for ectopic pregnancy [see comments]. *Lancet,* vol. **345**, pp1139-43.

Grech ED, Ramsdale DR [1993]. Angioplasty and acute myocardial infarction. *Lancet,* vol. **342**, pp191-192.

Greenfield S, Kaplan S, Ware JE [1985]. Academia and clinic: expanding patient involvement in care; effects on patient outcomes. *Annals of Internal Medicine,* vol. **102**, pp520-528.

Guyatt G, Drummond M, Feeny D, *et al.* [1986]. Guidelines for health technology assessment: therapeutic technologies. In: Feeny D, Guyatt G, Tugwell P. *Health Care Technology: Effectiveness, Efficiency and Public Policy,* Montreal: Institute of Research on Public Policy.

Hall J, Gerard K, Salkeld G, *et al.* [1992]. A cost utility analysis of mammography screening in Australia. *Social Science and Medicine,* vol. **34**, pp993-1004.

Hannan EL, Kilburn H, Bernard H, *et al.* [1991]. Coronary artery bypass surgery: the relationship between inhospital mortality rate and surgical volume after controlling for clinical risk factors. *Medical Care,* vol. **29**, pp1094-1107.

Hawkins J, Hudson C [1977]. *Shaw's Textbook of Operative Gynaecology.* London: Churchill Livingstone.

Henderson RA [1989]. The randomised intervention treatment of angina (RITA) trial protocol: a long term study of coronary angioplasty and coronary artery bypass surgery in patients with angina. *British Heart Journal,* vol. **62**, pp411-414.

Henry D [1992]. Economic analysis as an aid to subsidisation decisions: the development of Australian guidelines for pharmaceuticals. *PharmacoEconomics,* vol. **1**, pp54-67.

Henshaw RC, Naji SA, Russell IT, *et al.* [1993]. Comparison of medical abortion w'th surgical vacuum aspiration: women's preferences and acceptability of treatment. *British Medical Journal,* vol. **307**, pp714-717.

Hirsch NA [1994]. *Minimal Access Surgery: An Update,* Canberra: Australian Institute of Health and Welfare.

Hirsch NA, Hailey DM [1992]. *Minimal Access Surgery.* Canberra: Australian Institute of Health.

Hlatky MA [1995]. Patient preferences and clinical guidelines. *Journal of the American Medical Association,* vol. **273**, pp1219- 1220.

HM Treasury [1991]. *Economic Appraisal in Central Government: a Technical Guide*

for Government Departments. London: HMSO.

Hunink MGM, Wong JB, Donaldson MC, *et al.* [1995]. Revascularization for femoropopliteal disease: a decision and cost-effectiveness analysis. *Journal of the American Medical Association*, vol. **274**, pp165-171.

Hunt SM, McEwen J, McKenna SP [1986]. *Measuring Health Status*, Beckenham: Croom Helm.

Hunter JG [1991]. Laser or electrocautery for laparoscopic cholecystectomy. *American Journal of Surgery*, vol. **161**, pp345-9.

Hunter RW, McCartney AJ [1993]. Can laparoscopic assisted hysterectomy safely replace abdominal hysterectomy? *British Journal of Obstetrics and Gynaecology*, vol. **100**, pp932-934.

Jaeschke R, Sackett DL [1989]. Reseach methods for obtaining primary evidence. *International Journal of Technology Assessment in Health Care*, vol. **5**, pp503-519.

Jeans WD, Danton RM, Baird RN, *et al.* [1986]. The effects of introducing balloon dilatation into vascular surgical practice. *British Journal of Radiology*, vol. **59**, pp457-459.

Jenkinson C, Coulter A, Wright L [1993]. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *British Medical Journal*, vol. **306**, pp1437-1440.

Jenkinson C, Peto V, Coulter A [1994]. Measuring change over time: a comparison of results from a global single item of health status and the multi-dimensional SF-36 health status survey questionnaire in patients presenting with menorrhagia. *Quality of Life Research*, vol. **3**, pp317-321.

Jenkinson C, Peto V, Coulter A [1996]. Making snese of ambiguity: evaluation of internal reliability and face validity of the SF36 questionnaire in women presenting with menorrhagia. *Quality in Health Care*, vol. **5**, pp9-12.

Johannesson M [1993]. Commentary: the contingent valuation method. *Medical Decision Making*, vol. **13**, pp311-312.

Johannesson M [1994]. QALY's, HYE's and individual preferences - a graphical illustration. *Social Science and Medicine*, vol. **39**, pp1623-1632.

Johnnesson M [1995]. The ranking properties of healthy-years equivalents and quality-adjusted life-years under certainty and uncertainty. *International Journal of Technology Assessment in Health Care*, vol. **11**, pp40-48.

Johannesson M, Pliskin JS, Weinstein MC [1993]. Are healthy- years equivalents

an improvement over quality-adjusted life years? *Medical Decision Making*, **vol. 13**, pp281-286.

Johnson A [1994]. Minimally invasive surgery. Decreasing hospital stay is nothing new. (Letter). *British Medical Journal*, **vol. 308**, pp341.

