A COMPARISON OF WELFARIST AND EXTRA WELFARIST APPROACHES TO VALUING OUTCOMES IN MENORRHAGIA

by

SABINA SANGHERA

A thesis submitted to the
University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Health Economics Unit
School of Health and Population Sciences
College of Medical and Dental Sciences
University of Birmingham
October 2013

ABSTRACT

Extra-welfarist measures are recommended, by decision-makers, for use in economic evaluations. Hence they are commonly used to value outcomes in chronic conditions with episodic symptoms, such as menorrhagia. In menorrhagia, a woman's perceived change in quality-of-life (QoL) is the measure of treatment success and consequently, the primary clinical and economic outcome is change in QoL. This thesis presents findings of a comparison between welfarist and extra-welfarist approaches to valuing outcomes in menorrhagia, and aims to determine the value of levonorgestrel-releasing intrauterine system (LNG-IUS) compared to usual medical treatment for menorrhagia.

Findings from the systematic review demonstrated concerns that extra-welfarist measures may be unsuitable in menorrhagia due to their narrow health-related focus and that results depend on the timing of assessment, given the condition's episodic nature. The economic evaluation alongside the ECLIPSE trial showed that the extra-welfarist measures, EQ-5D and SF-6D, provide contrasting cost-effectiveness decisions. The welfarist willingness-to-pay (WTP) was shown to capture important aspects of wellbeing that are not captured by these extra-welfarist measures. Similar to SF-6D, the economic evaluation using WTP presented evidence against the use of the decision-maker recommended EQ-5D. It is argued that each measure provides information that should be considered by decision-makers when allocating healthcare resources.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisors, Tracy Roberts and Emma Frew, for their advice and support throughout the three years. In the final year particularly, I am extremely grateful to them for always making time to read my chapters despite their busy schedules. To my clinical supervisors, who are also the principal investigators of the ECLIPSE trial, Janesh Gupta and Joe Kai, for being willing to nest my PhD within the trial and for their comments on my research, also to the NIHR HTA for agreeing to fund it.

I would also like to thank the following people for assisting me throughout my PhD journey; Pelham Barton for sharing his technical expertise on the ECLIPSE analysis, Lee Middleton for his statistical advice and for giving me the ECLIPSE trial dataset, Cam Donaldson for his comments on the willingness-to-pay research protocol, Lisa Leighton and Laura Gennard for their administrative support with the willingness-to-pay study, also to Laura for transcribing the interviews, Lisa Gold for so kindly agreeing for me to join her at Deakin University on a placement in Australia, and also for her insightful comments on the research generally and, more specifically, the willingness-to-pay research.

Further thanks to the Birmingham Women's Hospital for allowing me to sit in the clinics to recruit the participants, to all the women who took the time to complete the questionnaire and attend an interview, without whom the research would not have been possible.

Much of this research has been presented at HESG, iHEA, ECHE, seminars within the Health Economics Unit at Birmingham and in Melbourne: I am grateful to all for their comments as they have helped to shape the ideas and have provided thought provoking discussions on the direction of the research.

Thanks to the HAPS PhD students for making the whole experience so enjoyable, especially to Sam, for proof reading some chapters and Paul, for so thoroughly proofreading the thesis.

Finally, I would like to thank my family for always supporting me and taking an interest in my research, especially my mum for her endless encouragement and Tut for proofreading the thesis with such enthusiasm!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CBA; Cost-benefit analysis

CEA; Cost-effectiveness analysis

CEAC; Cost-effectiveness acceptability curve

CUA; Cost-utility analysis

DCE; Discrete choice experiment

ECLIPSE; Effectiveness and Cost-Effectiveness of Levonorgestrel-Containing Intrauterine System in Primary Care against Standard Treatment for Menorrhagia

EQ-5D; EuroQol five dimension

GP; General practitioner

HUI; Health Utility Index

ICER; Incremental cost-effectiveness ratio

LNG-IUS; levonorgestrel-releasing intrauterine system

MAUI; Multi-attribute utility instrument

MCS; Mental composite scale

MMAS; Menorrhagia multi-attribute scale

NB; Net benefits

NHS; National Health Service

NICE; National Institute for Health and Care Excellence

OLS; Ordinary least squares

PCS; Physical composite scale

PSA; Probabilistic sensitivity analysis

QALY; Quality adjusted life year

QoL; Quality-of-life

SF-6D; Short-Form Six Dimension

SF-36/12; Short Form -36/12

SG; Standard gamble

TTO; Time trade off

VAS; Visual analogue scale

WTA; Willingness-to-accept

WTP; Willingness-to-pay

CHAPTER 1. INTRODUCTION

1.1 Economic Evaluation and Outcome Measures

Economic evaluations are carried out to determine the opportunity cost of using resources to provide a service or intervention to one group over another. As healthcare resources are scarce, it is important to assess the relative costs and benefits of interventions to enable decision-makers to allocate such scarce resources. Interventions are not recommended to be implemented in clinical practice without evidence on both the effectiveness and the cost-effectiveness relative to usual practice (NICE, 2008). To conduct an economic evaluation, information on the costs and benefits of competing interventions are considered.

In health economics there are two main alternative frameworks to conducting economic evaluations. These are welfarism and extra-welfarism. Each framework has a different theoretical and methodological basis for carrying out economic evaluations. Briefly, welfarists believe that the output of healthcare should be judged according to the extent to which it contributes to overall welfare, hence the aim of welfarism is to maximise social welfare (wellbeing). Whilst the aim of the extra-welfarist framework, on the other hand, has typically been to maximise health. Extra-welfarists tend to value the output of healthcare in terms of the contribution to health itself, not in terms of preferences for health compared to other goods. Consequently, the approach to conducting economic evaluations differs markedly in terms of the measures of costs and benefits used.

Currently, decision-makers, such as the National Institute for Health and Care Excellence (NICE), recommend the use of the extra-welfarist approaches to economic evaluations, specifically the cost-utility analysis (CUA) (NICE, 2008). In a CUA, the costs are measured in monetary terms and the benefits are measured as a quality adjusted life year (QALY). The QALY incorporates length and quality of life (QoL) in a single metric. The QALY is recommended to be generated using EQ-5D, which is a health-related QoL measure. However, when EQ-5D is not considered to be suitable for a condition, decision-makers will accept the QALY outcome derived from another health-related QoL measure, SF-6D (NICE, 2008).

When costs are measured in monetary terms and outcomes are expressed in natural units, which are specific to the condition under analysis (i.e. symptom avoided), the economic evaluation is termed a cost-effectiveness analysis (CEA). Whilst a CEA is an extra-welfarist evaluation, it is not recommended by decision-makers because the findings from these economic evaluations cannot be compared across conditions, due to the disease-specific nature of the outcome measure used. It should be noted that beyond the UK, the term CEA and CUA are typically used interchangeably, to mean a CUA.

An alternative economic evaluation, which is less commonly used in health, is known as a cost-benefit analysis (CBA). The CBA is a welfarist approach to economic evaluation. In this economic evaluation, both the benefits and costs are measured in monetary terms. The outcome measure typically used is willingness-to-pay (WTP). WTP provides a broad assessment of wellbeing and therefore encompasses benefits beyond health. The CBA is widely used in all other sectors, including transport, environment and education (Gafni, 2006). However, its application in healthcare is much more limited due to the perception of decision-makers that results are based on ability to pay, amongst other methodological issues (NICE, 2004).

Finally, when the outcomes and costs are presented in a disaggregated manner the economic evaluation is called a cost-consequence analysis. This analysis can be used when data from multiple outcome measures are available and when it is thought that each measure provides additional information that the decision-maker should take into consideration when allocating resources. For example, both extra-welfarist and welfarist costs and outcomes could be presented to decision-makers in a cost-consequence analysis.

For economic evaluation the value that individuals place on healthcare is typically assessed using measures that assess preferences for possible outcomes (preference based measures), rather than measures that assess health status (non-preference based measures) (Brazier et al, 1999). The welfarist approaches to valuing outcomes, such as WTP, are based on preferences, and this basis for valuing outcomes has been used in the case of the extra-welfarist CUA measures, EQ-5D and SF-6D. This issue is discussed further in section 3.3.

1.2 Valuing Outcomes in Menorrhagia

Menorrhagia is defined as “excessive menstrual blood loss which interferes with the woman’s social, emotional, physical and material quality of life” (NICE, 2007). It is a chronic condition but has symptoms that occur in episodes, which typically present for one week out of every month until menopause. Menorrhagia places a considerable burden on healthcare resources, with around 6% of women per year consulting their general practitioner (GP) in England and Wales (Office of Population Census and Surveys, 1995). Treatment is prompted predominantly by a woman’s subjective assessment of interference in her QoL, rather than solely by clinical assessment of volume of blood loss (Shapley et al., 2002). Women may change or cease treatment, according to their perception of effectiveness and relative to their contraceptive needs. The primary clinical and economic outcome measure used to assess treatment success is improvement in QoL.

Historically, women often progressed quickly to a surgical solution; either hysterectomy, resulting in the permanent cessation of bleeding and sterility, or since the 1990s, endometrial ablation, which uses electrical or thermal energy to destroy the endometrium, causing cessation of bleeding in 34% of women (Daniels et al., 2012). Non-hormonal and hormonal medical treatments are now available as first line therapy for women presenting with menorrhagia in primary care.

In 2007, NICE introduced guidelines for the levonorgestrel-releasing Intrauterine System (LNG-IUS) to be used to treat menorrhagia (NICE, 2007). It is shown in Chapter 4 that there is currently no reliable evidence on the cost-effectiveness of LNG-IUS compared to usual medical treatment and that the NICE recommendations for its use in clinical practice are based on limited evidence on cost-effectiveness (Stewart et al., 2001). The importance of using an outcome measure that accurately reflects women's concerns and experiences is explored throughout this thesis.

1.3 Aim and Objectives

The aim of this research is to (1) determine the value of LNG-IUS compared to usual medical treatment in menorrhagia, and (2) consider the suitability of current measures available for assessing the value of outcomes and interventions in menorrhagia.

The research has the following objectives;

- Systematically review the literature to; identify the current evidence from economic evaluations that have assessed LNG-IUS or usual medical treatment in menorrhagia, and to determine which economic outcome measures have been used and assessed in menorrhagia.

- Conduct an economic evaluation for the ECLIPSE trial using the currently recommended EQ-5D measure, and re-estimate this evaluation using SF-6D to draw a comparison between the cost-effectiveness findings of these extra-welfarist measures.

- Elicit WTP for LNG-IUS and usual medical treatment, from the theoretically preferred ex-ante perspective and the most commonly practiced ex-post perspective, followed by interviews to assess respondents’ understanding of the WTP exercise.

- Conduct a CBA comparing LNG-IUS and usual medical treatment, and compare the cost-effectiveness decisions between the welfarist CBA and the extra-welfarist CUAs.

1.4 Structure of Thesis:

Following on from this introductory chapter, the remaining 9 chapters are structured accordingly.

In **Chapter 2** the background of the clinical condition, menorrhagia, is outlined. This chapter describes the condition and its treatments, highlights the importance of assessing outcome measures particularly in menorrhagia, and describes the ECLIPSE trial, within which this PhD is nested.

Chapter 3 introduces the concepts of welfarism and extra-welfarism. The chapter outlines the theoretical underpinnings behind the two frameworks for conducting economic evaluations. The theoretical foundations, the methodological considerations associated with each of these frameworks and their use in economic evaluations are described.

In **Chapter 4** a systematic review of economic evaluations using LNG-IUS or usual medical treatment is described. This review was carried out to identify the available evidence within the area. That is, which intervention has been shown to be cost-effective in previous studies and to identify which outcome measures and decision models were used to determine these cost-effectiveness decisions.

Chapter 5 follows on from Chapter 4 by systematically reviewing the evidence on the outcome measures used in menorrhagia. The type of economic outcome measure used is identified, and the current evidence on the psychometric properties and the feasibility of the use of these measures in menorrhagia is presented.

The use of the extra-welfarist measures EQ-5D and SF-6D are explored in **Chapter 6.** In this chapter the model-based economic evaluation for the ECLIPSE trial is reported using EQ-5D. The economic evaluation is then re-estimated using SF-6D and the findings from the two extra-welfarist measures are compared and discussed.

The use of the welfarist WTP measure in menorrhagia is investigated in **Chapters 7, 8 and 9**;

In **Chapter 7** the methodology for the empirical welfarist WTP study is reported. The method for eliciting WTP from the ex-ante and the ex-post perspective is described, in addition to the methods used for the interviews. The chapter is concluded with a discussion related to the methodology used.

In **Chapter 8** the results of the WTP study are presented. The chapter is separated into three parts. In Part 1, the results from the ex-ante perspective are reported. In Part 2, the results from the ex-post perspective are outlined, and Part 3, refers to the findings from the interviews. Each part is concluded with a discussion and an overall discussion on the WTP findings is reported at the end of the chapter.

The findings from the ex-ante WTP study are then incorporated into a CBA which is reported in **Chapter 9.** The economic evaluation is carried out similar to the extra-welfarist economic evaluation to enable comparisons to be drawn.

In **Chapter 10** a comprehensive discussion on the key findings of the thesis that relate to the initial aim is reported. The main findings in relation to other research are presented along with reflections on the use of the welfarist and extra-welfarist measures in menorrhagia. The chapter is concluded with implications for policy-makers and further research recommendations.

CHAPTER 2. BACKGROUND OF CONDITION

2.1 Introduction

In this chapter the background to the condition of menorrhagia is reported. Since the objective of this thesis is to assess the suitability of outcome measures used in menorrhagia, the reasons for the focus on this condition are outlined. The chapter is structured as follows; first the nature of the condition is discussed, followed by the prevalence, types of treatments available and the clinical findings from the ECLIPSE trial.

2.2 Nature of the Condition

Menorrhagia is defined as "excessive menstrual blood loss which interferes with the woman's social, emotional, physical and material quality of life" (NICE, 2007). Menstrual bleeding typically begins at 12 years of age and ceases when menopause occurs, around 45-55 years old (NHS Choices, 2012). Menorrhagia occurs for a 'period' of the month, each month. The duration of menstruation typically lasts up to 7 days, but varies across women. Therefore, menorrhagia can be considered as a chronic condition that occurs in episodes.

Similar to the duration of bleeding, the volume of blood loss also varies across women but the 'average' volume of blood loss for a woman at reproductive age is 35ml per menstrual cycle however, the volume of blood loss is known to increase with age (O'Flynn & Britten, 2000). Previous attempts have been made to clinically define menorrhagia according to an objective volume of blood loss. The exact volume that the definition should be set to has been disputed over the years by researchers, but a definition of blood loss in excess of 80ml per cycle remained for many years (Fraser et al., 1984; Hallberg et al., 1966; Janssen, 1998; O'Flynn & Britten, 2000).

In theory, defining menorrhagia according to an objective volume of blood loss for clinical use to diagnose menorrhagia seems reasonable. However in practice, objective measurements of volume of blood loss are rarely taken and women are treated according to their subjective complaint (Sambrook & Cooper, 2005). Treatment was shown to be prompted predominantly by a woman's subjective assessment of interference in her quality of life (QoL), rather than solely by traditional clinical assessments of volume of blood loss (Shapley et al., 2002). Thus, women who complain of menorrhagia, but do not bleed in excess of 80ml per cycle, are still offered the same medical treatment as women that do bleed in excess of 80 ml per cycle.

Moreover, women's complaints of blood loss have been determined to be secondary to those of pain, physical sensation of blood loss, and tiredness (O'Flynn & Britten, 2000). Menorrhagia was also shown to impact on physical and psychological health, limiting work productivity, hindering social life and affecting family life and relationships (Bunkheila & Powell, 2008; Shaw et al., 1998). Further, Shaw et al (1998) also showed that impact on family life/ relationships was more important than physical wellbeing (Grant et al., 2000; Shaw et al., 1998). Following these findings and the difficulties in objectively identifying menorrhagia, in 2007 the National Institute of Health and Care Excellence (NICE) in the UK formally recognised that a clinical definition of menorrhagia based on volume of menstrual blood loss was inappropriate. The medical definition was changed to ensure that a woman's subjective assessment of her ability to cope with the blood loss and the impact of bleeding on her QoL are the key prompts for medical intervention. Therefore, the primary indicator of treatment success is now improvement in QoL (NICE, 2007).

2.3 Prevalence

Menorrhagia is one of the most common gynaecological conditions presenting in primary care (Protheroe, 2004). The condition places a considerable strain on healthcare resources as it affects 1.5 million women in England and Wales annually, resulting in 1 in 20 women, aged between 30-49, per year, consulting their GP with this problem (Effective Healthcare Bulletin, 1995; Office of Population Census and Surveys, 1995; Rees, 1991); with slightly fewer women of a younger age consulting their GP. Between 10-12% of all gynaecological referrals had been shown to be comprised of menorrhagia (Coulter et al., 1995; Grant et al., 2000). Approximately 50% of women referred to secondary care had a hysterectomy (Coulter et al., 1991; Grant et al., 2000), with 1 in 5 women expected to have a hysterectomy prior to turning 60 years old as a result of the condition (Bulmer, 2008). However, with the introduction of guidelines to promote pharmaceutical treatments as first line treatment, the number of hysterectomies carried out is reducing (Bulmer, 2008; RCOG, 1998). The relative impact on resources may be particularly great in menorrhagia, because an objective assessment and clinical criteria is not required to be met in order to obtain treatment, as women can demand treatment based on their perception of interference on their lives and not necessarily a clinical assessment.

2.4 Treatment

Menorrhagia can occur as a direct result of another condition, such as an infection, fibroids, or a coagulation (blood clotting) disorder (Woman's Health Medicine, 2005). In these cases, because menorrhagia is the secondary complaint, treatment is generally directed towards the underlying condition, which in turn eliminates menorrhagia. Menorrhagia can also occur when an underlying pathology does not exist. In these cases, as the cause of heavy bleeding is unknown, non-surgical treatments are used to help women manage the bleeding, rather than provide a cure for the bleeding. The second type of menorrhagia, with no underlying cause, is the main focus of this research as menorrhagia is the primary complaint from the women.

Historically, women usually progressed quickly to one of two surgical solutions to treat menorrhagia (Daniels et al., 2012). The first is hysterectomy, which is an expensive major operation resulting in the permanent cessation of bleeding and sterility. The second option, available since the 1990s, is endometrial ablation, which is a minor operation that uses electrical or thermal energy to destroy the endometrium, causing cessation of bleeding in 34% of women (Daniels et al., 2012). If not fully effective the first time, endometrial ablation may need to be repeated.

In 2007, the UK NICE guidelines (NCCWCH, 2007) for menorrhagia were published and recommended that, due to their invasive nature and expense, surgical intervention be the *last* course of treatment for menorrhagia. Surgical interventions are now only considered when pharmaceutical interventions are either unsuccessful, poorly tolerated, or the patient has a strong preference for surgery (Protheroe, 2004). As an alternative to surgical intervention, patients can manage their bleeding by taking pharmaceutical treatments. These include a range of alternative medications which can be taken orally, via injection, or inserted in the uterus, as is the case for the newly licenced levonorgestrel-releasing intrauterine system (LNG-IUS). As changes to the provision of treatment has been made, the clinical and cost-effectiveness of pharmaceutical treatments has become the primary focus of clinicians and decision-makers alike.

In this thesis a comparison of the alternative economic theoretical frameworks for valuing outcomes in menorrhagia will, in part, be assessed by measuring changes in outcomes associated with usual medical treatment or LNG-IUS. This is because, first, the research reported within this thesis is nested within the ECLIPSE trial and the comparator treatments of the trial include 'usual medical treatment' and 'LNG-IUS' and, second, because evidence on the clinical and cost-effectiveness of these competing treatment strategies is required to determine the first line treatment strategy in menorrhagia. These treatment strategies will be explained next.

2.4.1 Usual Medical Treatment

In this context, and according to the ECLIPSE trial definition, usual medical treatment includes either one of, or a combination of, five alternative treatments (NICE, 2007; RCOG, 1998). The alternative treatments are either non-hormonal which do not provide contraception; or hormonal treatments which do provide contraception.

i) Non-Hormonal Treatment

Non-hormonal treatments include tranexamic acid and mefenamic acid. In some cases a combination of these treatments are offered to patients to help alleviate symptoms. These non-hormonal treatments can also be combined with hormonal treatments. Non-hormonal treatments are not provided free of charge by the UK National Health Service (NHS), but the prescriptions for the medication are heavily subsidised by the government. These treatments are provided at the primary care level by a GP and are reviewed for effectiveness at 3 months.

Both mefenamic acid and tranexamic acid are oral tablets that are not taken continuously. Thus they are only taken during menstruation. Whilst both treatments reduce menstrual blood loss, mefenamic acid is claimed to be the least effective of the two treatments as blood loss is reduced by approximately 30% in comparison to the reduction of 50% observed when tranexamic acid is used (Zachariah & Fender, 2005). However, in addition to reducing blood loss, mefenamic acid also alleviates pain associated with menstrual cramps, which tranexamic acid does not (NICE, 2007; Protheroe, 2004). Potential side effects of both of these non-hormonal treatments include gastrointestinal side effects which are minor and uncommon.

ii) Hormonal Treatments

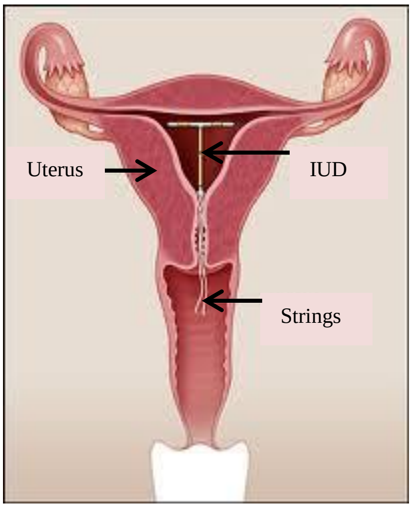
Hormonal treatments include combined oral contraceptives, injectable progestogen (depo-provera), and high dose progestogen (norethisterone). As each of the hormonal treatments provide contraceptive benefits they are provided free of charge to patients. Each of these treatments provide reversible contraceptive protection in addition to a reduction in menstrual blood loss. Similar to non-hormonal treatments, these hormonal treatments are prescribed at the primary care level by a GP. Combined oral contraceptives are reviewed at 3 months, whilst norethisterone and depo-provera are reviewed at 6 months.

A large range of these combined oral contraceptives are offered to patients with menorrhagia (Kadir, 2009). They are able to reduce blood loss by 50% and help to regulate the cycle (Zacchariah & Fender, 2005). Unlike the non-hormonal treatments, combined oral contraceptive tablets must be taken daily at approximately the same period of time, and in some cases, a seven day break from medication is required. Norethisterone should be taken orally from day 5 to day 25 of the menstrual cycle. Norethisterone reduces the volume of bleeding by up to 85% (NICE, 2007). However, the evidence on the effectiveness of norethisterone is poor and widely disputed (Protheroe, 2004). Due to the high dose required to reduce bleeding, norethisterone is commonly associated with side effects, which has led to reduced compliance (Kadir, 2009). Depo-provera is administered to the patient by the GP via injection every 6 months. This injected medication has been shown to cause complete cessation of blood loss in up to 85% of women (Bulmer, 2008). However, similar to norethisterone, the side effects of depo-provera, which can include irregular bleeding, lead to low compliance as patients discontinue treatment. Common side effects for all hormonal treatments include headaches, acne, nausea and weight gain, which typically subside over time.

2.4.2 Levonorgestrel-releasing Intrauterine System (LNG-IUS)

LNG-IUS is a hormone releasing intrauterine device which is inserted into the uterus by a healthcare provider, normally a GP (Figure 2.1). The LNG-IUS was originally licensed for contraception only, but its notable reduction in blood loss was identified and its potential use in menorrhagia was realised. The LNG-IUS is free of charge to patients, the insertion of the LNG-IUS typically takes a few minutes and the device remains active for up to 5 years. After 5 years the device can be replaced.

Figure 2.1 Image of intrauterine device (IUD) in uterus (Gralapp, 2006)



LNG-IUS is claimed to be the most effective alternative treatment to hysterectomy, as it minimises blood loss and has the ability to eliminate menstrual pain (Kadir, 2009). The reduction in blood loss ranges from 71% to 96% (NICE, 2007). However, the LNG-IUS has an extensive side effect profile which is extremely common within the first 6 months of fitting. These side effects are thought to be a large reason for patients seeking alternative treatment. Side effects include increased bleeding, increased irregular bleeding, bleeding between menstruation, bleeding following intercourse and device expulsion. If women are able to withstand these side effects, after 6-9 months 90% of women have been shown to be pleased with the treatment (Bulmer, 2008).

In all of the above cases, when administration of these non-surgical treatments is stopped, the beneficial effects are also no longer seen. Hence, unlike surgical treatments, these pharmaceutical treatments are not curative. For this reason, women often seek to change, or cease treatment, according to their perception of effectiveness, and relative to their contraceptive needs or their reproductive life stage.

2.5 The ECLIPSE Trial

To identify evidence on the clinical and cost-effectiveness of the two treatments LNG-IUS and usual treatment, the largest randomised controlled trial in menorrhagia and the first in the UK to compare these treatments is currently underway. The acronym ECLIPSE stands for The Effectiveness and Cost-Effectiveness of Levonorgestrel-Containing Intrauterine System in Primary Care against Standard Treatment for Menorrhagia (ECLIPSE). The (ECLIPSE) trial is a National Institute of Health Research Health Technology Assessment (NIHR HTA) funded randomised controlled trial.

Clinical trials are used to gather data on the effectiveness of new interventions. There are several types of clinical studies, but randomised controlled trials are currently considered the gold standard (Sibbald & Roland, 1998). Randomised controlled trials typically provide unbiased evidence on the treatment-effect relationship because they are carried out in controlled conditions which ensure that there are no systematic differences between the compared treatment groups (Sibbald & Roland, 1998). Randomised controlled trials are known to provide the most rigorous method of establishing whether a relationship exists between the treatment and outcome, which in turn means that the findings of economic evaluations, alongside these trials, are amongst the most robust (Sibbald & Roland, 1998). Hence economic evaluations are commonly conducted alongside randomised controlled trials to identify the cost-effectiveness of a 'new' intervention, relative to treatment as usual. Evidence on cost-effectiveness alongside evidence on effectiveness is required by decision-makers when making decisions about resource allocation.

2.5.1 Study Design

Full details of the trial are reported in Gupta et al (2013). Briefly, 571 women with menorrhagia from 63 UK centres were randomised between February 2005 and July 2009. Women between 25 and 50 years of age presenting to their GP with menorrhagia, occurring over at least three consecutive cycles, provided written informed consent to participate. Women were randomised to having a LNG-IUS fitted, or usual medical treatment, chosen by the GP and the woman based on contraceptive needs or desire to avoid hormonal treatment.

Usual medical treatment options included mefenamic acid, tranexamic acid, norethisterone, a combined oestrogen/progestogen or a progestogen only oral contraceptive pill (any formulation), or methoxyprogesterone acetate injection (NCCWCH, 2007; RCOG, 1998). The particular medical treatment was specified prior to randomisation. Treatments were reviewed by a GP at 6 weeks and 3 months. Subsequently, treatments could be changed or discontinued due to perceived lack of benefit, side effects, change in contraception need, referral for endometrial ablation or hysterectomy as per usual practice (NCCWCH, 2007; RCOG, 1998). Treatment changes reported by patients were confirmed with the GP.

The primary outcome measure used in the clinical analysis was a disease-specific QoL measure named Menorrhagia Multi-Attribute Scale (MMAS). Briefly, the MMAS questionnaire is comprised of six attributes including ‘practical difficulties’, ‘social life’, ‘psychological health’, ‘physical health’, ‘work and daily routine’, and ‘family life/relationships’ (Shaw et al., 1998). Each attribute has been weighted according to patient preferences using counters, which are considered to be ‘importance points’. The patients disease-specific QoL is captured on a 0-100 scale (Shaw et al., 1998). Zero represents the worst severity of the condition and 100 represents the best state for the condition. The methods for scoring and weighting measures are discussed further in section 3.3.2 in Chapter

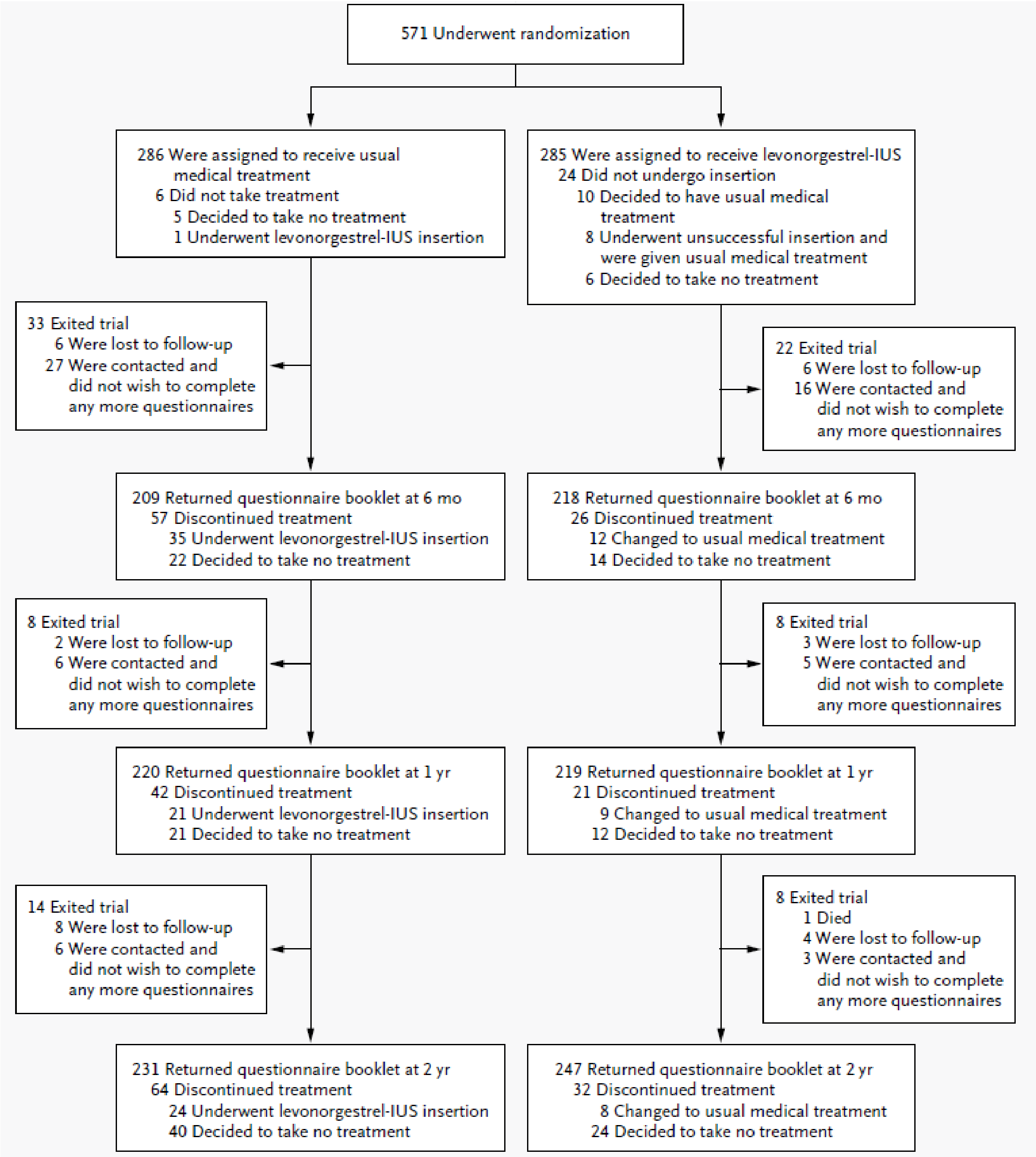
3. Secondary outcomes such as EQ-5D and SF-36 (described in more detail in the next chapter), impact on sexual activity and number of surgical interventions were also recorded. Data were collected at baseline, 6 months, 1 year and 2 years.

Currently the 2 year interim analysis on the clinical effectiveness has been carried out and published (Gupta et al., 2013). A 5 year analysis and potential 10 year analysis are planned for the future. In this section, the clinical findings from the trial so far will be summarised.

2.5.2 Trial Results

The trial had an (478/571) 84% response rate at the 2 year time point but in 9% of these cases, the MMAS was not completed (Gupta et al., 2013). The number of women in each treatment arm, at each time point, is detailed in Figure 2.2.

Figure 2.2 Follow up of ECLIPSE patients (reproduced from Gupta et al 2013)



According to Gupta et al (2013) at the 2 year time point, the chance of women remaining in the LNG-IUS arm was approximately twice as likely as those randomised to usual medical treatment remaining in that arm. A significant improvement in the primary outcome, MMAS, was observed in both treatment arms at all time points compared to baseline (Gupta et al., 2013). However, when compared to usual medical treatment, the degree of improvement was significantly greater in the LNG-IUS arm at every time point of assessment. The average difference was 13.4 points [95% CI 9.9 -16.9] $p<0.001$, which suggests an improvement in two or three of the MMAS domains indicating a change from being greatly affected to minimally affected or minimally affected to unaffected (Gupta et al, 2013). 22% of women underwent surgery, either hysterectomy or endometrial ablation, at the 2 year time point and there was no statistically significant difference in the rate of surgery between the two treatment arms (Gupta et al., 2013). In the trial based analysis the secondary outcomes for general QoL revealed the following: EQ-5D scores increased significantly from baseline in both treatment groups; a significant difference in scores between treatments was not detected ($p=0.38$); the scores from SF-36 were reported according to the instrument attributes and similar to EQ-5D showed a significant improvement from baseline; for SF-36 in all but one domain, 'mental health', the score for LNG-IUS was significantly greater than the score for usual medical treatment.

Consequently, the clinical results of the ECLIPSE trial showed that LNG-IUS produces a significantly greater improvement in disease-specific QoL compared to usual medical treatment. The results of the cost-utility analysis alongside the ECLIPSE trial are reported in Chapter 5.

2.6 Conclusion

Menorrhagia is a common condition that has a considerable impact on healthcare resources. Objective assessments of volume of blood loss are not deemed to be suitable for the condition. The medical definition of menorrhagia has been changed to ensure that a woman's perception of blood loss and the impact of bleeding on her QoL are the primary prompts for medical intervention. Therefore, the patients' subjective assessment of their ability to cope with the bleeding is important, rather than an objective assessment of volume of blood loss. As surgical interventions are no longer considered as first line treatment for menorrhagia, the clinical and cost-effectiveness of non-surgical treatment strategies, LNG-IUS and usual medical treatment, must be explored. Given that LNG-IUS is a relatively newly licensed treatment for menorrhagia, and very little evidence is currently available, the ECLIPSE trial aimed to explore this very issue and found that at the 2 year time point LNG-IUS is most effective when a clinical outcome measure is used.

As the primary indicator of treatment success is improvement in QoL it is deemed important to identify which QoL measures have been used, and to determine whether current measures accurately capture women's concerns and experiences. This assessment of outcome measures will ensure that robust recommendations of economic evidence are provided to decision-makers. As part of the empirical work, the economic evaluation for the ECLIPSE trial comparing LNG-IUS and usual medical treatment is presented in Chapter 6. The use of economic evaluations, and outcome measures, from the alternative frameworks within menorrhagia will then be explored in subsequent chapters.