Jones-Lee MW [1976]. *The Value of Life: An Economic Analysis*. Chicago: University of Chicago Press.

Kahn HA, Sempos CT [1989]. *Statistical Methods in Epidemiology*. New York: Oxford University Press.

Kaplan RM [1995]. Utility assessment for estimating quality- adjusted life years. In: Sloan FA, *Valuing Health Care. Costs, Benefits and Effectiveness of Pharmaceuticals and other medical technologies*. Cambridge: Cambridge University Press.

Kaplan RM, Anderson JP [1990]. The general health policy model: an integrated approach. In: Splilker B (ed), *Quality of Life Assessments in Clinical Trials*. Raven Press.

Keeler AB, Cretin S [1983]. Discounting of Life- Saving and Other Nonmonetary Effects. *Management Science*, **vol. 29**, pp300- 306.

Kesteloot K, Penninckx F [1993]. The costs and effects of open versus laparoscopic cholecystectomies. *Health Economics*, **vol. 2**, pp303-312.

Kind P, Rosser R, Williams A [1982]. Valuation of quality of life: some psychometric evidence. In: Jones-Lee MW (ed), *The value of life and safety*. Amsterdam: North Holland.

Klein R [1989]. *The Politics of the NHS*, London: Longman.

Koopmanschap MA, Rutten FFH [1994]. The impact of indirect costs on outcomes of health care programs. *Health Economics*, **vol. 3**, pp385-393.

Koopmanschap MA, van Ineveld BM [1992]. Towards a new approach for estimating indirect costs of disease. *Social Science and Medicine*, **vol. 34**, pp1005-1010.

Krabbe PFM, Essink-Bot ML, Bonsel GJ [1996]. On the equivalence of collectively and individually collected responses: standard gamble and time trade-off judgements of health states. *Medical Decision Making*, **vol. 16**, pp120-132.

Labelle RJ, Churchill DN, Martin S, *et al.* [1987]. Economic evaluation of extracorporeal shock wave lithotripsy, percutaneous ultrasonic lithotripsy, and standard surgical treatment of urolithiasis - a Canadina perspective. *Clinical and*

Investigative Medicine, vol. **10**, pp86-95.

Laffel GL, Fineberg HV, Braunwald E [1987]. A cost-effectiveness model for coronary thrombolysis/reperfusion therapy. *Journal of the American College of Cardiology*, vol. **10**, pp79B-90B.

Lam CM, Murray FE, Cuschieri A [1996]. Increased cholesystectomy rate after the introduction of laparoscopic cholecystectomy in Scotland. *Gut*, vol. **38**, pp282-4.

Lancet [1993]. Surgical innovation under scrutiny. *Lancet*, vol. **342**, pp187-188.

Laupacis A, Feeny D, Detsky AS, *et al*. [1992]. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Canadian Medical Association Journal*, vol. **146**, pp473-481.

Leese B, Hutton J, Maynard A [1992]. A comparison of the costs and benefits of recombinant human erythropoietin (epoetin) in the treatment of chronic renal failure in 5 European countries. *PharmacoEconomics*, vol. **1**, pp346-356.

Legorreta AP, Silber JH, Costantino GN, *et al*. [1993]. Increased cholecystectomy rate after the introduction of laparoscopic cholecystectomy. *Journal of the American Medical Association*, vol. **270**, pp1429-1432.

Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, *et al*. [1984]. Describing health states: methodologic issue in obtaining values for health states. *Medical Care*, vol. **22**, pp543-552.

Loewenstein GF, Prelec D [1993]. Preferences for sequences of outcomes. *Psychological Review*, vol. **100**, pp91-108.

Loomes G [1995]. The myth of the HYE. *Journal of Health Economics, vol. 14*, pp1-7.

Loomes G, McKenzie L [1989]. The use of QALYs in health care decision making. *Social Science and Medicine*, vol. **28**, pp299- 308.

Luce BR, Elixhauser A [1990]. *Standards for Socioeconomic Evaluation of Health Care Products and Services*. Berlin: Springer-Verlag.

Luft HS, Garnick HW, Mark DH *[1990]. Hopsital Volume, Physician Volume and Patient Outcomes: Assessing the Evidence*. Ann Arbor, MI: Health Administration Press Perspectives.

Luft HS, Hunt SS, Maerki SC [1987]. The volume-outcome relationship: practice makes perfect or selective-referral patterns. *Health Services Research*, vol. **22**, pp157-182.

Lundorff P, Thorburn J, Hahlin M, *et al.* [1991]. Laparoscopic surgery in ectopic pregnancy: a randomised trial versus laparotomy. *Acta Obstet Gynecol Scand*, vol. 70, pp343-348.

Luoto R, Kaprio J, Uutela A [1994]. Age at natural menopause and sociodemographic status in Finland. *American Journal of Epidemiology*, vol. 139, pp64-76.

MacIntyre IMC [1991]. Tribulations for clinical trials - poor recruitment is hampering research. *British Medical Journal*, vol. 302, pp1099-1100.