In the next chapter, a background to economic evaluation is provided. The theoretical foundations behind economic evaluations and the outcome measures, described.

CHAPTER 3. THEORETICAL FOUNDATIONS AND APPLICATION OF ECONOMIC EVALUATION

3.1 Introduction

In this chapter a discussion on some of the theoretical and methodological principles of economic evaluation is provided. As mentioned in Chapter 1, healthcare resources are scarce and economic evaluations are used by decision-making bodies such as the UK National Institute for Health and Care Excellence (NICE) to aid decisions about resource allocation. Economic evaluations represent a framework used to compare the costs and benefits, or outcomes, of alternative options for resource use (Drummond et al., 2005). In a normal market the value of a good is determined directly by observing how an individual responds to changes in price and quantity (McIntosh et al, 2010). For example, whether the individual is willing to pay £40 but not £50 reveals how much they value that good. Therefore, whether the individual consumes the good is related to their willingness and ability to pay (Morris et al, 2007). However in healthcare, the conditions that define a normal market are not met and market failure is said to occur. Consequently, in the UK, the government intervenes to ensure an efficient allocation of resources and provides healthcare free at the point of consumption (Morris et al, 2007). Hence, it is not possible to measure individual's preferences or value for healthcare by assessing their response to changes in price. It is then necessary to use other measures to assess the benefits of a treatment or programme, which can be measured directly or indirectly and are discussed in section 3.3. Economic evaluation provides values for healthcare services (McIntosh et al, 2010). The type of economic evaluation used is driven by the theoretical underpinnings of the approach. The most suitable method to use is widely debated.

The aim of the thesis is to assess the valuations of the outcomes used in economic evaluations of interventions for menorrhagia. In this chapter, a comprehensive discussion of the two main theoretical frameworks for conducting economic evaluations is first presented. The methodological considerations associated with the application of these frameworks to economic evaluations are then reported. The reporting is deliberately concise as a wealth of literature has already been written within the area (Birch & Donaldson, 2003; Brouwer et al., 2008; Drummond et al., 2005; Morris et al., 2007).

3.2 Theoretical Foundations

Normative economics is a type of analysis to assess the most desirable resource allocation according to a set of assumptions or value judgements. The assumptions provide a basis for ordering different states of resource allocation according to desirability, enabling an assessment of the value of providing the resources to one group and not another (Boadway & Bruce, 1984). The investigation into determining (ranking) which states are most desirable is traditionally known as welfare economics. Further, some researchers do now differentiate normative economics into welfarism and non-welfarism to allow for the additional methods developed for ranking alternative states (Morris et al., 2007; Tsuchiya & Williams, 2003).

Broadly, within health economics, there are two frameworks from which economic evaluations can be conducted. These include ‘welfarism’ and ‘extra-welfarism’. Extra-welfarism is considered to be a type of non-welfarism and is the most widely accepted alternative to welfarism. Other such non-welfarism frameworks, or the decision-maker approach, have also been suggested but all of these approaches have the same objective as extra-welfarism (discussed in 3.2.2). The welfarist and extra-welfarist framework differ in the methodological basis for conducting economic evaluations, as the value judgements used for ranking different states of resource allocation differ. In the next sections each framework will be discussed in turn.

3.2.1 Welfarism

Welfare generally refers to well-being and the key value judgment behind welfare economics is to maximise social welfare (wellbeing), by maximising utility (McIntosh et al., 2010). The definition of 'utility' or 'utilities' varies widely, but in welfarism it is typically considered a measure of strength of preference which is related to desirability (Coast, 2009). The impact of a change in resource allocation is judged by the value that individuals place on the alternative states of the world and this value for the change is captured in terms of changes in *utilities*. In other words, value judgements about the desirability of different options of resource allocation are judged by how the changes affect individual utility. This approach, where the evaluative space is confined to impact on utilities, was first termed 'the welfarist approach' by Sen in 1977 (Sen, 1977).

Welfare economics is grounded in consumer choice (preference) theory, which assumes that individuals will have preferences and will therefore be able to rank alternative states of the world with the aim of maximising their own utility, within the restraints of their income (McIntosh et al., 2010). Thus the welfarist decision rule of maximising utility is rooted in theory. The utility change that occurs as a result of alternative states of resource allocation (the welfare change) is dependent on the individuals' preference for the alternative states. The strength of preference for the welfare change determines the amount the individual values the change (i.e. how it changes their utility). Following a change in welfare (resource allocation) individuals' utilities are aggregated to determine whether that change in resource allocation results in an improvement in social welfare (McIntosh et al., 2010). As utility cannot be measured directly, money is used as a proxy for the change in utility.

The key value judgement, of maximising utility (social welfare), was presented as the Pareto principle. The principle asserts that a state of the world is better than another if an individual is made better off without making another individual worse off (Mitchell & Carson, 1989). This state of the world is considered to be an improvement in social welfare, hence there is a Pareto improvement; ergo a policy that can achieve this should be implemented. The Pareto principle provides a method for drawing comparisons between different states of the world according to whether a Pareto improvement is observed. However, it has been widely argued that in reality it is unlikely that a change in resource allocation could make one person better off without making another worse-off (Coast, 2004; Mitchell & Carson, 1989). Due to these limitations and the lack of translation to real life, a compensation principle was introduced to operationalise the approach. The compensation principle by Kaldor and Hicks states that there is a potential Pareto improvement if the gainers, from the change in resource allocation, could hypothetically compensate the losers (Hicks, 1939).

Hicks then described two key methods to use to measure the impact of a change in welfare, which are (i) equivalent variation and (ii) compensating variation (Hicks, 1939). The aim of these methods is to identify the level of compensation that ensures utility levels are unchanged. A third method known as consumer surplus which is the difference between what an individual is willing to pay and what they actually pay (Jonnesson & Jonsson, 1991), is less widely used as it is thought to be less appropriate in a policy decision-making context (McIntosh et al., 2010).

Equivalent variation measures the amount of money required to keep the individual at the utility level post change (O'Brien & Gafni, 1996). Hence the change has already occurred and the measurement is from an ex-post perspective. The compensating variation measures the amount of money required to keep the individual at the initial (before change) utility level. Thus it is measured prior to the change occurring, from an ex-ante perspective, to obtain the expected change in utility following the provision of resources (O'Brien & Gafni, 1996).

The Kaldor criterion of the Kaldor-Hicks compensation principle states that a programme results in an increase in efficiency and should be implemented if the gainers can compensate the losers sufficiently so that everyone would be better off following the change (Kaldor, 1939). Hence the criteria of the compensation principle (whether a programme would be implemented) is equivalent to whether the sum of the values derived from the compensating variation is greater than zero, as explained by McIntosh et al (2010). The compensating variation relates to the gainers being able to compensate the losers and still maintain the initial utility level. If this is the case, the programme should be implemented. The (compensating variation and Kaldor criterion) values are relative to the initial utility level and therefore the compensating variation (ex-ante) is the theoretically preferred approach (McIntosh et al., 2010). The equivalent variation is related to the amount of money that the losers must give to the gainers to compensate them for the forgone utility gain from the lack of intervention. In this case everyone would be better off without the change and therefore the programme should not be implemented. The compensation principle is operationalised in economic evaluations through the use of the cost-benefit analysis (CBA) (Mishan, 1972). The Kaldor-(Hicks) criterion is implemented in a CBA using methods associated with the compensating variation. Although the compensation principle is widely accepted, it is argued that as the compensation is hypothetical, the change would involve a redistribution of health in favour of the gainers at the expense of the losers (Little, 1949; Morris et al., 2007).

In welfarist theory, an improvement in social welfare is measured as an improvement in utility alone. The effect that the consumption of healthcare has on individual utility is valued in the utility function but the provision of healthcare itself is not (Brouwer & Koopmanschap, 2000). Health itself does not enter the utility function but the consequence of health, what the consumption of healthcare and subsequent improved health enables an individual to do, is reflected in the utility function, in the form of the ability to, for example, go back to work, to continue with daily routine, feeling better and so on. Thus welfarism considers overall wellbeing and enables the individual to incorporate both health and non-health benefits into their decision in so far as the effect on their utility (Donaldson et al., 1997a). Culyer (1991) argued that a downfall of the welfarist approach is that benefits associated with non-goods characteristics, such as the process of care, cannot be incorporated into the analysis but Birch and Donaldson (2003) explain that characteristics of non-goods can and have been incorporated into welfarist studies, in terms of process utility which is the utility associated with the process of care, in so far as its consequence on utility.

Some researchers have intimated that welfarism requires value judgments of alternative states of the world to be based on the individuals who are affected by the change in resource allocation only, as individuals are thought to be the best judge of their own welfare (Brouwer et al., 2008; Coast, 2004). This notion has been criticised for not accounting for individuals who are willing to sacrifice their own utility, in order to improve the utility of another. However, Birch and Donaldson (2003) outline that this is not a strict criterion of welfarism. It should be noted that caring externalities and option externalities have been incorporated into studies, so the values of individuals who are not directly affected by the change in resource allocation, can be taken into consideration (Birch & Donaldson, 2003). Further, the way in which healthcare is funded determines who the affected individuals are, following a resource change. From a UK perspective, as healthcare provision is primarily publicly financed through taxation, all citizens contribute to the National Health Service. It is therefore argued that all citizens would be affected as they either incur the losses or benefits associated with a change in resource allocation (NICE, 2008). Brazier et al (2005) explains that the welfarist approach is unclear in this respect and that to use a welfarist framework does not necessarily mean to use patient values in a publicly financed system, as they are not the only affected individuals. O'Brien and Gafni (1996) state that in a publicly financed system the entire population makes up the affected individuals. Shackley and Donaldson (2000) outline whose values should be used according to each context. This notion of whose values should be incorporated is discussed further in section 3.3.1.

The welfarist approach addresses issues with allocative efficiency (Currie et al., 1999). That is, as the outcome measure is expressed in monetary terms, comparisons of resource use can be made across industry sectors and comparisons are not confined to healthcare. Thus, information can be used to determine how resources should be allocated to the healthcare, transport, environment and education sectors. However, as issues of distribution and efficiency are separated in the welfarist approach, it is argued that welfarism rarely takes into account equity and distribution concerns (Morris et al., 2007; Olsen & Smith, 2001). The Kaldor-Hicks compensation principle presents the most efficient allocation of resources and the redistribution of utilities, or compensation, can be researched by others and applied by decision-makers (Coast, 2009; Kaldor, 1939). The issue of 'ethical' or equitable distribution can be solved by decision-makers. Work has been undertaken to apply distributional weights, but these are rarely used in practice (Donaldson, 1999).

Despite the extensive use of the welfarist framework in transport and environmental economics and its strong theoretical underpinnings in consumer choice theory, the use of the welfarist approach in health economics is widely debated (Brouwer et al., 2008). It is often argued to be too narrow due to its focus on utility alone. Sen has argued that utility alone is not sufficient to capture wellbeing, and that healthcare and health should be valued as an end in itself, not for the benefits that health can derive (Sen, 1980). The quality of the utility derived and the individual's capability to convert the healthcare received into wellbeing were argued to be necessary to take into consideration. Further, an individual may have adapted to their current situation and therefore adjusted their expectations of utility, which would subsequently alter their ability to value and desire goods (Brouwer et al., 2008; Cohen, 1993).

It is also often argued that as judgements on different states of allocation are based on an individual's strength of preference for a state, which is measured in monetary terms, it is ultimately based on ability to pay, therefore the allocation of resources are argued to be skewed towards the rich (Coast, 2004; Gold et al., 1996). This is one of the key arguments against the use of welfarism in healthcare. It is argued that healthcare should be distributed in a fair manner, i.e. based on need, not on strength of preference (Brouwer et al., 2008). Healthcare is considered to be a merit good that should be provided to society based on need (Musgrave, 1959). Despite the distributional weights available, this argument of not basing allocation on anything other than need is backed by decision-makers. Hence, the claimed discomfort of the public and decision-makers of valuing health in monetary terms is argued to be one of the main reasons that have led to the reduced focus of welfarism in healthcare and the development of other methods for valuing different states of resource allocation (Brouwer et al., 2008, Coast et al., 2008b).

Brouwer et al (2008) described a final ‘seed’ that led researchers to consider other approaches for valuing different states of resource allocation, that of allowing decision-makers to be the sources of values in public decision-making. One such approach, which was developed out of the perceived problems associated with welfarism, was extra-welfarism.

3.2.2 Extra-welfarism

The exact definition of extra-welfarism is unclear. Other prior approaches such as the ‘non-welfarist’, or the ‘decision-maker approach’ all differ in their exact application but the key value judgement behind all three of these approaches, as applied in health economics, is the same.

In 1991, Culyer (1991) wrote about an approach, now known as extra-welfarism, which attempted to overcome the perceived problems associated with the welfarist approach and incorporate Sen’s writings. The approach aimed to consider non-utility information, such as capabilities and functioning, in the evaluative space. Health would be valued in its own right and not for how it makes the individual feel. Other information in addition to utilities would be taken into consideration when comparing states of the world. Therefore the initial aim of extra-welfarism was to broaden the valuation scope, enabling more than utility to be taken into account. Brouwer et al (2008) also differentiate extra-welfarism from welfarism because outcomes do not necessarily need to be weighted according to preferences. Indeed, more than solely the individual’s views are said to be considered in extra-welfarism and comparisons between individuals on a range of wellbeing measures can be made. Birch and Donaldson

(2003) are among many researchers who have reported the flaws in extra-welfarism and have attempted to refute some of these differences in their paper which set out to identify what is 'extra' in extra-welfarism.

Culyer (1991) deemed health to be one of the most relevant characteristics to take into consideration when valuing healthcare, as an individual's endowment of health (health characteristic) could be directly taken into consideration when comparing resource allocation for healthcare. Therefore outcomes were not solely weighted based on preferences. Although originally aiming to broaden the evaluative scope, in practice the extra-welfarist framework tends to focus only on health and has led to changes in health being considered instead of changes in utility (Hurley, 1998). The key value judgement behind extra-welfarism is to maximise health. Capabilities are not considered, but only the functionings, i.e. being in good health or being mobile, are considered rather than what the individual is capable of doing as a result of healthcare (Coast et al., 2008b). The extra-welfarist approach is argued to have an even narrower evaluative space than welfarism and is considered to be a partial implementation of Sen's capability approach (Coast et al., 2008a). The extra-welfarist approach is operationalised through cost-utility analysis (CUA) (discussed in section 3.2.2). As outcomes are based on health, the evaluations cannot assess efficiency in terms of how much should be allocated to healthcare over other sectors, as in welfarism, but instead these evaluations can assess technical efficiency in terms of determining which resource allocation produces the greatest benefit at the same cost, or the same benefit at the lowest cost (Morris et al., 2007). In summary, the aim of extra-welfarism is to maximise health, by using community values to assess the impact of the change in resource allocation on health status.

The extra-welfarist approach is widely implemented in health economics and is recommended by decision-makers (NICE, 2004). Although extra-welfarism is routinely implemented in practice, it has been criticised to lack any substantial theoretical underpinnings (Birch & Donaldson, 2003; Coast, 2004; Tsuchiya & Williams, 2001). Health alone is argued not to be appropriate by many researchers in the field (Birch & Donaldson, 2003; Coast et al., 2008b; Ryan, 1999). Several interventions have benefits that either lie outside of health or are not confined to health, and current health related measures are likely to underestimate the benefits, which will have a negative impact on decision-making (Coast et al., 2008b). Further, the primary value judgement behind extra-welfarism of maximising health has been questioned (Anand & Wailoo, 2000; Coast, 2009). It has been argued that as the aim of healthcare is to improve health, then resources should be allocated in a manner that maximises health (Brouwer et al., 2008). However, this principle of maximising health, in particular, does not coincide with any theory and it has been widely reported that society does not agree with such a stance (Dolan et al., 2005). The actual criteria for distributing healthcare in practice is unclear. Coast (2009) explains that currently, distribution is not based on need because once health is produced it cannot be redistributed to ensure a fair allocation, as in welfarism. Instead, once health is produced, it is simultaneously consumed. She states that the extra-welfarists are "*endorsing the ethical position that the total sum of health produced within the healthcare system is what matters, no matter how that health is distributed*" (Coast, 2009 pp 789). For these reasons Birch and Donaldson (2003) also state that the same criticisms of welfarism, which are related to the hypothetical compensation and decisions based on potential compensation, apply to extra-welfarism. Coast suggests an alternative, possibly more appropriate, rule for the distribution of healthcare based on providing a *"threshold level of health […], rather than maximising the total level of health…"* (Coast 2009 pp 791).

These ‘limitations’ of welfarism, and extra-welfarism, led to the development of the capability approach which aims to more fully implement Sen’s work in health economics. This work is currently in the development stages. Specifically the capability approach focuses on what interventions enable the individual to do in terms of their capabilities, whether they choose to or not, aiming to ensure an ethical and equitable distribution of healthcare (Coast, 2008b). As the capability approach is not the focus of this thesis its theoretical underpinnings or the suggested outcomes will not be discussed any further. A summary of the differences between welfarism and extra-welfarism are presented in Table 3.1.

Table 3.1 Summary of welfarism and extra-welfarism

Welfarism	**Extra-welfarism**
Maximise social welfare (utility)	Maximise health
Affected individuals values should be considered	Societal values should be considered
Consequence of health and healthcare on utility is the most important outcome	Health itself is valued and is the most important outcome
Distribution and equity are considered to be separate	Distribution and equity are not separable
Allocative efficiency	Technical efficiency
Incorporates benefits wider than health*	Incorporates only health benefits

*But has been argued to be narrow due to the impact of change in resource only being incorporated in so far as its consequences on utility

3.3 Methodological Considerations

As the theoretical background to welfarism and extra-welfarism has now been described the methodological considerations associated with the application of these approaches through economic evaluation will be discussed next.