MacRae KD [1989]. Pragmatic versus explanatory trials. *International Journal of Technology Assessment in Health Care*, vol. 5, pp333-339.

Magos A, Baumann R, Turnbull AC [1989]. Transcervical resection of the endometrium in women with menorrhagia. *British Medical Journal*, vol. 298, pp1209-1212.

Makinen AM, Nordback IH [1995]. Cholecystectomy: comparison of minilaparotomy and laparoscopy. *Int Surg*, vol. 80, pp99-101.

Malenka DJ, Roos N, Fisher ES, *et al.* [1990]. Further study of the increased mortality following transurethral prostatectomy: a chart-based analysis. *Journal of Urology*, vol. 144, pp224-228.

Manyonda IT, Varma TR [1991]. The advent of minimally invasive procedures in gynaecological surgery. *Maternal and Child Health*, pp320-324.

Marshall D, Hailey D, Hirsch N, *et al.* [1994]. *The Introduction of Laparoscopic Cholecystectomy in Canada and Australia*. Canberra: Australian Institute of Health & Welfare.

Mason J, Drummond M, Torrance G [1993]. Some guidelines on the use of cost effectiveness league tables. *British Medical Journal*, vol. 306, pp570-572.

Maynard A [1991]. Developing the health care market. *Economic Journal*, vol. 101, pp1277-1286.

McCloy R [1992]. Through the Keyhole. *Health Service Journal*, vol. 12th November, pp26-27.

McMahon AJ, O'Dwyer PJ, Russell IT, *et al.* [1992]. Laparoscopic versus open cholecystectomy and the need for a randomised trial: a survey of surgeons and ethical committees in the British Isles. *Journal of Laparoendoscopic Surgery*, vol. 2, pp277-280.

McMahon AJ, Russell IT, Baxter JN [1994]. Laparoscopic versus minilaparotomy cholecystectomy: a randomised trial. *Lancet*, **vol. 343**, pp135-138.

McNeil BJ, Weichselbaum R, Pauker SG [1978]. Fallacy of the five-year survival in lung cancer. *New England Journal of Medicine*, **vol. 299**, pp1397-1401.

McNeil BJ, Weichselbaum R, Pauker SG [1981]. Speech and survival: trade-offs between quality and quantity of life in laryngeal cancer. *New England Journal of Medicine*, **vol. 305**, pp982-987.

McNeil BJ, Pauker SG, Sox HC, *et al.* [1982]. On the elicitation of preferences for alternative therapies. *New England Journal of Medicine*, **vol. 306**, pp1259-1262.

Measurement and Valuation of Health Group [1994]. *The Measurement and Valuation of Health. First Report on the Main Survey.* York: CHE, University of York.

Medical Research Council [1993]. Health technology assessment in surgery: the role of the randomised controlled trial. *Medical Research Council*, London: MRC.

Mehrez A, Gafni A [1989]. Quality-adjusted life years, utility theory and healthy-years equivalents. *Medical Decision Making*, **vol. 9**, pp142-149.

Mehrez A, Gafni A [1991]. The health-years equivalent: how to measure them using standard gamble approach. *Medical Decision Making*, **vol. 11**, pp140-146.

Mehrez A, Gafni A [1993]. Healthy-years equivalents versus quality-adjusted life years: pursuit of progress. *Medical Decision Making*, **vol. 13**, pp287-292.

Messahel FM [1995]. Post cholecystectomy admission to the intensive care unit. Comparison between open, mini-lap and laparoscopic techniques. *Anaesthesia*, **vol. 50**, pp901-4.

Messina MJ [1995]. Laparoscopy-assisted vaginal hysterectomy: cost analysis and review of initial experience in a community hospital. *JAOA*, **vol. 95**, pp31-36.

Miller JN, Colditz GA, Mosteller F [1989]. How study design affects outcomes in comparisons of therapy. II: surgical. *Statistics in Medicine*, **vol. 8**, pp455-466.

Ministry of Health [1994]. *Ontario guidelines for economic analysis of pharmaceutical products.* Ontario: Ministry of Health.

Mishan E [1971A]. *Cost-Benefit Analysis.* London: Allen and Unwin.

Mishan EJ [1971B]. Evaluation of life and limb: a theoretical approach. *Journal of Political Economy*, **vol. 79**, pp687-705.

Miyamoto JM, Eraker,SA [1985]. Parameter estimates for a QALY utility model. *Medical Decision Making*, **vol. 5**, pp191-213.

Mohide EA, Torrance GW, Streiner DL, *et al.* [1988]. Measuring the wellbeing of family caregivers using the time trade-off technique. *Journal of Clinical Epidemiology*, **vol. 41**, pp475-482.

Moody FG [1992]. The changing health care economy: impact on surgical techniques. In: Gelijns AC, *Technology and Health Care in an Era of Limits*. Wahington DC: National Academy Press.

Mooney G, McGuire A [1988]. Economics and medical ethics: an economic viewpoint. In: Mooney G, McGuire A (eds), *Medical Ethics and Economics in Health Care*. Oxford: Oxford University Press.

Mooney G, Olsen JA [1991]. QALYs: where next? In: McGuire A, Fenn P, Mayhew K (eds). *Providing Health Care: The Economics of Alternative Systems of Finance and Delivery*, Oxford University Press.

Mosteller F, Falotico-Taylor J [1989]. *Quality of Life and Technology Assessment*, Washington DC: Institute of Medicine.

Mulley AG [1989]. Assessing patients' utilities can the ends justify the means? *Medical Care*, **vol. 27**, ppS269-S281.

Nease RF, Owens DK [1994]. A method of estimating the cost- effectiveness of incorporating patient preferences into practice guidelines. *Medical Decision Making*, **vol. 14**, pp382-392.

Netten A [1994]. *Unit Costs of Community Care*. Canterbury: PSSRU, University of Kent.

Neumann PJ, Weinstein MC [1991]. The diffusion of new technology: costs and benefits to health care. in: Gelijns AC, Halm EA. *The Changing Economics of Medical Technology*, Washington: National Academy Press. pp21-34.

Nezhat C, Admon D, Nezhat CH, *et al.* [1994]. Hospital cost comparison between abdominal, vaginal, and laparoscopy-assisted vaginal hysterectomies. *Obstetrics and Gynaecology*, **vol. 83**, pp713-716.

NHS Centre for Reviews and Dissemination [1995]. Relationship between volume and quality of health Care: a review of the literature. *CRD Report 2*, York: NHS Centre for Reviews and Dissemination, University of York.

NHS Management Executive [1993]. *Costing for Contracting*. Leeds: Department of Health.

NIH Consensus Development Panel [1993]. Gallstones and laparoscopic cholecystectomy. *JAMA*, **vol. 269**, pp1018-1024.

Nishioka NS, Richter JM [1987]. Endoscopic therapy of bleeding peptic ulcers: a cost-benefit analysis. *Gastrointestinal Endoscopy*, **vol. 33**, pp277-83.

Nord E [1991]. The validity of a visual analogue scale in determining social utility weights for health states. *International Journal of Health Planning and Management*, **vol. 6**, pp234-242.

O'Brien BJ, Anderson DR, Goeree R [1994]. Cost-effectiveness of enoxaparin versus warfarin prophylaxis against deep-vein thrombosis after total hip replacement. *Journal of Canadian Medical Association*, **vol. 150**, pp1083-1090.

O'Brien B, Gafni A [1996]. When do "dollars" make sense. Toward a conceptual framework for contingent valuation studies in health care. *Medical Decision Making*, **vol. 16**, pp288-299.

O'Brien B, Viramontes JL [1994]. Willingness to pay: a valid and reliable measure of health state preference? *Medical Decision Making*, **vol. 14**, pp289-297.

O'Brien BJ, Novosel S, Torrance G, *et al.* [1995]. Assessing the economic value of a new antedepressant. A willingness to pay approach. *PharmacoEconomics*, **vol. 8**, pp34-45.

O'Connor AMC, Boyd NF, Warde P, *et al.* [1987]. Eliciting preferences for alternative drug therapies in oncology: influence of treatment outcome description, elicitation technique and treatment experience on preferences. *Journal of Chronic Diseases*, **vol. 40**, pp811-818.

OECD [1996]. *OECD Health Data 1996*. Paris: OECD.

Office of Technology Assessment [1982]. *Strategies for Medical Technology Assessment*, Washington DC: OTA.

Oster G, Epstein AM [1987]. Cost-effectiveness of antihyperlipemic therapy in the prevention of coronary heart disease. *Journal of the American Medical Association*, **vol. 258**, pp2381-2387.

Paimela H, Tuompo PK, Perakyla T, *et al.* [1991]. Peptic ulcer surgery during the H2 receptor antagonist era: a population-based epidemiological study of ulcer surgery in Helsinki from 1972 to 1987. *British Journal of Surgery*, **vol. 78**, pp28-31.

The Panel on Cost-Effectiveness in Health and Medicine [1996]. *Cost-Effectiveness in Health and Medicine*. Washington: Office of Public Health and Science, US Public Health Service.

Parsonage M, Neuberger H [1992]. Discounting and health benefits. *Health Economics*, **vol. 1**, pp71-6.

Passamani E [1991]. Clinical trials - are they ethical. *New England Journal of Medicine*, **vol. 324**, pp1589-1592.

Pauly MV [1995]. Valuing health benefits in monetary terms. In: Sloan FA, *Valuing Health Care. Costs, Benefits and Effectiveness of Pharmaceuticals and Other Medical Technologies*. Cambridge: Cambridge University Press.

Pearce D, Markandya A, Barber EB [1989]. *Blueprint for a Green Economy*. London: Earthscan Publications.

Pearson V [1994]. *Minimal access surgery: a review*. Bristol: Health Care Evaluation Unit, University of Bristol.