3.3.1 Welfarist Cost-Benefit Analysis and Willingness-to-Pay

CBAs are found in abundance in environmental economics, and other sectors but are less commonly conducted within health economics (Gafni, 2006). The more widespread use in environmental economics is likely to be attributable to the contingent valuation approach being primarily developed in environmental economics, and the National Oceanographic and Atmospheric Administration (NOAA) providing recommendations and guidelines for its use. Furthermore, contingent valuation is one of the only methods that can account for existence value which is particularly relevant to environmental economics, and is related to the benefit derived from knowing an environmental resource exists regardless of whether you want to visit (Mitchell & Carson, 1989). The less frequent use in healthcare is partly due to UK decision-makers, such as NICE, rejecting the welfarist approach in health arguing that decisions should not be based on ability to pay but on need, as mentioned previously, and recommending the use of the alternative extra-welfarist type of economic evaluation, namely CUA (discussed in 3.3.2). It is also likely to be related to the empirical limitations of getting people to attach a monetary value to health. Furthermore, not only are there issues with the use of the theoretical underpinnings of the welfarist approach, there are also methodological concerns with the application of the approach.

The welfarist principles are applied to decision-making through the CBA type of economic evaluation. Like all economic evaluations the CBA assesses the costs and outcomes related to a change in resource allocation, which could for example be the provision of a healthcare intervention. As mentioned previously, in welfarism, money is used as a proxy for measuring changes in the utility (outcome) following a welfare change. Thus in CBA both costs and outcomes are expressed as monetary values. A CBA aids policy-makers' decisions about whether an intervention is worth implementing. This would be the case if the benefits outweighed the cost, resulting in a positive net social benefit, producing a welfare gain.

As mentioned in section 3.1 the healthcare market does not resemble a normal market and therefore individual's preferences for healthcare cannot be measured by assessing their response to changes in price. Other measures such as stated preferences and revealed preferences can be used. The welfare change observed from the provision of an intervention can be measured directly through the contingent valuation stated preference technique. Alternative methods for eliciting preferences include the stated preference discrete choice experiment. These choice experiments present the individual with a scenario of various attributes and attribute levels of healthcare (i.e. the comparison of two treatments) and measure the amount an individual is willing to pay for a change in the level of an attribute of healthcare (McIntosh et al., 2010). The use of stated preference discrete choice experiments is growing in the literature and is at the development stages for eliciting preferences in health. Indirect methods, such as revealed preferences, are methods of eliciting the value for healthcare by assessing the influence the improved health has on another good (McIntosh et al., 2010). Revealed preferences are rarely used in health economics. As the welfarist contingent valuation method is the focus of this thesis, it will be discussed in more detail.

Direct and indirect elicitation methods

In the contingent valuation technique, an individual's utility is measured directly in terms of the amount of money they would be willing to give up (receive), to (not) receive an intervention, in order to compensate those individuals that lose out following the welfare change (O'Brien & Gafni, 1996). Hence, this compensation would restore social welfare to the original utility level prior to the change in resource allocation. The outcome measure typically used is known as willingness-to-pay ('WTP'), which is a measure of the amount of money the gainers are willing to give up for the treatment (Johannesson & Jonsson, 1991; Olsen & Smith, 2001). An alternative, less commonly used outcome is called willingness-to-accept ('WTA') where the amount of money the losers would be willing to receive to incur the welfare loss and forgo the intervention is assessed (Olsen & Smith, 2001; Johannesson & Jonsson, 1991). WTP is measured rather than WTA because a welfare gain from the treatment is more often assessed (Sach et al., 2007). In practice, healthcare decisions are related to the addition of a new programme and tend to measure welfare gain (WTP). WTP is also associated with more conservative answers as the responses are constrained to the respondents' income (Arrow et al., 1993; Diener et al., 1998). Hence outcomes of CBAs in terms of WTP will be discussed.

Elicitation perspective

The changes in utility, as captured by WTP, following a change in provision of healthcare can be measured by either the equivalent variation or the compensating variation. The 'correct' perspective to use is dependent on whether there is any uncertainty in the outcome (O'Brien & Gafni, 1996). That is, if uncertainty exists the ex-ante perspective should be used and where the values are given under conditions of certainty the ex-post perspective is said to be most appropriate. However, what is considered to be ex-ante and ex-post varies in the literature. It was initially stated that where there is uncertainty, the approach to be considered should be ex-ante (Klose, 1999). This can be taken to mean that those at risk of becoming ill and those at risk of requiring treatment are ex-ante (Mcintosh et al., 2010, O'Brien & Gafni, 1996). Therefore some argue that even if patients are considered but have yet to have treatment, they are ex-ante as there is uncertainty around the effect of treatment on their utility function (McIntosh et al., 2010; O'Brien & Gafni, 1996). Shackley and Donaldson (2000) define the ex-post perspective as patients that are diseased, in the process of waiting for treatment or currently having treatment.

Whose values?

As mentioned previously, whose values should be used to value healthcare has also been debated. O'Brien and Gafni (1996) state that use value (patients alone) is not consistent with theory as the entire population is affected not just patients. Thus both use and non-use (non-patients) values should be used to account for option value and caring externalities. Smith (2003) and Sach et al (2007) determined that use value is most commonly used as patients' values are used, in spite of use and non-use being theoretically preferred in CBAs.

Elicitation techniques

Several methods for eliciting WTP have evolved. Table 3.2 outlines the main methods and provides details of the issues associated with each elicitation method.

Table 3.2 WTP elicitation formats

Elicitation format	Elicitation method	Methodological issues
Open-ended questions the respondent is asked their maximum WTP for treatment	Questionnaire/ interview	• Large sample required • Requires a-priori information on distribution of WTP
A bidding game the respondent is asked whether they would WTP a certain amount for treatment and this amount is varied until the maximum WTP is identified	Interview	• Provides value cues – starting point bias • Large sample
Closed-ended question where the respondent is presented with a value and asked if they would be WTP	Questionnaire/ interview	• Provides value cues – starting point bias and yea saying • Requires a-priori information on distribution of WTP
Payment scale where a vertical list of monetary values are presented to the respondents and they are asked to circle their maximum WTP	Questionnaire/ interview	• Provides value cues – range bias • Possibly Requires a-priori information on distribution of WTP

WTP; willingness-to-pay

There are a range of variations to these elicitation methods which are outlined in McIntosh et al (2010). Each elicitation method has advantages and disadvantages: some are related to introducing bias by providing value cues to patients (all but open-ended), others are related to requiring very large sample sizes to obtain an appropriate distribution of WTP (bidding game/ closed-ended), or requiring a-priori information on the distribution of WTP to be able to ask the questions (bidding game, closed-ended, in some cases payment scale) (Johannesson & Jonsson, 1991; McIntosh et al., 2010). The introduction of bias through the provision of value cues is an issue in WTP studies because the respondent's final WTP values are said to be influenced by the values presented, and therefore the final WTP value may not reflect their true WTP (Whynes et al, 2004). Instead the values elicited are related to the elicitation method used. However methods to overcome or limit the potential for bias are suggested in McIntosh et al (2010). A disadvantage of the open-ended question format is that it results in a large number of protest answers or non-response due to the way in which the question is asked as it provides little guidance (Arrow et al., 1993; Donaldson, 1997b). Protest answers relate to the respondent misunderstanding the exercise and refusing to provide a WTP value. The payment scale has been shown to have a higher completion rate than the open-ended format (Whynes et al., 2003). The payment scale (card) and bidding game are the most commonly used elicitation formats (Sach et al., 2007).

It has also been found that the maximum average WTP differs depending on the elicitation method (Frew et al., 2003) and a consensus has not been reached on which elicitation method is most suitable. This is one of the obstacles to implementing WTP. Another obstacle is the lack of agreement on whether an interview or questionnaire format is most suitable. The use of interviews is thought to introduce interviewer bias, whereby individuals may not provide their real WTP values and may be influenced by the interviewing situation (Johannesson & Jonsson, 1991; McIntosh et al., 2010). Recommendations from environmental economics are often used as a basis to suggest that interviews are the most suitable format. However, it can be argued that interviews are not required for the following reasons: healthcare goods are not as alien to respondents as goods in environmental economics and therefore do not need such a rich context of the programme or the market to elicit meaningful WTP values and also because private payment is also taken for some interventions (Smith, 2003; Smith, 2006).

Whether WTP should be elicited as a one off, yearly or monthly payment or in terms of out of pocket cost or taxation is another issue of debate. There appears to be no consensus on which payment vehicle should be used but it has been found that taxation leads to a greater number of protest responses (Donaldson, 1997a; Klose, 1999). The out of pocket payment is the most commonly used type when comparing pharmaceutical interventions (Sach et al., 2007). The type of payment vehicle used is dependent on whether healthcare is privately or publicly financed, and whether option value, that is providing a valuation to be able to have the intervention if the respondent should need it, is elicited.

As outlined in McIntosh et al (2010) other more general issues that apply to all applications of WTP are related to individuals providing inconsistent responses which can be related to:

- Strategic bias where a respondent values a treatment but provides a protest answer or very low value because they think they will have to pay, or the respondent provides a high WTP value to ensure the intervention is implemented (Johannesson & Jonsson, 1991)

- The embedding effect has been observed where individuals are unable to differentiate between two treatments and instead provide a value for any treatment (Kahnemann & Knetsch, 1992)

- The warm glow effect can also occur where true preferences are not elicited but the respondent provides WTP values for a good cause and receives moral satisfaction from the act (Kahnemann & Knetsch, 1992).

Approach to Costing

According to welfare (CBA) theory as the benefits are broad and can include both health and non-health in so far as they impact on utility, broad costs impacted on healthcare and society are also argued to be important to take into consideration in the analysis (McIntosh et al., 2010). Societal costs include costs to employers which are related to patient time off work due to illness (productivity costs) and patient costs associated with travel for healthcare, amongst others (Gray et al., 2010). Very few studies have carried out a full CBA in healthcare which includes costs associated with healthcare and society. A recent (partial) CBA has been carried out by Haefaeli et al (2008) in healthcare as part of their feasibility study. The use of societal costs is widely debated and the most appropriate method of handling certain societal costs, such as productivity costs is not entirely clear. The reason for such a debate is related to decision-makers recommending that as healthcare resources are allocated within a fixed healthcare budget, costs beyond healthcare are not of primary importance to decision-makers (NICE, 2008). However, these costs are important in other countries, such as the Netherlands, and are a part of the reference case. It should be noted that productivity costs are more commonly and more readily incorporated into CBAs, relative to CUAs, as both costs and outcomes are reported in monetary units. As the collection of productivity costs includes an additional burden on patients, and decision-makers do not recommend societal costs, they are likely to be rarely collected which will mean that they will not be readily available to analysts. Further, it should be noted that the incorporation of societal costs is likely to be more important for certain conditions, such as those which require carers because societal costs are likely to be more prominent in these cases (McIntosh et al., 2010).

Finally it is widely reported that CBAs lack any reporting guidelines (McIntosh et al., 2010). The development of formal guidelines may lead to an increase in CBAs in healthcare.

3.3.2 Extra-welfarist Cost-Effectiveness Analysis and Quality-Adjusted Life Years

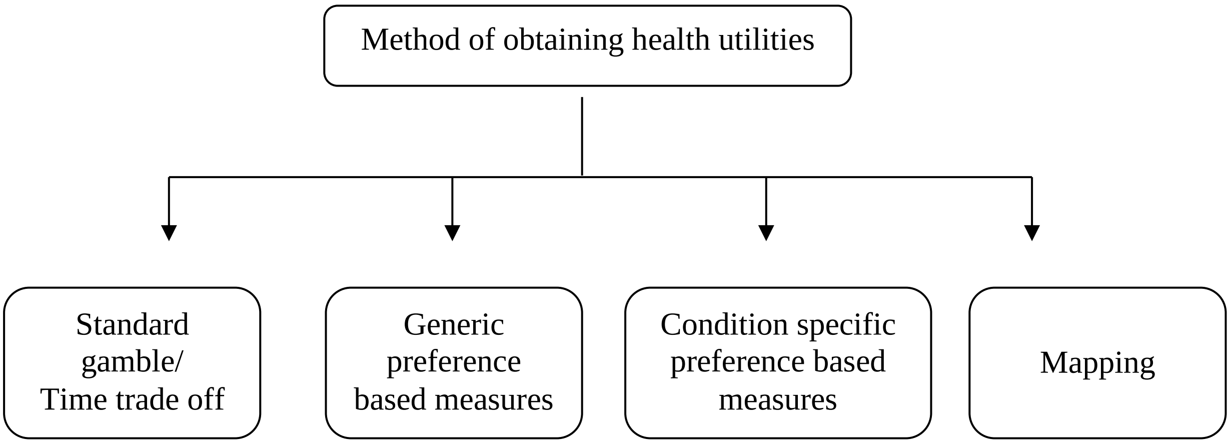
Similar to approaches of WTP and CBA, which comprise the welfarist framework, there are also methodological issues with the application of the extra-welfarist approach. Whilst within the extra-welfarist framework utility is substituted for health as a measure of assessment of welfare change, the notion of utility does still exist as (health) utility-based measures can be used to obtain the measure of change in health following a change in resource allocation. A broad range of measures are available to assess change in health-related quality of life (QoL), including those specific to particular conditions, and those that are generic. NICE explicitly recommend the use of generic outcomes to achieve a level of consistency when making resource allocation decisions across healthcare conditions. Specifically, NICE recommend the use of the quality-adjusted life year (QALY) as the unit of outcome (NICE, 2008).

The extra-welfarist framework is applied to the decision-making process through the use of economic evaluation either by using the cost-effectiveness analysis (CEA) or the CUA. Internationally these terms are used interchangeably, but in the UK a CUA assesses costs and outcomes, where the outcome measure used is a QALY. The QALY is a measure that incorporates both quality and quantity of life. The patient's QoL is measured on a (utility) scale where 1 represents full health, 0 death and values below zero reflect states worse than death (Drummond et al., 2005). The utility value is then combined with information on length of life to estimate QALYs. Unlike the CBA, in CUAs only costs are measured in monetary terms as outcomes are measured in QALYs. Data on costs and QALYs are combined to compare the associated costs and outcomes of the new intervention against standard practice. Interventions that lie within an arbitrary threshold of £20,000 to £30,000 per QALY are considered to be cost-effective and should be implemented in clinical practice (NICE, 2008).

The advantages of the QALY are related to health utilities being derived using the same 0 (death) to 1 (full health) anchors, so the QALYs gained across a range of conditions can be compared (Gold et al., 1996). Further, the length of time spent in a state is assessed in combination with the health related utility associated with the state. However, further to concerns discussed in the theoretical section about using health alone in the evaluative space, Tsuchiya and Dolan (2005) identified issues specific to the use of QALYs. These include concerns with the assumption that the value of a health state is constant irrespective of time spent in the health state and that the value of the health state is independent of subsequent or prior health states.

Possible methods for eliciting utilities and generating QALYs are outlined in Figure 3.1 and are discussed next.

Figure 3.1 Summary of key methods to derive utilities in extra-welfarism

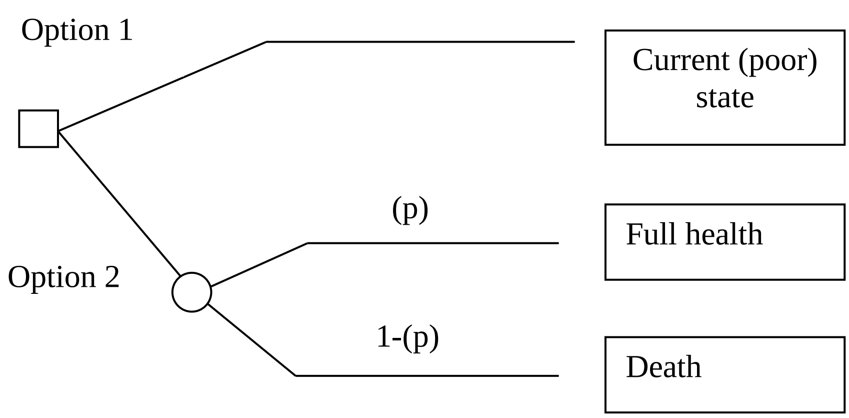


Direct elicitation methods

The utility value for different health states is derived through the use of preference-based measures. This valuation or weighting of health states is related to utility theory and therefore produces (health) utilities. Individuals are provided with a description of a health state and then directly or indirectly express their preferences for that health state. The recommended measures used to elicit these preferences are those that are choice based including time trade off (TTO) and standard gamble (SG), with standard gamble being the most in line with the von- Neumann-Morgenstern utility theory (Drummond et al., 2005; von Neumann-Morgenstern, 1944). A third measure, which will not be discussed in detail, is the rating scale and the most common is the visual analogue scale (VAS) (Parkin & Devlin, 2006). The VAS is not choice based and is thought to provide weaker evidence than the other measures (Brazier & Rowen, 2011). Choice based measures such as SG and TTO are the preferred methods and are recommended by NICE because they incorporate opportunity cost into the assessment, with SG also taking into account risk attitudes.