Petitti DB [1994]. *Meta Analysis, Decision Analysis and Cost- Effectiveness Analysis. Methods for Quantitative Synthesis in Medicine*. New York: Oxford University Press.

Peto V, Coulter A, Bond A [1993]. Factors affecting general practitioners' recruitment of patients into a prospective study. *Family Practice*, **vol. 10**, pp207-211.

Petrou S, Malek M, Davey PG [1993]. The reliability of cost- utility estimates in cost-per-QALY league tables. *PharmacoEconomics*, **vol. 3**, pp345-353.

Phipps JH, Lewis BV, et al. [1990]. Treatment of functional menorrhagia by radiofrequency-induced thermal endometrial ablation. *Lancet*, **vol. 335**, pp374-376.

Pliskin JS, Shepard DS, Weinstein MC [1980]. Utility functions for life years and health status. *Operations Research*, **vol. 28**, pp206-224.

Pocock SJ [1983]. *Clinical Trials: A Practical Approach*, Chichester: Wiley.

Pocock SJ, Henderson RA, Rickards AF, *et al.* [1995]. Meta- analysis of randomised trials comparing coronary angioplasty with bypass surgery. *Lancet*, **vol. 346**, pp1184-1189.

Propper C [1988]. Estimation of the value of time spent on NHS waiting lists using stated preference methodology. *Centre for Health Economics Discussion Paper No. 49*, York: CHE, University of York.

Raju KS, Auld BJ [1994]. A randomised prospective study of laparoscopic vaginal hysterectomy versus abdominal hysterectomy each with bilateral salpingo-oophorectomy. *British Journal of Obstetrics and Gynaecology*, **vol. 101**,

pp1068-1071.

Ratcliffe J [1995]. The measurement of indirect costs and benefits in health care evaluation: a critical review. *Project Appraisal*, **vol. 10**, pp13-18.

Reddick EJ, Olsen DO [1989]. Laparoscopic laser cholecystectomy. A comparison with mini-lap cholecystectomy. *Surgical Endoscopy*, **vol. 3**, pp1-3.

Redelmeier DA, Heller DN [1993]. Time preference in medical decision making and cost-effectiveness analysis. *Medical Decision Making*, **vol. 13**, pp212-217.

Richardson J [1994]. Cost utility analysis: what should be measured? *Social Science and Medicine*, **vol. 39**, pp7-21.

Richardson J, Hall J, Salkeld G [1989]. Cost utility analysis: the compatibility of measurement techniques and the measurement of utility through time. In: Smith CS (ed), *Economics and Health: Proceedings of the 11th Conference of Health Economics*. Melbourne: Monash University.

Richardson J, Hall J, Salkeld G [1996]. The measurement of utility in multiphase health states. *International Journal of Technology Assessment in Health Care*, **vol. 12**, pp151-162.

Richardson RE, Bournas N, Magos AL [1995]. Clincial practice: is laparoscopic hysterectomy a waste of time? *Lancet*, **vol. 345**, pp36-41.

Ridley S, Biggam M, Stone P [1991]. Cost of intensive therapy - a description of methodology and initial results. *Anaesthesia*, **vol. 46**, pp523-530.

RITA Trial Participants [1993]. Coronary angioplasty versus coronary artery bypass surgery: the randomised intervention treatment of angina (RITA) trial. *Lancet*, **vol. 341**, pp573-580.

Robinson R [1993]. Cost-utility analysis. *British Medical Journal*, **vol. 307**, pp859-862.

Roos LL [1989]. Nonexperimental data systems in surgery. *International Journal of Technology Assessment in Health Care*, **vol. 5**, pp341-356.

Roos NP, Wennberg JE, Malenka DJ, *et al.* [1989]. Mortality and reoperation after open transurethral resection of the prostate for benign prostatic hyperplasia. *New England Journal of Medicine*, **vol. 320**, pp1120-1124.

Royal College of Surgeons of England [1992]. *Reports of the Working Party for Day Case Surgery*. London: Royal College of Surgeons of England.

Royal College of Obstetricians and Gynaecologists (RCOG) Medical Audit Unit

[1991]. *Third Bulletin*. Manchester: RCOG Medical Audit Unit.

Rubins HB [1994]. From clinical trials to clinical practie: generalizing from participant to patient. *Controlled Clinical Trials*, vol. **15**, pp7-10.

Ruta DA, Garratt AM, Chadha YC, *et al.* [1995]. Assessment of patients with menorrhagia: how valid is a structured clinical history as a measure of health status? *Quality of Life Research*, vol. **4**, pp33-40.

Rutherford AJ, Glass MR [1990]. Management of Menorrhagia. *British Medical Journal*, vol. **301**, pp290-291.

Ryan M, Shackley P [1995]. Viewpoint: assessing the benefits of health care: how far should we go? *Quality in Health Care*, vol. **4**, pp207-213.

Sackett DL, Torrance GW [1978]. The utility of different health states as perceived by the general public. *Journal of Chronic Disease*, vol. **31**, pp697-704.

Sacks H, Chalmers TC, Smith H [1982]. Randomised versus historical controls for clinical trials. *American Journal of Medicine*, vol. **72**, pp233-240.

Schoemaker PJH [1982]. The expected utility model: its variants, purposes, evidence and limitations. *Journal of Economic Literature*, vol. **20**, pp529-563.

Scottish Hysteroscopy Audit Group [1995]. A Scottish audit of hysteroscopic surgery for menorrhagia: complications and follow up. *British Journal of Obstetrics & Gynaecology*, vol. **102**, pp249-254.

Sculpher MJ [1994]. Phase II Medical Laser Technology Assessment. An economic evaluation of the Nd:YAG laser in gastroenterology. *HERG Research Report 14*, Uxbridge: HERG, Brunel University.

Sculpher MJ [1993]. A preliminary economic evaluation of the diode laser in ophthalmology. *Lasers in Medical Science*, vol. **8**, pp163-169.

Sculpher MJ, Bryan S, Dwyer N, *et al.* [1993]. An economic evaluation of transcervical endometrial resection versus abdominal hysterectomy for the treatment of menorrhagia. *British Journal of Obstretrics and Gynaecology*, vol. **100**, pp244-252.

Sculpher MJ, Buxton MJ [1993]. Th episode-free day as a composite measure of effectiveness. An illustrative economic evaluation of formoterol versus salbutamol on asthma therapy. *PharmacoEconomics*, vol. **4**, pp345-352.

Sculpher MJ, Drummond MF, Buxton MJ [forthcoming]. The iterative use of economic evaluation as part of the process of health technology assessment. *Journal of Health Servies Research and Policy*,

Sculpher MJ, Dwyer N, Byford S, *et al.* [1996A]. Randomised trial comparing hysterectomy and transcervical endometrial resection: effect on health-related quality of life and costs two years after surgery. *British Journal of Obstetrics and Gynaecology*, vol. **103**, pp142-149.

Sculpher MJ, Michaels J, McKenna M, *et al.* [1996B]. A cost-utility analysis of laser-assisted angioplasty for peripheral arterial occlusions. *International Journal of Technology Assessment in Health Care*, vol. **12**, pp104-125.

Sculpher MJ, Seed P, Henderson RA, *et al.* [1994]. Health service costs of coronary angioplasty and coronary artery bypass surgery: the randomised intervention treatment of angina (RITA) trial. *Lancet*, vol. **344**, pp927-929.

See WA, Cooper CS, Fisher RJ [1993]. Predictors of laparoscopic complications after formal training in laparoscopic surgery. *Journal of American Medical Association*, vol. **270**, pp2689-2692.

Shanahan D, Knight M [1992]. Laparoscopic cholecystectomy. *British Medical Journal*, vol. **304**, pp776-777.

Sheldon TA, Faulkner A [1996]. Vetting new technologies. *British Medical Journal*, vol. **313**, pp508.

Shepperd S, Coulter A, Farmer A [1995]. Using interactive videos in general practice to inform patients about treatment choices: a pilot study. *Family Practice*, vol. **12**, pp443-447.

Shiell A, Seymour J, Cameron S [1995]. QALYs, risk-adjusted QALYs and HYEs: is there a difference? Paper presented to the UK Health Economists Study Group: University of Aberdeen.

Showstack JA, Rosenfeld KE, Garnick DW, *et al.* [1987]. Association of volume with outcomes of coronary artery bypass graft surgery. *Journal of the American Medical Association*, vol. **257**, pp785-789.

Siddle N, Sarrel P, Whitehead M [1987]. The effect of hysterectomy on the age at ovarian failure: identification of a subgroup of women with premature loss of ovarian function and literature review. *Fertility and Sterility*, vol. **47**, pp94-100.

Sidney S, Quesenberry CP, Sadler MC, *et al.* [1992]. Reoperation and mortality after surgical treatment of benign prostatic hypertrophy in a large prepaid medical care program. *Medical Care*, vol. **30**, pp117-125.

Singer A, Almanza R, Gutierrez A, *et al.* [1994]. Preliminary clinical experience with a thermal balloon endometrial ablation method to treat menorrhagia. *Obstetrics and Gynaecology*, vol. **83**, pp732-734.

Sloan F (ed) [1995]. *Valuing Health Care: Costs, Benefits and Effectiveness of Pharmaceuticals and Other Medical Technologies*. Cambridge: Cambridge University Press.

Smith R [1991]. Injuries to common bile duct during laparoscopic cholecystectomy. *British Medical Journal*, **vol. 303**, pp1475.

Sonnenberg FA, Beck JR [1993]. Markov models in medical decision making. *Medical Decision Making*, **vol. 13**, pp322-338.

Special Medical Development Project Team [1990]. *Diabetic Retinopathy Screening Project: Sepecial Medical Development at Exeter, Oxford and Sheffield. HERG Research Report No. 8*. Uxbridge: HERG, Brunel University.

Spilker B [1990]. *Quality of Life Assessments in Clinical Trials*, New York: Raven Press.