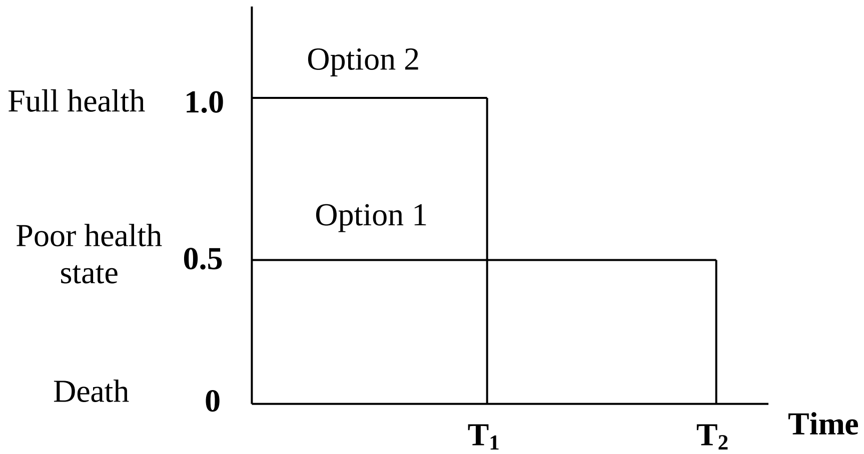
Standard gamble uses risk and probability to assign weights. The technique presents an individual with a choice between two alternative health states. As shown in Figure 3.2, a poor health state is presented to the individual; they can either remain in this state or have an intervention that has a probability of returning to a health state in full health, and a probability of immediate death. Utility values are assigned to the described poor health state by identifying the probability at which the individual is indifferent between the alternative options (Drummond et al., 2005).

Figure 3.2 Illustration of SG (based on Drummond et al., 2005)



Time trade-off differs to SG in that uncertainty and probability is not used (Torrance, 1976). As shown in Figure 3.3, an individual is presented with a poor health state for a certain amount of time, and a health state in full health for a shorter period of time. The time in full health is varied to identify the point at which the individual is not willing to trade off any more years of life in full health to move out of the poor health state. The utility value is then obtained by calculating the amount of time at which the individual is indifferent and the utility for the poor health state is estimated by the amount of time in full health state divided by the amount of time in the poor health state (Drummond et al., 2005). For example, 20 years in full health is equivalent to 40 years in poor health (20/40=) 0.5 utility value.

Figure 3.3 Illustration of TTO (based on Drummond et al., 2005)



In the UK, preferences are elicited using values derived from the community, as NICE states, utilities should be based on the general public preferences and not the condition-specific patient's preferences alone (NICE, 2008).

When both SG and TTO are used, the respondent is asked to imagine they are in the health states and provide their own values, similar to use values for WTP. Similar to the welfarist

WTP, the elicitation of preferences using SG and TTO have also been found to produce different results (Tsuchiya et al., 2006). The way in which the SG and TTO are presented also lead to different results not only between the methods but also within the same choice based technique. This effect is also observed with WTP (Dolan, 1997).

Indirect elicitation methods – generic measures

Typically, utility values are not derived for every possible health state associated with all diseases and conditions using these direct methods. Instead multi-attribute utility instruments (MAUI) such as the EuroQol-5 Dimension (EQ-5D) are used. Currently EQ-5D is recommended by decision-makers (NICE, 2013). Generally, these multi-attribute systems are generic and therefore present descriptions of health states that can be used and compared across conditions. Thus, weights, based on preferences, are assigned to alternative characteristics of health states and are then used to optimally allocate resources in order to maximise health (Coast et al., 2008b).

A questionnaire is constructed with a set of health related domains or attributes (i.e. mobility, pain, self-care, as in EQ-5D) and each domain has a set of levels (3 or 5 for EQ-5D) to describe the extent of the problem associated with each domain. Information from patients was then used to generate a set of descriptions for a range of health states. The general population then valued a sub-set of these health states using either SG or TTO (EQ-5D uses TTO) and an algorithm was applied to generate values for the remaining health states. Thus, each possible health state associated with the questionnaire has a utility value assigned to it. The patient with the condition under investigation is then able to complete the questionnaire and an analyst can attach a utility value for the patient’s health state by using the appropriate algorithm, which accounts for public preferences, to obtain an overall EQ-5D utility index score for the patient (Dolan et al., 1995). This overall utility score is then multiplied by the length of time the patient spends in the health state in order to produce a QALY.

Currently, there is a range of generic health related utility-based QoL instruments (multi-attribute systems) available that can be used in all conditions. These measures are used in CUAs and include, EQ-5D (5L and 3L); Short Form 6 dimension (SF-6D), Health Utility Instrument (HUI2 &3); the Assessment of QoL (AQOL), and 15D (Drummond et al., 2005). Unlike the other preference based measures discussed previously, SF-6D was not originally designed as a preference based measure. That is, SF-6D was developed from a non-preference based measure (SF-36). SF-6D was developed by obtaining preference weights, using SG from the general population, for the various attributes to assess the strength of preference for the different health states (Brazier et al., 1998). Hence, SF-6D is generated by mapping (explained later) the non-utility based Short Form-36 (SF-36) or SF-12 on to SF-6D. Table 3.3 describes the measures that have UK value sets for the adult population.

Table 3.3 Generic preference based measures with adult UK value sets

Instrument	Attribute	Recall period	Preferences	Scoring
EQ-5D *(hrQoL)*	1. Mobility 2. Self care 3. Usual activities 4. Pain/ Discomfort 5. Anxiety/Depression	• Health today	TTO	Each item rated on a 3 or 5 point scale depending on measure used: 1. No problems 2. Some problems 3. Unable Or 1. No pain/anxiety 2. Moderate pain/anxiety 3. Extreme pain/anxiety **5L** 1. No problems 2. Slight problems 3. Moderate problems 4. Severe problems 5. Unable
SF-36/ SF-12 *(hr QoL)*	1. Physical functioning 2. Role physical 3. Bodily pain 4. General health 5. Vitality 6. Social functioning 7. Role – emotional 8. Mental health	• In general • Past 4 weeks	N/A	Must first be mapped onto SF-6D in order to be used in a cost-utility analysis Can be summarised into; 1. Mental health (MCS) 2. Physical Health (PCS) SF-36 has 36 questions, whilst SF-12 covers the same attributes but has only 12 questions
SF-6D *(hrQoL)*	1. Physical functioning 2. Role limitations 3. Social functioning 4. Pain 5. Mental health 6. Vitality	N/A – mapped from SF-36/SF-12	SG	Each item rated on a 4-6 point scale Algorithm provided by John Brazier
HUI2 *(hrQoL)*	1. Sensory 2. Mobility 3. Emotion 4. Cognition 5. Self-care 6. Pain	• Current – specific recall period or • Usual health	VAS mapped onto to SG	Each item rated on a 4-5 point scale

Hrqol; health related quality of life, N/A; not applicable, TTO; time trade off, SG; standard gamble, VAS; Visual analogue scale. (Adapted from Brazier & Rowen, 2011)

Despite NICE recommending EQ-5D to generate QALYs as a reference case, it is recognised that EQ-5D may not be appropriate in all cases (NICE, 2008). When EQ-5D is not appropriate NICE recommend that a choice based method be used (Brazier & Rowen, 2011; NICE, 2008). EQ-5D may not be suitable because it may not be sufficiently sensitive or relevant for all conditions. Brazier et al (2004) have shown that ceiling effects have been observed, where many respondents have scored the highest level of EQ-5D. It is also recognised that some researchers and clinicians do not consider EQ-5D to be a suitable measure, as clinicians are typically interested in condition specific outcomes (Brazier & Rowen, 2011). It has been found that the two most commonly used preference based measures SF-6D and EQ-5D do not generate the same results. The utility values for the same patient have been shown to differ and the reasons for this are likely to be related to the measures capturing different aspects of QoL and the valuations for the utilities being derived by different methods i.e. SG for SF-6D and TTO for EQ-5D (Brazier et al., 2004, Whitehurst & Bryan, 2011). Whitehurst et al (2011) also showed that at a group level, the mean values differ which provides warnings against drawing comparisons between studies that have used SF-6D and those that have used EQ-5D. Furthermore, the degree of the cost-effectiveness of competing interventions has also been shown to differ across the two measures (Davis et al., 2012; Sach et al., 2009).

Indirect elicitation methods – disease-specific measures

When a generic utility based measure is not considered to be suitable for a condition, disease-specific utility based measures are increasingly developed (Drummond et al., 2005). Due to their nature, disease-specific measures are more sensitive to changes than generic measures. There is a range of condition specific preference based measures available. For NICE to use the condition specific measure it must:

1. Focus on condition-specific health related QoL, not just on symptoms of the condition

2. The valuations of the states must be provided from the UK general population

3. Choices should be based on SG or TTO

4. The measures must be anchored by (0) death and (1) full health as generic measures, to be able to generate QALYs (Brazier & Rowen, 2011).

However, condition-specific utility-based measures are typically anchored by full health (1) and the worst possible state for the condition (0), rather than full health (1) and death (0) as in generic measures. The disease-specific Menorrhagia Multi-attribute Scale (MMAS), discussed in Chapter 2, falls into this category (Shaw et al., 1998). These types of measures cannot be used to generate QALYs, which are currently recommended by decision-makers, as mortality has not been taken into consideration when the preference based technique was conducted. Further, patient values, rather than the general population values are typically used, despite NICE's recommendations for UK general population values. A specific obstacle for the wide use of MMAS is that the choice based method used was not based on SG or TTO. Currently economic evaluations which use measures such as MMAS, which cannot generate QALYs, are considered to be sophisticated CEAs (Drummond et al., 2005). At the very least, as preferences for the different health states have been identified, incremental changes on the disease-specific scale can be considered to be more meaningful than changes on non-preference based disease-specific measures. An increase in score does mean a preferable change to a 'better' state than previously, enabling scores across respondents to be more readily compared (Brazier, 2005).

Preference based condition specific measures for cancer and asthma that do meet NICE's criteria have been developed (Brazier & Rowen, 2011; Yang et al., 2011). Recent studies have placed valuations for health states on the typical 0 (death) to 1 (full health) scale using the UK general population values, enabling the generation of QALYs (Brazier & Rowen, 2011). In these cases the disease-specific utility based measures provide the benefits of utility measures, enabling comparisons to be made across conditions, and the sensitivity required for the specific condition (Drummond et al., 2005). However, much research has yet to be carried out within this area before it can become a viable alternative to generic utility-based measures. Concerns with condition specific preference based measures typically include their limited ability to incorporate co-morbidities and side effects and some measures can also be criticised for overly focussing on domains that are specific to the condition and not on domains that are unaffected by the condition, which could lead respondents to overstate values (Brazier & Rowen, 2011).

Mapping

Mapping is typically considered to be the process of obtaining values for a utility based instrument from a non-utility based measure (Brazier et al., 2010). In cases where generic preference based measures are not appropriate for a particular condition and condition specific preference based measures are not available, non-utility based measures can be used to generate utilities by mapping. This can be carried out by either mapping, also known as 'cross-walking', the results of the condition-specific measures or indicators onto preference based measures, such as EQ-5D. A robust mapping method requires the collection of both the utility based measure (i.e. EQ-5D) and the condition specific measures, or indicators, (i.e. MMAS) to generate an algorithm that predicts the relationship between the measures, for example, enabling MMAS to be mapped onto EQ-5D. This algorithm can then be applied to any study where a condition-specific measure or indicator has been used and a utility based measure has not. As utilities are not directly elicited and mapping relies on the robustness of the algorithm, it is recognised to be the 'second-best' approach to generating utilities (Longworth & Rowen, 2013). A further limitation of mapping is that a prior sample study must be available to establish the relationship between the measures before a condition specific measure can be mapped onto a utility based instrument. Further, in some cases it is not possible to carry out mapping where there is insufficient overlap in the descriptive content of the measures (Longworth & Rowen, 2013).

Box 1 outlines the key recommendations provided by NICE for valuing health states and generating QALYs.

Box 1. NICE recommendations (adapted from Brazier & Rowen, 2011)

Key NICE recommendations for valuing outcomes:

1. Choice based method using either SG or TTO and UK general population values in a **generic MAUI** to generate QALYs
    - Specifically EQ-5D to enable comparability *(the use of any other technique must be justified including SF-6D)*

- **Condition specific MAUI** (valued using SG/TTO) using UK general population values
- **Mapping** based on empirical data, not opinion
- Additional analyses using condition-specific measures and patients values can be used where justification is provided.

MAUI; multi-attribute utility instrument, SG; standard gamble, TTO; time trade off.

It is clear from the discussion presented here on extra-welfarist measures, that more applied research has been carried out and reported in extra-welfarism than in welfarism, as prescriptive guidelines have been produced on how to report and conduct CUAs. Many CUAs have been carried out in the vast majority of disease areas to value interventions for decision-makers. These evaluations are primarily taken from a healthcare perspective and not a societal perspective, for reasons mentioned previously, despite some researchers advocating the use of the societal perspective for a CUA to ensure decision-makers are presented with all available information (Brouwer et al., 1997; Sculpher, 2001).

Table 3.4 presents a summary of the outcomes and economic evaluations associated with the welfarist and extra-welfarist framework.

Table 3.4 Summary of application of welfarist and extra-welfarist approaches

Framework	**Economic evaluation**	**Efficiency**	**Outcome**	**Method of obtaining preferences**	
				Directly	**Indirectly**
Extra-welfarism	CUA:	productive efficiency	QALY*	• TTO • SG	• EQ-5D* • SF-6D • Preference based condition specific • Mapping
Welfarism	CBA:	allocative efficiency	Monetary amount	• CV • DCE	• Revealed preferences

*recommended by decision-makers. CV; contingent valuation, QALY; quality adjusted life year, CUA; cost-utility analysis, CBA; cost-benefit analysis; DCE, discrete choice experiment

3.3.3 Vehicles for Economic Evaluation

Decision-modelling can also be used as a vehicle to carry out an economic evaluation. The costs and outcomes are entered into the model, and it is used to determine the cost-effectiveness of alternative interventions. The use of decision-models can be thought of as *'providing a framework for decision-making under conditions of uncertainty'* (Drummond et al., 2005 pp 276). Decision-modelling is increasingly applied to healthcare decisions due to the advantages which are outlined by Buxton et al (1997). For example, when economic evaluations of competing interventions are required and economic data are not collected alongside a trial, decision-models can be used to synthesise data on costs and effects from multiple sources (Buxton et al., 1997). Further, when trial data are available but the trial time horizon does not reflect the full disease process, decision models can be used to project costs and outcomes beyond the trial time horizon (Buxton et al., 1997). Hence they can be used to determine the long-term cost-effectiveness of interventions. Drummond et al (2005) also noted that decision-models can be used to identify gaps in current evidence by assessing uncertainty, therefore informing further research. It is stated by Briggs et al (2006) that decision-models can be used in both CBAs and CUAs, however the current evidence would suggest that decision-models have primarily been used in an extra-welfarist framework. Hence Brennan et al (2006) state that the model-based economic evaluation output typically includes the costs of the competing treatments and the QALYs accrued by each intervention.

Decision-makers have recognised the benefits of decision-models and have incorporated their use into the guidelines for economic evaluations (NICE, 2004). NICE state that decision-models can be used to inform decision-making where appropriate. The assumptions of the model and its justification must be presented in the report along with an assessment of the structural uncertainty of the model and the uncertainty associated with the model input. Guidelines of good practice in modelling have been developed by Philips et al (2004) and are expected to be followed. However the structure of the model is dependent on the ability of the analyst to reflect the disease pathway adequately and the availability of data (Brennan et al., 2006).

There are several types of models, which include those that model cohorts of patients, such as decision trees and Markov models where the average values for all patients is obtained and those that model at the individual patient level, such as individual sampling models (Brennan et al., 2006; Briggs et al., 2006). Each type of model has its own properties and assumptions. Each will be discussed in turn next:

1) The **decision tree** is one of the most commonly used decision models and is also one of the most simplistic. The alternative treatment options for the disease are first

presented and from there on the subsequent pathways or events associated with that treatment, such as improvement in health or death, are presented with the associated probabilities of each occurring (Briggs et al., 2006). The decision tree then presents the subsequent pathway associated with each event, i.e. improved health or death. Hence the probability of subsequent events occurring is dependent on previous events. The mean value is then calculated by multiplying the probabilities for each event by the costs and outcomes for each treatment (Brennan et al., 2006). However, there are limitations associated with the decision-tree as it does not allow for recurrent events. The amount of time spent in each state is also not taken into account and the decision-tree does not readily allow for complex disease pathways, with multiple consequences, that occur over a long time horizon (Briggs et al., 2006).

2) A **Markov model** reflects several health states or consequences of treatment, where the probability of transitioning between health states occurs according to the time cycle of the model (Briggs et al., 2006). The time cycle reflects a set period of time in which transitions between health states can occur i.e. on a monthly basis. Each state is then given an average cost and outcome, which is summed and multiplied by the proportion of people in each state. Markov models are therefore used for more complex disease pathways and when recurrent events are required to be modelled. However, the limitation of these models is the Markov assumption that patient history is not taken into consideration, as the prior health states and the amount of time since transitioning to the subsequent states are not taken into account (Briggs et al., 2006). These assumptions can be overcome, but the analysis becomes complex and the limitations are outlined by Briggs et al (2006).

3) **Individual sampling models** are the most complex of the three key models discussed here and build upon the limitations of the previous models. As mentioned previously, individual sampling models are used to model patients at the individual level. Hence, future transitions to health states can be dependent on the patient's history, which in turn will have an effect on costs and QoL (Briggs et al., 2006). However, these models do require a vast amount of data, are computationally complex and it has been suggested that they do not easily allow for a comprehensive assessment of uncertainty in the parameters to be carried out through sensitivity analysis (Briggs et al., 2006).

The uncertainty of results obtained from decision-models can be assessed in sensitivity analyses. These can be either deterministic or probabilistic. In a deterministic sensitivity analysis, model parameters are changed individually. Hence one parameter is changed and the others are not (Andronis et al., 2009). In probabilistic sensitivity analyses, model parameters are altered simultaneously according to a given distribution. Each parameter is assigned a distribution and a value for each parameter is randomly drawn from the distribution and the effect on the cost-effectiveness of varying multiple model parameters can then be assessed (Briggs, 2000).