Spiro HM [1992]. Diagnostic laparoscopic cholecystectomy. *Lancet*, **vol. 339**, pp167-168.

Steiner CA, Bass EB, Talamini MA, *et al.* [1994]. Surgical rates and operative mortality for open and laparoscopic cholecystectomy in Maryland. *New England Journal of Medicine*, **vol. 330**, pp403- 408.

Still RM, Walsh DJ [1992]. A pilot study to assess the learning curve in transcervical endometrial resection; possible implications for postgraduate accreditation. *Gynaecological Endoscopy*, **vol. 1**, pp111-113.

Stirrat GM, Dwyer N, Browning J [1990]. Planned trial of transcervical resection of the endometrium versus hysterectomy (Letter). *British Journal of Obstetrics and Gynaecology*, **vol. 97**, pp459.

Stirrat GM, Farndon J, Farrow SC, *et al.* [1992]. The challenge of evaluating surgical procedures. *Annals of the Royal College of Surgeons of England*, **vol. 74**, pp80-84.

Stoker ME, Vose J, O'Mara P, *et al.* [1992]. Laparoscopic cholecystectomy: a clinical and financial analysis of 280 operations. *Archives of Surgery*, **vol. 127**, pp589-595.

Stoker DL, Spiegelhalter DJ, Wellwood JM [1994]. Laparoscopic versus open inguinal hernia repair: randomised prospective trial. *Lancet*, **vol. 343**, pp1243-1245.

Streiner DL, Norman GR [1989]. *Health Measurement Scales: a Practical Guide to their Development and Use*, Oxford: Oxford University Press.

Strull WM, Lo B, Charles G [1984]. Do patients want to participate in medical decision making? *Journal of American Medical Association*, vol. 252, pp2990-2995.

Sutherland HJ, Llewellyn-Thomas H, Boyd NF, *et al.* [1982]. Attitudes toward quality of survival: the concept of "maximal endurable time". *Medical Decision Making*, vol. 2, pp299-309.

Sutton C [1993]. Laparoscopic surgery in gynaecology. *Hospital Update*, pp143-152.

Sutton CJG, Ewen SP, Whitelaw N, *et al.* [1994]. Prospective, randomized, double-blind, controlled trial of laser laparoscopy in the treatment of pelvic pain associated with minimal, mild, and moderate endometriosis. *Fertility and Sterility*, vol. 62, pp696-700.

Sutton CJG, Ewen SP [1994]. Thinning the endometrium prior to ablation: is it worthwhile? *British Journal of Obstetrics and Gynaecology*, vol. 101, pp10-12.

Taylor KM, Margolese RG, Soskolne CI [1984]. Physicians' reasons for not entering eligible patients in a randomised clinical trial of surgery for breast cancer. *New England Journal of Medicine*, vol. 310, pp1363-1367.

Thompson SG [1994]. Systematic Review: why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal*, vol. 309, pp1351-1355.

Times [1992]. Woman died after keyhole surgery. *Times*, vol. December 8th, pp4.

Torrance GW [1976]. Social preferences for health states: an empirical evaluation of three measurement techniques. *Socio- Economic Planning Sciences*, vol. 10, pp129-136.

Torrance GW [1986]. Measurement of health state utilities for economic appraisal - a review. *Journal of Health Economics*, vol. 5, pp1-30.

Torrance GW [1987]. Utility approach to measuring health-related quality of life. *Journal of Chronic Diseases*, vol. 40, pp593- 600.

Torrance GW, Thomas WH, Sackett DL [1972]. A utility maximisation model for evaluation of health care programs. *Health Services Research*, vol. 7, pp118-32.

Torrance GW, Feeny DH, Furlong WJ, *et al.* [1996]. Multiattribure utility function for a comprehensive health status classification system: Health Utilities Index Mark 2. *Medical Care*, vol. 34, pp702-722.

Trondsen E, Reiertsen O, Andersen OK, *et al.* [1993]. Laparoscopic and open

cholecystectomy. *European Journal of Surgery*, **vol. 159**, pp217-221.

Tulandi T [1986]. Salpingo-ovariolysis: a comparison between laser surgery and electrosurgery. *Fertility and Sterility*, **vol. 45**, pp489-491.

Tunis SR, Bass EB, Steinberg E [1991]. The use of angioplasty, bypass surgery and amputation in the management of peripheral vascular disease. *New England Journal of Medicine*, **vol. 325**, pp556-562.

van der Linden W [1980]. Pitfalls in randomised surgical trials. *Surgery*, **vol. 87**, pp258-262.

Veith FJ, Gupta SK, Wengerter KR, *et al.* [1990]. Changing arteriosclerotic disease patterns and management strategies in lower-limb-threatening ischemia. *Annals of surgery*, **vol. 212**, pp402-414.

Vessey MP, Villard-Mackintosh L, McPherson K, *et al.* [1992]. The epidemiology of hysterectomy: findings in a large cohort study. *British Journal of Obstetrics and Gynaecology*, **vol. 99**, pp402- 407.