3.4 Conclusion

In this chapter the two main theoretical foundations behind economic evaluations and their methodological considerations have been discussed. The concerns related to applying both welfarist and extra-welfarist theoretical frameworks in practice outlined and the similarities in the issues associated with the application of the frameworks, described. It is explained that due to the perceived reliance of the welfarist framework on ability to pay, amongst other things, the extra-welfarist approach is recommended in the UK.

As the condition of menorrhagia has been described in Chapter 1 and the potential outcome measures that can be used to assess changes in QoL for economic evaluation outlined here, in the next chapter a systematic review on the economic evaluations conducted in menorrhagia using utility-based measures is provided.

CHAPTER 4. SYSTEMATIC REVIEW: ECONOMIC EVALUATIONS IN MENORRHAGIA

4.1 Introduction

This chapter presents a systematic review of pharmaceutical interventions for menorrhagia. The objective of the review is to assess the evidence on cost-effectiveness for the different pharmaceutical interventions for menorrhagia, including levonorgestrel-releasing intrauterine system (LNG-IUS). The focus is on pharmaceutical interventions because surgical interventions are no longer recommended as first line treatment for menorrhagia. It is necessary to conduct this systematic review in order to collate the existing evidence on cost-effectiveness, identify any limitations of studies, conflicting results or methodologies, and areas of uncertainty that should be investigated prior to conducting the economic evaluation alongside the ECLIPSE trial. It is intended that the evaluations or model structures identified in this review could then be used to inform the economic evaluation for the ECLIPSE trial.

The National Institute of Health and Care Excellence (NICE) issued clinical guidelines on heavy menstrual bleeding in 2007. As an economic evaluation is reported in the NICE guidelines, it is critiqued along with the other studies identified, but first the main recommendations from the NICE guidelines are briefly described in section 4.1.1. Following the summary of the NICE guidelines, the methods of the review are described including eligibility criteria, search strategy and data extraction criteria. The results are then reported as a critique of the studies, followed by the discussion and conclusion.

4.1.1 NICE Guidelines

It is important to summarise and critique the NICE guidelines for menorrhagia because they are developed to provide recommendations on which treatments should be used in clinical practice. To provide these recommendations, NICE commissioned a systematic review of the literature on treatment effects and cost-effectiveness in menorrhagia to be carried out. Typically, in areas where there is little evidence available on cost-effectiveness, an economic evaluation is conducted and reported as part of the NICE guidelines. The NICE review identified one study that considered the cost-effectiveness of LNG-IUS compared to a hysterectomy (Hurskainen et al., 2004). No studies considering the cost-effectiveness of the other pharmaceutical treatments (combined oral contraceptives, oral progestogens, tranexamic acid, mefenamic acid and norethisterone) in menorrhagia were found. An economic evaluation was therefore conducted to determine the cost-effectiveness of all the currently available treatments for menorrhagia (NICE, 2007).

The results in the NICE guidelines showed that when the need for contraception is not taken into account and all of the pharmaceutical treatments (LNG-IUS, mefenamic acid, tranexamic acid and combined oral contraceptives) are compared against watchful waiting, LNG-IUS is most cost-effective (NICE, 2007). When hormonal treatment (LNG-IUS and combined oral contraceptives) is compared to surgical treatment (hysterectomy and endometrial ablation), LNG-IUS is the most cost-effective. When contraception is not sought and non-hormonal treatment (mefenamic acid and tranexamic acid) is compared to watchful waiting, tranexamic acid is found to be the most cost-effective intervention. The results of the sensitivity analysis showed that the policy decision did not change in any case. A summary of the NICE recommendations is presented in Box 4.1.

Box 4.1 Summary of NICE recommendations (NICE, 2007)

<u>Summary of NICE Recommendations</u>

1. LNG-IUS as first line treatment;
2. Second line treatment is either tranexamic acid, mefenamic acid or combined oral contraceptives;
3. Third line is oral progestogens (norethisterone), or injected progestogen;
4. Where pharmaceutical treatments are unsuccessful, the woman does not want to conceive and QoL is severely impaired, surgical interventions should be considered.
    a. First line surgical treatment is endometrial ablation
    b. A hysterectomy should only be considered as a final option.

LNG-IUS; levonorgestrel-releasing intrauterine system, QoL; quality of life

4.2 Methods

The guidelines by the Centre for Review and Dissemination and Cochrane Collaboration for Reviews were followed in the review reported in this chapter (CRD, 2009; Higgins & Green, 2009).

The literature is updated using June 2006 as a starting point. Reasons for updating the literature are explained in section 4.5.2. In the process of updating the review in June 2013, two other systematic reviews within the area were identified (Blumenthal et al., 2011; Gemzell-Danielsson et al., 2013). After reading these reviews it was still deemed necessary to carry out the systematic review reported in this chapter as neither had assessed the quality of the available evidence. Further details of these published reviews are discussed in section 4.5.2 of the discussion.

4.2.1 Eligibility Criteria

Since the objective of the review was to assess the evidence on cost-effectiveness for pharmaceutical treatments in menorrhagia, studies were included in the review if they met the following criteria:

i. Report an economic evaluation, such as a cost-utility analysis (CUA) (termed cost-effectiveness analysis in international studies), cost-benefit analysis or effectiveness studies that assess costs and include a measure of quality of life (QoL), willingness-to-pay (‘WTP’)

ii. The interventions compared must include either LNG-IUS, combined oral contraceptives, tranexamic acid, mefenamic acid, norethisterone or long-acting progestogens for menorrhagia. Therefore, if a pharmaceutical intervention is compared to a surgical intervention the study is eligible for inclusion in the review.

- The population of interest was women with menorrhagia
- Studies that focused on adolescents (less than 15 years old) were not included. If menorrhagia is experienced in adolescents, it is likely that it will subside after puberty and the typical population of sufferers are predominantly women approaching menopause (Duckitt & Shaw, 1998)
- Studies involving patients with adenomyosis and endometriosis were excluded because menorrhagia is typically secondary to another symptom
- Studies including women with bleeding disorders such as Von Willebrand disease were excluded because the relationship between the impact of this on QoL compared to the impact of menorrhagia alone on QoL has yet to be established and treatment for the condition differs (James et al., 2004; Kujovich, 2005).

In all other cases, studies were included if menorrhagia was the presenting complaint.

4.2.2 Databases and Search Terms

In November 2010, the following databases were searched from 2006 onwards; Medline (1946 onwards), EMBASE (1947 onwards), PsycINFO (1967 onwards), Social Science Index, SSCI, Database of Abstracts of Reviews of Effects and NHS EED. Reference lists from eligible studies that were selected were also reviewed. The review was updated in June 2013.

The search terms used were heavy menstrual bleeding **or** menorrhagia **or** HMB **and** cost **or** economic **and** levonorgesterel releasing intrauterine system **or** LNG-IUS **or** levonorgesterel intrauterine system **or** oral contraceptive **or** tranexamic acid **or** mefenamic acid **or** progestogen **or** norethisterone. Both index terms and free text terms were used. Search terms used for each database including Boolean operators are presented in Appendix 1.1. The results were managed using Reference Manager Software (version 12).

4.2.3 Selection of Papers for Review

The review of the papers followed a two-stage method used by Mugford (1996) which has subsequently been established and described elsewhere (Bricker et al., 2000; Roberts et al., 2002). Firstly each study was categorised based on the title and abstract using the eligibility criteria. Then the full paper was obtained if the title and abstract was deemed relevant. Where no abstract was available, the full paper was obtained. All studies were reviewed in both stages by this author. To validate this process, 20% of the titles and abstracts were independently reviewed by two other researchers (Tracy Roberts and Emma Frew). All of the investigators were blind to the categorisation decisions.

***Stage 1** – Initial Categorisation of studies*

The following criteria were used to assess the relevance of each study in the first stage of the review:

A) Primary research is reported and includes an economic evaluation

B) The study discusses the economic aspects of care and contains useful primary or secondary cost or utilisation data

C) The study may have useful information but does not obviously fall into (A) or (B)

D) The study has no relevance

Studies in category (A) and (B) were considered relevant to the systematic review. 10% of the studies in category (C) were reviewed for relevance and those studies in category (D) were not reviewed any further.

***Stage 2** – Further categorisation*

The studies categorised in (A), (B), and (C) were further categorised after reading the entire article into the following categories:

1. Economic Evaluation

2. Cost study which reports QoL

3. Review article, describes results of previously published economic data

4. Not relevant

Studies classified into A(1), A(2), B(1) and B(2) were carried forward and assessed according to their quality.

4.2.4 Data Extraction and Management

The quality of the economic evaluations was assessed according to the recommended criteria presented in Drummond et al (2005) as it is more comprehensive than other available checklists. Where a decision model was used, the study was assessed according to the recommended criteria outlined in the checklist by Philips et al (2004). Data were then extracted from all of the studies with an indication of the quality of the study. The data extraction criteria based on the key areas of the checklists is presented in Table 4.1. In addition to the key criteria in the checklists, the country of the study, funding source and whether the mean or median values were used as data sources in the model were also assessed, though the country of study and funding source were not used to assess the quality of the studies but were extracted for discussion purposes. Data extraction was managed in Microsoft Excel (v2010).

Table 4.1 Data extraction criteria

Criteria	*Justification*
Country	*It is important to determine generalisability and whether the data sources are appropriate for the country of the study.*
Currency	*Was the currency reported suitable based on the data source used? Are the findings generalisable to other settings?*
Primary or secondary study?	*The findings from primary data can be considered to be more robust, provided the study is carried out appropriately.*
Data source	*Where is data taken from for costs, outcomes and transition probabilities? Was the source suitable in terms of country? The validity of the study is limited if evidence is taken from unsuitable sources and extrapolated.*
Costs	*Was the most current cost year for the study used? Were all data inflated to the same price year? Were discount rates applied as is widely recommended?*
Perspective of the analysis	*Healthcare service or societal perspective? Are the data used consistent with perspective?*
Outcome measures	*Are QALYs or WTP the main outcome measure? How were QALYs derived, was the same valuation method used for all utility inputs? There is evidence to show that the results of the different valuation methods are not comparable.*
Mean or Median values used	*Are utility data based on medians or mean values? Is justification provided for the value used?*
Model type	*The use of the most suitable model type will ensure a valid economic evaluation.*
Model structure	*How was the model structure determined? Theory? Prior models? Expert opinion?*
Model time horizon	*Is the model time horizon suitable for the theory of the treatments and menorrhagia?*
Model cycle	*Does the model time cycle allow for costs and outcomes to be captured accurately?*
Key model assumptions	*Are they suitable? Determines validity of findings.*
Uncertainty analysis	*Has a comprehensive sensitivity analysis been carried out using one-way and probabilistic sensitivity analysis?. Were appropriate sampling distributions used?*
Funding source	*Was the study funded by the manufacturer of the pharmaceutical interventions or was it independent?*
Interventions	*Which pharmaceutical interventions have been assessed in the evaluation?*
Study size	*In primary data studies was the sample sufficient to provide robust results?*
Most cost-effective intervention	*Which intervention was shown to be most cost-effective?*

QALY; quality adjusted life year, WTP; willingness-to-pay

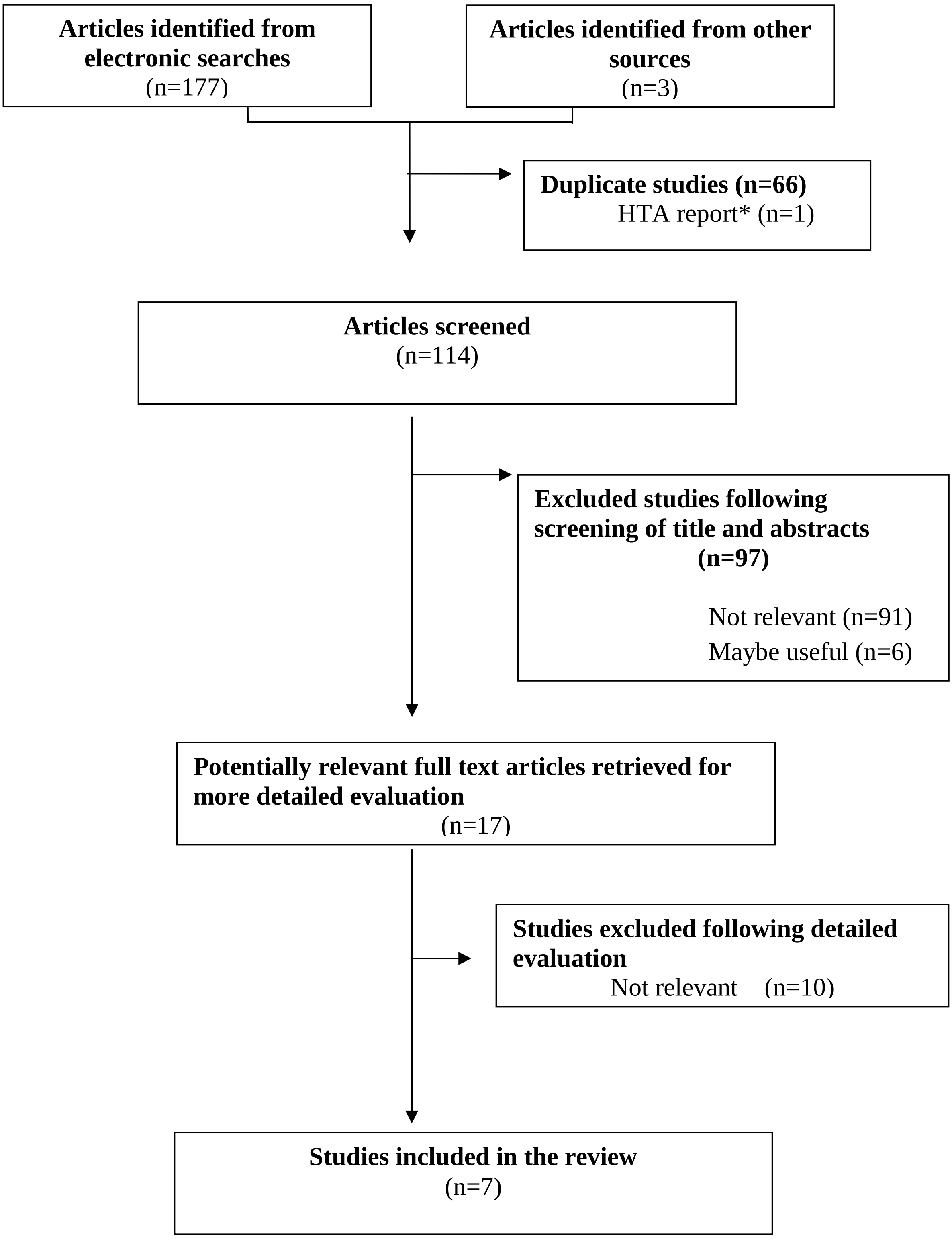
4.3 Results

The literature search identified 180 articles. 114 articles were screened and 17 of these were considered to be potentially relevant. Figure 4.1 illustrates the flow of papers from the initial electronic search to the number included in the review. Seven studies were assessed according to the quality criteria. Six of these studies were economic evaluations and used secondary data (Clegg et al., 2007; Ganz et al., 2013; Lete et al., 2011; NICE, 2007; Roberts et al., 2011; You et al., 2006). The one remaining study used primary data on costs and QoL but did not combine this information (Brown et al., 2006). All studies used decision-model based economic evaluations. The most robust study was carried out by Roberts et al (2011). This was an academic study that used secondary data and was carried out from a UK health service perspective. Roberts et al (2011) found hysterectomy to be the most cost-effective intervention, compared to endometrial ablation and LNG-IUS. They did not consider other non-surgical interventions. However, the utility values used in the model were combined from studies in different countries (Finland and the UK) and included values from EQ-5D and a direct TTO valuation method. These limitations will be discussed in section 4.4.

Of all the studies that reached the quality assessment stage of the review, it can be seen that only one study has used primary data to inform the model (Brown et al., 2006). However, the outcomes were not presented in terms of quality adjusted life years (QALYs) as the SF-36 values were not mapped onto SF-6D. Three studies were carried out in the UK, including the NICE guidelines (Clegg et al., 2007; NICE, 2007; Roberts et al., 2011) and all of the studies used a decision model. Two studies found hysterectomy to be most cost-effective (You et al., 2006; Roberts et al., 2011) whilst the remaining studies found LNG-IUS to be most cost-effective.

The seven studies that reached the quality assessment stage will be assessed in this section. Details of the seven studies can be found in Tables 4.2 and 4.3. The discussion of the articles will be divided into the following areas: model features, model inputs, study perspective, effectiveness measure, mean/median, sensitivity analysis and funding source. It is necessary to discuss the key issues that arose in each of the studies that reached the quality assessment stage of the systematic review, thereby ensuring these issues are not repeated in subsequent economic evaluations. Further details of the issues related to the quality of the studies are presented in Appendix 1.2. The more specific concerns with the NICE guidelines will then be discussed separately as the findings reported in the guidelines carry a considerable weight when informing clinical practice.

Figure 4.1 Flow of papers through study



*Both the HTA report and the publication relating to the same cost-effectiveness study were identified.

Table 4.2 Data extracted from relevant studies

Study	Year	Final Classification	Country	Currency	Data Source	Perspective	Effectiveness Measure	Mean/ Median	Model Type	Uncertainty Analysis	Funding Source
Clegg et al	2007	A1	UK	UK £	Secondary data	Healthcare provider	QALYs: EQ-5D & TTO	Mean & Median	Markov	One-way & PSA	Manufacturer of LNG-IUS
NICE	2007	A1	UK	UK £	Secondary data	Healthcare provider	QALYs: EQ-5D & RAND36	Mean	Markov	One-way	
Roberts et al	2011	A1	UK	UK £	Secondary data	Healthcare provider	QALYs: EQ-5D & TTO	Mean	Markov	One-way & PSA	Academic
You et al	2006	A1	Hong Kong	US $	Secondary data	Healthcare provider	QALYs: TTO (Sculpher)	Median	Markov	One-way & PSA	Academic
Brown et al	2006	A2	New Zealand	NZ $	Primary data	Direct & Indirect	SF-36	Mean	Decision tree	One-way	Academic
Ganz et al	2013	A2	US	US $	Secondary data	US payer perspective	QALYs: EQ-5D	Mean	Individual sampling	One-way & PSA	Manufacturer of LNG-IUS
Lete et al	2011	A2	Spain	ESP €	Secondary data	Healthcare provider	QALM: TTO (claimed EQ-5D); SFM	Mean	Markov	One-way & PSA	Manufacturer of LNG-IUS

LNG-IUS; levonorgestrel-releasing intrauterine system, PSA; probabilistic sensitivity analysis, QALY; quality adjusted life year, QALM; quality adjusted life month; SFM; symptom free months, TTO; time trade off

Table 4.3 Data extracted from relevant studies continued

Study	Interventions	Model cycle	Model time horizon	Study size	Most cost-effective
Clegg et al	LNG-IUS/ Hysterectomy/ EA	1 month	5 years	n/a	LNG-IUS
NICE	LNG-IUS/Surgery/ Oral	3 months	5 years	n/a	LNG-IUS
Roberts et al	Hysterectomy/ EA/ LNG-IUS	1 month	10 years	n/a	Hysterectomy
You et al	Hysterectomy/EA/LNG-IUS/ Oral	1 year	5 years	n/a	Hysterectomy
Brown et al	LNG-IUS Vs TBEA	n/a	2 years	70	LNG-IUS
Ganz et al	LNG-IUS/surgery/non-surgical	3 months	5 years	n/a	LNG-IUS
Lete et al	LNG-IUS/combined oral contraceptive/progestogens	6 months	5 years	n/a	LNG-IUS

EA; endometrial ablation, LNG-IUS; levonorgestrel-releasing intrauterine system, TBEA; thermal balloon endometrial ablation

4.3.1 Model Features

A Markov model was used in five of the studies identified, except Brown et al (2006) and Ganz et al (2013). Brown et al (2006) used a decision tree which is not entirely appropriate as the condition and the treatments under assessment are continuous. Hence LNG-IUS is not a cure for menorrhagia and if a decision tree is used the changes in QoL that occur and the associated recurring resource use is not captured. Therefore the treatments and the monthly condition cannot be accurately modelled by a decision tree as it does not allow for recurrent events and is far too simplistic. Ganz et al (2013) used the individual sampling model. The Markov model or, the more complex, individual sampling models are more appropriate for menorrhagia as they enable the recurrent events to be captured and can more accurately model the pathways.

There is a notable heterogeneity in time horizons used across the studies. The NICE guidelines and four of the other studies used a 5 year time horizon, whilst Brown et al (2006) adopt a 2 year time horizon and Roberts et al (2011), a 10 year time horizon. It is reasonable to expect that given the condition is chronic and continues until menopause, a longer time horizon would be most suitable. However there are advantages and disadvantages associated with each time horizon used. The 5 year time horizon commonly used seems to be reasonable given that the LNG-IUS is effective and licenced for 5 years. However as the condition continues until menopause and the LNG-IUS can be replaced and used for another 5 years, this time horizon may not capture all impact on resources. In contrast, the 2 year time horizon used by Brown et al (2006) is based on primary data and could therefore be considered to be more reliable but relatively short since future changes in resource use and QoL would not be captured. If this latter study was to be on-going and the analysis reported the first of many analyses to be carried out at various intervals, this interim analysis could be deemed acceptable. However, a subsequent evaluation is not mentioned in the paper. The evaluation by Brown et al (2006) could be considered to be more reliable in terms of data but the results are limited due to the short time horizon used. Roberts et al (2011) use a 10 year time horizon based on secondary data and use various assumptions to allow extrapolation of data to such a time horizon. The findings may prove to be more insightful when robust primary data are available because uncertainties associated with such broad assumptions need not arise. For example, the contraceptive needs and preferences for treatment are not likely to remain constant as they may change over such a long length of time. The length of the time horizon, then, of 10 years, is most suitable given the chronic nature of the condition. However, a clear trade-off, between the use, and availability, of more robust primary data in a short time horizon versus the less reliable, but more suitable, long-term analysis based on secondary data taken from several different sources is evident.

The time cycle used in the Markov model varies across all studies. Roberts et al (2011) and Clegg et al (2007) use a monthly time cycle. NICE (2007) and Ganz et al (2013) use a 3 month cycle. You et al (2006) use a yearly cycle and Lete et al (2011) use a 6 month cycle. The suitability of a time cycle of 3 months or more for the model is questionable. A cycle of 3 months or greater assumes that a patient does not change health states for the given time period. The NICE guidelines state that a 3 month cycle was chosen due to available evidence (NICE, 2007). It is also presumed that a 3 month cycle was chosen because treatment of both tranexamic acid and mefenamic acid should cease if no improvement is observed after 3 months of administration. However, if there is no improvement after 1 or 2 months, or if the treatment is associated with side effects, it is unlikely that a patient will continue to administer the treatment. This point is particularly pertinent in the case of LNG-IUS, as it is associated with numerous side effects within the first 6 months, such as inter-menstrual bleeding, post-coital bleeding or device expulsion. Such effects will more than likely result in changes in health states and treatment and would impact on QoL and resource use each month. By using a time cycle of more than a month, the monthly changes and impact on resource associated with a change of treatment, repeat prescriptions and follow-up appointments will not be captured resulting in an inaccurate economic evaluation. Hence, the 6 month and 1 year time cycles used by You et al (2006) and Lete et al (2011) will not capture important changes in QoL or effects on resource use leading to a poor economic evaluation. Whilst the studies using 3 month cycles will be slightly more accurate than these two studies with longer time cycles, they will still not be entirely accurate and capture all important changes (Ganz et al 2013; NICE 2007). The monthly time cycle used by Roberts et al (2011) and Clegg et al (2007), is most appropriate because menstruation occurs monthly and data on effectiveness and costs will be modelled according to the cyclical nature of the disease.

4.3.2 Model Inputs –Data Sources

Six of the seven studies were based on limited data as data on costs and utilities were taken from small trials in different countries. Furthermore, in some cases data from trials with different population characteristics were synthesised and in other cases data sources were unclear. As mentioned previously, Brown et al (2006) used primary data as they conducted a clinical trial within New Zealand but a sample size of 70 women was used. It is unlikely that the results of this study could be generalisable as it could be argued that the sample size used may not accurately reflect the population of menorrhagia sufferers. Further, three of the studies (Lete et al., 2011; Roberts et al., 2011; You et al., 2006) used utility values for the health states ‘well’ and ‘recurrent menorrhagia’ based on a study by Sculpher (1998). These utilities are based on patients who are referred to surgery. This point will be discussed in detail later but it is important to note that the QoL of these sufferers may be worse than a sufferer who had not been referred to surgery. Furthermore, some of these studies made assumptions to adjust the utility values from the secondary data to obtain values for the LNG-IUS, convalescence and other new health states that were more relevant for the newer treatments (Clegg et al., 2007; Roberts et al., 2011; You et al., 2006). Two studies estimated utility values from secondary data but provided a limited explanation of how the estimated values were derived, meaning that the results were not reproducible (Clegg et al., 2007; Lete et al., 2011). It is presumed that assumptions were made due to limited data on these alternative health states being available. Whilst fairly robust assumptions can be made on the process of care and associated resource use, it is unclear whether assumptions regarding the actual utility values for several health states of the model that have not been valued previously would be suitable. In some cases where assumptions were made, justifications for these values were not provided. It would be beneficial to elaborate on these methods to improve the transparency of the study.

Being a primary study, the study by Brown et al (2006) was the only one to use values from the country of interest. Four studies used utility values from a Finnish study, despite the country under analysis not being Finland (Clegg et al., 2007; Ganz et al., 2013; NICE, 2007; Roberts et al., 2011). Lete et al (2011) was a Spanish study and used UK values from Sculpher (1998) due to a value set not being available for Spain. There is evidence to suggest that different countries have different value sets which are not transferable, and therefore only values obtained from the country under analysis should be used where possible to ensure an appropriate evaluation is carried out (Knies et al., 2009). This issue will be discussed further in section 4.6.

Effectiveness Measure

A range of instruments have been used to assess the QoL of a woman for menorrhagia and a woman treated with menorrhagia. In some studies combinations of different measures have been used in the same analysis. The SF-36 instrument was used in the only study with primary data by Brown et al (2006). Sculpher (1998) was referenced for some utility values in the studies by NICE (2007), You et al (2006), Clegg et al (2007), Lete et al (2011) and Roberts et al (2011). Sculpher (1998) used the direct time trade-off (TTO) method to value QoL in menorrhagia sufferers. In addition, Clegg et al (2007), NICE (2007) and Roberts et al (2011) also used utility values from an earlier Finnish clinical trial by Hurskainen et al (2004) which used the EQ-5D instrument. There are concerns with using utility values from different valuation methods. Despite both Sculpher's values and EQ-5D being based on TTO, the way in which the values were elicited are entirely different. The TTO used by Sculpher (1998) uses patient values to directly elicit utility values for health states that are specific to menorrhagia, whilst the utility values in EQ-5D are valued indirectly by TTO using values from the general population for general health states. Therefore, the utility values obtained from the two methods are not comparable and should not be used together in the same analysis. Data then, from a variety of instruments should not be used, in the same way that data from different countries should not be employed in a study that only considers the effect on one country.

Mean/Median Values

Data on utilities or costs are commonly summarised as either the mean and median values. The arithmetic mean, or average, is defined as "the sum of all scores in a distribution divided by the total number of cases" (Argyrous, 2000 pp 126). The median is defined as the middle score of a set of data that is rank-ordered from lowest to highest (Argyrous, 2000).

In the articles identified in the systematic review, various summary measures have been used. Clegg et al (2007) have combined both mean and median utility values in the same analysis, that is from one source they use mean values and from another, median values. You et al (2006) and Lete et al (2011) both employ median values and the remaining studies, mean values. Roberts et al (2011) use mean values for their primary outcome and use median values in the sensitivity analysis to demonstrate the extent to which the results of an economic evaluation can change depending on the summary measure used. There appears to be some confusion about which summary measure should be used. It is therefore important to provide a discussion on why this issue occurs and which summary measure should be used.

It has been argued that the type of summary statistic used should be dependent on the distribution (or spread) of the data (Altman, 1990). In many cases datasets have a symmetrical distribution which is a bell shape, where the majority of scores lie in the centre of the distribution and fewer and fewer scores are seen as the scores move away from the centre. In symmetrical distributions, the mean and median values are identical. However, in certain cases, the distribution can be skewed, particularly in the case of resource use. The majority of people will lie within a certain range of resource use, but there will be some people that utilise many more resources causing the distribution to be skewed towards one end of the scale. In this case, the mean and median values would differ drastically. When data are right skewed, the median would be greater than the mean, and vice versa, as the mean is greatly influenced by outliers (extreme results). It has therefore been described as a "misleading notion of the average" (Argyrous, 2000 pp128). As the median is simply the middle value in the dataset and not the sum of all the values, it is not affected by the outliers. Many statisticians argue that the median should be used when the data are skewed as it is a more representative summary of the data. This issue has led to some confusion when conducting economic evaluations, as some health economists have used the median value when the distribution of data are skewed and others have used the mean. In the NICE guidelines, when determining the mean length of menstruation, NICE state that “the use of the mean is questionable, given the skewed distribution” and therefore the median is used to define the summary value (NICE, 2007 pp21). However, this statement is not related to utility values and it is possible that the opinion may differ when data from an economic evaluation are under consideration.

Numerous authors including Thompson and Barber (2000), Torrance (1986) and Roberts et al (2011) argue that the mean should be used to summarise any dataset that is used for an economic evaluation, regardless of the shape of the distribution. It is argued that each person’s utility value should be valued equally, giving equal weight to each person (Torrance, 1986). If every individual person’s utility was not given equal weight and the median was used as a summary measure, it could be argued that a representative estimate of the population’s utility value has not been obtained. Hence, it is argued that those people that belong to this population and have extremely high or extremely low utility values will not be taken into account. A recent article by Roberts et al (2011) highlighted that numerous studies that have conducted economic evaluations of surgical interventions for menorrhagia have used median values. They highlight that the articles do not justify the use of the median over the mean, and explain that the only possible justification could be due to inconsistencies when mean values are reported in the literature. For example, in the article by Sculpher (1998) ‘well post ablation’ had a lower <u>mean</u> value than ‘convalescence following ablation’, which is unexpected. However, the <u>median</u> value for ‘well post ablation’ was greater than ‘convalescence post ablation’. As mentioned previously, Roberts et al (2011) re-calculated their evaluation and used median utility values to demonstrate that the result of the most cost-effective intervention differs depending on the summary statistic employed. Thompson and Barber (2000) also argue that the arithmetic mean is the most “informative measure” as the other summary measures do not provide information on the overall cost or utility value that will be derived by treating all patients. Instead, the median value provides data on the ‘typical’ value for an individual.

Conversely, Dolan (2000) argue that neither the mean nor median is appropriate. They initially state that if the majority of data are found within a certain range of values and very few lie outside of this range, it would be much more intuitive to use the median value, as the mean will be ‘pulled’ up or down much more than is representative of the population. They then discuss the notable differences that using each summary statistic will have when determining resource allocation, stating that the less severe states would be negatively skewed and the more severe states positively skewed. Therefore, the benefits of moving between states, i.e. from more severe to less severe, will be less if the mean was used than if the median was used. It is then concluded that neither measure is an accurate representation of the population (Dolan, 2000).

Although it seems that neither of the measures, mean or median, are particularly appropriate for summarising datasets, it is clear that one set method must be agreed upon to improve the consistency of economic evaluations. A key argument for using the mean is that all individuals’ utility values should be treated equally. It would not be appropriate to ignore certain individual’s values, as the results would be unrepresentative of the population. Therefore, it is argued that the most appropriate method to use would be the mean to summarise the data, but the median value could also be presented.

4.3.3 Perspective of Economic Evaluation

Although the perspective of the economic evaluation is not used as criterion to pass or fail a study, it is worth discussing the perspective used in each study. All the studies identified except Brown et al (2006), considered direct costs borne by the healthcare provider. In the UK, the decision-makers, NICE, recommend that a healthcare provider perspective be used as they are primarily interested in trying to allocate healthcare resources within the healthcare budget. Therefore, only costs borne by the healthcare provider are recommended when the economic evaluation is conducted. Hence all UK studies used a healthcare provider perspective (Clegg et al., 2007; NICE, 2007; Roberts et al., 2011). However, as discussed in Chapter 3, there is also a debate around whether the societal perspective should be taken into consideration. The societal perspective includes impact on the healthcare provider and society through changes in productivity and out of pocket costs incurred by the patient, such as travel to appointments, amongst other things. The study by Brown et al (2006) carried out in New Zealand, incorporated societal costs such as lost income. As menorrhagia is known to significantly impact the QoL of sufferers on a continuous cyclical basis, it is likely that the condition would impact on productivity (work) every month. Also, some women will be expected to pay for their prescribed pharmaceutical medication which will add to the costs that should be taken into consideration from the societal perspective. Therefore, in this sense, Brown et al (2006) have conducted a more comprehensive economic evaluation than the other studies as they have considered both the healthcare provider and societal costs. However, the likelihood of studies carrying out the societal perspective will be dependent on the decision-makers recommendations.

4.3.4 Sensitivity Analysis

All studies conducted a one-way sensitivity analysis to assess the uncertainty of the findings. In a one-way sensitivity analysis, model parameters are altered individually and the effect of this change on the overall cost-effectiveness result is determined. To conduct a robust sensitivity analysis and provide a more accurate assessment of the uncertainty of the findings, it is necessary to conduct a probabilistic sensitivity analysis ('PSA') where model parameters are changed simultaneously. Five out of the seven studies conducted this additional PSA (Clegg et al., 2007; Ganz et al., 2013; Lete et al., 2011; Roberts et al., 2011; You et al., 2006). You et al (2006) conducted the PSA and specified the triangular distribution for the cost and utility parameters but did not justify the use of the triangular distribution over the traditional gamma distribution for costs and beta distribution for utility values. A PSA was not conducted in the NICE guidelines (2007). It is particularly important to assess the uncertainty associated with the results of the NICE guidelines for two reasons. First because poor data were available and, second, because these recommendations from the guidelines inform clinical practice, emphasising the importance of ensuring the results are robust.

4.3.5 Funding Source

Finally, although the source of funding was not used to pass or fail a study, funding sources for each of the studies must be mentioned. The manufacturer of LNG-IUS is the funding source for three out of the seven studies (Clegg et al., 2007; Ganz et al., 2013; Lete et al., 2011). LNG-IUS was shown to be the most cost-effective intervention in all three of these studies. In these cases there is a conflict of interest, which the authors acknowledge, as the funders of the study have a vested interest in the outcome of the study. An outcome of LNG-IUS being most cost-effective, as was shown in all of these studies, would lead to an increase in company profits. While the cost-effectiveness findings from the academic study conducted by Brown et al (2006) and the NICE guidelines (2007) concurred with these industry funded studies, by finding LNG-IUS to be most cost-effective, two of the other academic studies determined that hysterectomy was the most cost-effective (Roberts et al., 2011; You et al., 2006). The mixed results indicate the need for further investigation into the most cost-effective intervention for treating menorrhagia.