Vilos GA, Pispidikis JT, Botz CK [1996]. Economic evaluation of hysteroscopic endometrial ablation versus vaginal hsyterectomy for menorrhagia. *Obstsetrics and Gynaecology*, **vol. 88**, pp1-5.

Vitale GC, Collet D, Larson GM, *et al.* [1991]. Interruption of professional and home activity after laparoscopic cholecystectomy among French and American patients. *American Journal of Surgery*, **vol. 161**, pp396-398.

von Neumann J, Morgenstern O [1944]. *Theory of Games and Economic Behavior.* New York: Princeton University Press.

Vondeling H, Mathus-Vliegen E, Banta D [1991]. Palliation of advanced colon cancer. In: Banta D, Shou I. *Lasers in Health Care: Effectiveness, Cost-Effectiveness and Policy Implications*, Denmark: Academic Publishing.

Voyles CR, Meena AL, Petro AB, *et al.* [1990]. Electrocautery is superior to laser for laparoscopic cholecystectomy. *American Journal of Surgery*, **vol. 160**, pp484.

Vroegindeweij D, Kemper FJM, Tielbeek AV, *et al.* [1992]. Recurrence of stenoses following balloon angioplasty and Simpson atherectomy of the femoro-popliteal segment. A randomised comparative 1-year follow-up study using color flow Duplex. *European Journal Vascular Surgery*, **vol. 6**, pp164-171.

Wakker P [1996]. A critisism of healthy-years equivalents. *Medical Decision Making*, **vol. 16**, pp207-214.

Warner MM, Pugh S, Riley C, *et al.* [1993]. Blurring the Boundaries: the Future of

Hospital and Primary Care: The Case of Gastrointestinal Diseases. Cardiff: Welsh Health Planning Forum and the Department of Medicine, University of Wales College of Medicine.

Warner P [1994]. Preferences regarding treatments for period problems: relationship to menstrual and demographic factors. *Journal of Psychosomatic Obstetrics and Gynaecology*, **vol. 15**, pp93-110.

Weinstein MC, Fineberg H *et al.* [1980]. *Clinical Decision Analysis*. Philadelphia: WB Saunders Company.

Weinstein MC, Stason WB [1977]. Foundations of cost- effectiveness analysis for health and medical practices. *The New England Journal of Medicine*, pp716-721.

Weintraub WS, Jones EL, King SB, *et al.* [1990]. Changing use of coronary angioplasty and coronary bypass surgery in the treatment of coronary artery disease. *American Journal of Cardiology*, **vol. 65**, pp183-188.

Welsh Health Planning Forum [1994]. *Health and Social Care 2010. Report on phase one*. Cardiff: Welsh Health Planning Forum.

Wennberg JE [1990]. What is outcomes research. Gelijns AC. *Modern Methods of Clinical Investigation*, Washington: Institute of Medicine. pp33-46.

Wennberg JE, Roos N, Sola L, *et al.* [1987]. Use of claims data systems to evaluate health care outcomes. *Journal of the American Medical Association*, **vol. 257**, pp933-936.

Wennberg JE, Mulley AG, Hanley D, *et al.* [1988]. An assessment of prostatectomy for benign urinary tract obstruction. *Journal of the American Medical Association*, **vol. 259**, pp3027-3030.

Wickham JEA [1994]. Minimally invasive surgery: future developments. *British Medical Journal*, **vol. 308**, pp193-197.

Wickham JEA [1987]. The new surgery. *British Medical Journal*, **vol. 295**, pp1581-1582.

Williams A [1985]. Economics of Coronary Artery Bypass Grafting. *British Medical Journal*, **vol. 291**, pp326-329.

Williams A [1995]. The measurement and valuation of health: a chronicle. *Centre for Health Economics Discussion Paper 136*, York: CHE, University of York.

Wilson JM, Dunn EJ, Wright CB, *et al.* [1986]. The cost of simultaneous surgical standby for percutaneous transluminal coronary angioplasty. *Thoracic and Cardiovascular Surgery*, **vol. 91**, pp362-370.

Wolfe BM, Gardiner BN, Leary BF, *et al.* [1991B]. Endoscopic cholecystectomy:an analysis of complications. *Archives of Surgery*, **vol. 126**, pp1192-1198.

Wolfe BM, Gardiner B, Frey CF [1991A]. Laparoscopic cholecystectomy: a remarkable development. *Journal of the American Medical Association*, **vol. 265**, pp1573-1574.

Wong JB, Sonnenberg FA, Salem DN, *et al.* [1990]. Myocardial revascularization for chronic stable angina. Analysis of the role of percutaneous transluminal coronary angioplasty based on data available in 1989. *Annals of Internal Medicine*, **vol. 113**, pp852- 871.

Woods JR, Saywell RM, Nyhuis AW, *et al.* [1992]. The learning curve and the cost of heart transplantation. *Health Services Research*, **vol. 27**, pp219-238.

Zelen M [1990]. Randomised consent designs for clinical trials: an update. *Statistics in Medicine*, **vol. 9**, pp645-656.