4.4 Discussion Related to NICE Guidelines

In this section a summary and critique of the NICE guidelines is reported. It is deemed necessary to lend a whole section to appraising the NICE guidelines because, as mentioned, recommendations that arise from these guidelines are used to inform clinical practice. Hence an assessment of the quality of the guidelines is particularly important. In addition to the issues mentioned previously, several issues specific to the NICE guidelines will be discussed in this section to highlight the importance of ensuring that the guidelines are as accurate as possible.

4.4.1 Analyses

In the NICE guidelines two of the three analyses conducted were not deemed to be entirely appropriate. When ‘all pharmaceutical treatment’ or ‘non-hormonal treatment’ is compared to watchful waiting, each intervention is compared to no treatment. For example, mefenamic acid is compared to no treatment, and tranexamic acid is compared to no treatment and so on. Indirect comparisons of the cost-effectiveness results of each intervention against no treatment are then made to determine the order of cost-effectiveness of each treatment. Rather than indirectly comparing the cost-effectiveness of each intervention, it would be more meaningful to obtain the decision model results from a direct comparison by comparing tranexamic acid against mefenamic acid. NICE stated that a direct comparison between treatment groups is not conducted, because a rank order of treatment did not previously exist. However, in order to overcome this obstacle, several different analyses comparing a range of different interventions could have been conducted.

Furthermore, a literature search was carried out by the NICE guideline team and they found that no economic evidence supporting or opposing the case for cost-effectiveness of the oral progestogen, norethisterone, and injected progestogen was available. However, neither norethisterone nor injected progestogens were considered in the decision model as possible interventions for treating menorrhagia and therefore no results on their cost-effectiveness were obtained. Despite this, NICE do recommend that oral progestogen and injected progestogen should be used as a third line treatment for menorrhagia. This recommendation has no economic merit and could lead to a situation where clinicians are more inclined, than previously, to prescribe these treatments even though their cost-effectiveness has not been determined.

4.4.2 ECLIPSE Committee Points

Many of the key limitations of the evaluation reported in the NICE guidelines were identified by the ECLIPSE trial team in their trial protocol, where the methods for assessing the effectiveness and cost-effectiveness of LNG-IUS against usual medical treatment in primary care is reported, as detailed in Chapter 1. It should be noted that the rationale for the ECLIPSE study was largely based on the limited findings of the NICE guidelines which highlighted the lack of direct data available on non-surgical interventions for menorrhagia. The objective of the ECLIPSE trial therefore was to provide the first opportunity to capture primary data on utilities and resource use on menorrhagia within the UK.

It is argued by the ECLIPSE team that the source of data used to populate the model reported in the NICE guidelines for the LNG-IUS intervention is particularly poor. One study conducted in Finland by Hurskainen et al (2004) with a small number of participants is heavily drawn upon to provide data for the model for LNG-IUS. It is recognised that at the time, this was the only study that provided information on cost-effectiveness for LNG-IUS in relation to menorrhagia, yet the small sample size and the country of the study limits the generalisability to the population of women with menorrhagia within the UK. As the perspective of the NICE guidelines is the UK, National Health Service (NHS), the use of data from a Finnish clinical trial to estimate utility values to populate the model would not be consistent with the UK perspective, as clinical practice and utilities derived from patients in Finland may not translate to UK clinical practice. There is emerging literature suggesting that the transferability of the EQ-5D instrument across countries is limited (Knies et al., 2009). This is because the EQ-5D and similarly SF-6D have national value sets, and results from international studies should be adjusted to the appropriate value set for the country. It was found that a difference between national value sets can be perceived in the preferences for the attributes and that these differences were due to methodological and cultural differences (Knies et al., 2009).

Secondly, as the study by Hurskainen et al (2004) is set in secondary care, rather than primary care, each of the women in the study had been referred by their general practitioner (GP) or clinic to a hospital. It must be questioned whether it would be appropriate to extrapolate this information into the primary care setting. One could argue that the very fact that these women have been referred to hospital is because the women believe that their condition is particularly severe and that it cannot be managed by pharmaceutical treatment alone, assuming it was offered beforehand. Furthermore, as the women recruited into the trial were on the waiting list to have a hysterectomy, the utility values can be considered to be biased towards surgery and the women themselves may have already have been resolved to the idea of a permanent solution, or cure. Consequently, using a value from this Finnish trial for the NICE model parameter ‘proportion of women who have surgical treatment following failed pharmaceutical treatment’ for LNG-IUS in the guidelines is inappropriate because these cases may be unrepresentative of all sufferers of menorrhagia, and would result in an overestimation of cost-effectiveness. However, it should be noted that NICE did recognise that using a utility value from Sculpher (1998), derived from women who had been referred to surgery, may overestimate the extent of the problem in primary care because these may be severe cases. But, the one-way sensitivity analysis undertaken by NICE to test this assumption showed that even if the utility decrement associated with menorrhagia is low, i.e. it is considered a serious problem for the woman, the relative results of the model were unchanged. However, a significant point highlighted by the ECLIPSE team is that only a one-way sensitivity analysis and threshold analysis was conducted. A thorough analysis of uncertainty using PSA was not carried out. Thus, even though NICE recognised that there were very little data available to inform the guidelines, a full analysis of the uncertainty of results was not attempted.

4.5 Discussion

This review has highlighted the lack of data available on the cost-effectiveness of alternative interventions for menorrhagia. The findings from the NICE guidelines demonstrate the need for a large pragmatic clinical trial, as the recommendations were based largely on one primary study, argued to be of poor quality, which was carried out in Finland. Overall, the NICE guidelines highlight the sheer lack of effectiveness and cost-effectiveness data within this area. This is especially emphasised by the value for treatment success rate for the combined oral contraceptive treatment, being estimated by “taking high and low estimates of the effectiveness value obtained for other therapies and using a triangular distribution" (NICE, 2007 pp 111). It seems that, from a decision-makers perspective, to provide truly useful and accurate clinical guidelines for menorrhagia, it would have been much more beneficial to emphasise the importance of conducting a large UK based randomised controlled trial which considered the range of interventions under consideration, rather than providing clinical guidelines based on poor data.

Six subsequent publications following the NICE guidelines were identified. Only one of these studies was a primary study but the costs and outcomes, in terms of QoL, were reported separately and not presented as a CUA (Brown et al., 2006). This systematic review also demonstrates that the available cost-effectiveness results conflict with each other as five out of the seven studies identified in the review found LNG-IUS to be the most cost-effective intervention for treating menorrhagia and two did not. However, three of these studies that found LNG-IUS to be cost-effective were funded by the manufacturers of LNG-IUS. The shortage of UK based studies is also revealed as only two studies, in addition to the NICE guidelines, were conducted in the UK and neither of these used primary data.

One study was considered to be more robust and more relevant for the UK, than the others (Roberts et al., 2011). This study was a UK based economic evaluation conducted using secondary data to populate a decision model. This study found hysterectomy to be the most cost-effective intervention, when compared to LNG-IUS and endometrial ablation. This finding contrasts with those reported in the NICE guidelines. LNG-IUS was deemed the most cost-effective intervention in the NICE guidelines and it was also recommended that endometrial ablation should be conducted prior to a hysterectomy as it was the least cost-effective treatment. The study by Roberts et al (2011) was much more robust than the NICE study because individual patient level data from numerous clinical trials were used in the meta-analysis when synthesising evidence. Hence much more data were incorporated than the

NICE study which increases the reliability of the findings. Once again, this comparison highlights the need for a large pragmatic clinical trial that considers all interventions for treating menorrhagia to be carried out.

4.5.1 Strengths and Limitations

The strength of this review is that a comprehensive update of the literature on all of the economic evaluations and cost studies, which report QoL, in menorrhagia is provided. Second, the quality of the identified studies, including the NICE guidelines, is assessed using a recommended checklist which, to current knowledge, has not been done before. A limitation of the search is the inability to obtain and review the search terms and search methodology used in the NICE guidelines. Although best efforts were made to acquire the search terms from the authors of the NICE report a response was not received. However, it is believed that the search terms used are appropriate as the disease-specific search terms are similar to other systematic reviews, that have not reviewed economic evaluations in menorrhagia, and the economic search terms are similar to those used by Roberts et al (2002).

4.5.2 Comparison with Other Studies

It was not deemed necessary, by the research team, to conduct a systematic review of the literature prior to 2006 because it is expected that the review as part of the NICE guidelines had comprehensively assessed all studies prior to this time point in order to compose the guidelines. Hence an update of the literature was required to determine whether any new articles of economic evidence within menorrhagia had emerged.

Further, as mentioned in the methods, two systematic reviews were published previously. One was carried out by Blumenthal et al (2011) and the other by Gemzell-Danielsson et al

(2013). Neither review appeared to be a comprehensive systematic review of the studies that have assessed cost-effectiveness in menorrhagia. Both reviews focussed on LNG-IUS. The review by Gemzell-Danielsson et al (2013) is a more general review of LNG-IUS for all requirements (i.e. contraception) and in a paragraph dedicated to menorrhagia, the authors simply report which intervention is the most cost-effective in previous studies, referring to Blumenthal et al (2011). Whilst the review by Blumenthal et al (2011) is more specific reporting on the cost-effectiveness, and effectiveness, of LNG-IUS related to menorrhagia, a critical appraisal of the studies using the recommended checklists for economic evaluations was not provided to assess the quality of the included studies. Additionally, several more relevant databases, such as NHS EED and the Social Science Index, were searched to ensure that all studies were identified. A further three additional studies, which reported on the cost-effectiveness of LNG-IUS for menorrhagia, were identified in the review reported in this chapter.

4.5.3 Further Research

The findings from the review reported in this chapter have shown that there is a need for a good quality economic evaluation using primary data to be conducted in the UK. This will ensure that robust information on the cost-effectiveness of alternative non-surgical interventions for menorrhagia is presented to decision-makers and that meaningful results can inform clinical practice. The need for more reliable evidence should be met when the CUA alongside the UK ECLIPSE randomised controlled trial, reported in Chapter 6, is published as the methods used address the issues identified here to provide a robust measure of cost-effectiveness of non-surgical interventions for menorrhagia.

4.6 Conclusion

In this chapter a systematic review of economic evaluations in menorrhagia is reported. It is found that there are mixed results on which treatment is most cost-effective and there is very little reliable data available to date on the cost-effectiveness and effectiveness, in terms of QoL, of pharmaceutical interventions for menorrhagia. Only one primary study, since the 2006 is identified but a small sample size is used which limits the generalisability of the findings. Further, this one primary study does not carry out a full CUA as the results from SF-36 are not mapped onto SF-6D to generate QALYs. All of the other studies identified use secondary data that are taken from a different country or combined different utility valuation methods in the same analysis. The findings reported in this chapter indicate a need for a robust randomised trial with a large sample size to determine the clinical and cost-effectiveness of LNG-IUS compared to pharmaceutical treatment. In the next chapter a systematic review on the use and psychometric assessment of economic outcome measures in menorrhagia is reported.

CHAPTER 5. AN ASSESSMENT OF ECONOMIC MEASURES USED IN MENORRHAGIA: A SYSTEMATIC REVIEW

5.1 Introduction

In this chapter, a systematic review of the economic measures (utility-based quality of life (QoL)) and contingent valuation) used in menorrhagia is presented. This systematic review has been accepted for publication (see Sanghera et al., 2013). The aim of such a systematic review is to assess the appropriateness and relative success of these measures in studies of menorrhagia. It is necessary to conduct this systematic review because (i) it has been recently formally recognised that QoL is the key indicator of treatment success in menorrhagia (NICE, 2007). Therefore, it is important to identify which QoL measures have been used and whether they are deemed to be adequate based on the literature. (ii) A recent study by Roberts et al (2011) has suggested that QoL or utility values available for menorrhagia are inadequate and need to be reassessed because the utility values associated with treatments were the main cause of uncertainty in an economic evaluation. Until the evidence on the utility values associated with menorrhagia is strengthened, robust recommendations of economic evidence cannot be provided to decision-makers.

To strengthen evidence, it is necessary to ensure that the valuation approach accurately reflects women's experiences and preferences. Consequently, this systematic review provides a narrative review of all economic measures (utility-based QoL and contingent valuation) that have been used in the valuation of outcomes associated with menorrhagia, and evaluates their psychometric properties and clinical utility. First, the methods for the literature search and study selection are described. Second, the results are presented according to the instrument under assessment and finally the chapter ends with conclusions on the use of QoL measures in menorrhagia and prospects for further work.

5.2 Methods

An initial scoping search was conducted in December 2011 to identify previous systematic reviews and the most appropriate search terms. Two relevant reviews were identified (Blumenthal et al., 2011; Clark et al., 2002). Details on the objectives of these previous reviews are discussed later in section 5.4.2. As neither of these reviews answered the question of the research reported here, it was necessary to conduct a full systematic review.

The guidelines by the Centre for Review and Dissemination and Cochrane Collaboration for reviews were followed using a narrative synthesis (CRD, 2009; Higgins & Green, 2009).

5.2.1 Eligibility Criteria

Since the objective of the review was to assess the appropriateness and relative success of the use of economic measures in menorrhagia, studies were included in the review if they met the following criteria:

- Either:
    - i. Use economic measures (such as EQ-5D, SF-36, SF-6D, disease-specific utility based measures, willingness-to-pay (WTP)) in relation to menorrhagia in a clinical trial or observational study or
    - ii. Discuss the development/ describe economic measures used in women with menorrhagia or assess economic measures used in menorrhagia.

The term, economic measures, refers to utility-based QoL instruments that have the potential to be used in an economic evaluation in order to report the results in terms of incremental quality adjusted life year (QALYs). These include EQ-5D, SF-36, TTO, SF-6D, and also any disease specific utility-based measures and visual analogue scale (VAS) in order to be comprehensive. Additionally, contingent valuation measures such as WTP which are used in cost-benefit analyses are also considered under the term 'economic measures'. The following should be noted:

- The population of interest was women with menorrhagia undergoing assessment of QoL
- Studies that focused on adolescents (less than 15 years old) were not included. If menorrhagia is experienced in adolescents it is likely that it will subside after puberty and the typical population of sufferers are predominantly women approaching menopause. (Duckitt & Shaw, 1998)
- Studies involving patients with adenomyosis and endometriosis were excluded because menorrhagia is typically secondary to another symptom
- Studies including women with bleeding disorders such as Von Willebrand disease were excluded because the relationship between the impact of this on QoL compared to the impact of menorrhagia alone on QoL has yet to be established and treatment for the condition differs (James et al., 2005).

In all other cases, studies were included where menorrhagia was the presenting complaint. Restrictions on study design were not applied as data from any study type would address the review question.

In January 2012 the following databases were electronically searched; Medline (1948 onwards), EMBASE (1947 onwards), Social Science Citation Index, Science Citation Index, PsychInfo (1967 onwards), CINAHL, NHS EED, DARE and HTA. Reference lists from eligible studies that were selected were also reviewed. The review was updated in August 2013.

Search terms were identified from search filters in the databases and previous systematic reviews (Blumenthal et al., 2011; Clark et al. 2002). The search terms included were heavy menstrual bleeding **or** menorrhagia **and** questionnaires **or** quality of life **or** outcome assessment **or** psychometry **or** psychological tests **or** psychometrics **or** interview **or** instrument (see Appendix 2.1). The search was limited to female and human studies. Index terms and free text were combined where possible. The results were managed using Reference Manager Software (version 12).

5.2.2 Study Selection

The review was conducted using a two stage process which followed the method of Mugford (1996) and has subsequently been established and described elsewhere (Bricker et al., 2000). The title and abstract of studies were screened according to the eligibility criteria and the full paper was obtained if appropriate. Where no abstract **and** no full text was available, the study was excluded. Where no abstract was available, the full paper was obtained. All studies were reviewed in both stages by this author. To validate this process, 20% of the titles and abstracts were independently reviewed by two other investigators (Tracy Roberts and Emma Frew). All of the investigators were blind to the categorisation decisions.

Stage 1 *– Initial Categorisation of studies*

The following criteria were used to assess the relevance of each study in the first stage of the review:

A) Primary research is reported and includes an economic measure (utility-based QoL or contingent valuation), which is used in an economic evaluation or cost-study

B) The study uses an economic measure in a clinical trial or observational study

C) The study may have useful information but does not obviously fall into (A) or (B)

D) The study has no relevance

Studies in category (A) and (B) were considered relevant to the systematic review, those studies in category (C) and (D) were not reviewed any further. 10% of the studies in category (C) were reviewed for relevance.

Stage 2 *– Further categorisation*

After reading the full article, the studies categorised in (A) and (B) were further categorised into the following categories:

1. An economic measure is assessed or used in an economic evaluation/ cost study

2a. Effectiveness study – An economic measure is used as a primary outcome

2b. Effectiveness study – An economic measure is used as a secondary outcome

3. The study discusses and/or describes QoL instruments

4. Review article

5. The study has no relevance

5.2.3 Data Extraction and Management

It was not deemed appropriate to use a strict quality criterion where articles can pass or fail since the objective was to determine which instruments have been used and consider how they have been used from a health economics viewpoint. Data were extracted and discussed based on the criteria outlined in Table 5.1. A narrative synthesis of data was taken, as a meta-analysis was not appropriate given the nature of the study question (CRD, 2009; Pirkis et al., 2005). Thus a discussion regarding the instrument's use is provided.

Included studies are broadly categorised into two groups; those that assessed the psychometric properties and the feasibility of economic outcome measures and those that used the measures. The full criteria used to judge the psychometric properties of the instruments are presented in Table 5.1. These criteria were considered to include all key psychometric properties and were based on a previously published study (Pirkis et al, 2005). Briefly, the instruments were assessed according to the *validity*, whether the instrument measures what it is designed to measure, *reliability*, defined as the ability to provide consistent scores and *sensitivity*, the extent to which the measure captures clinical changes. The extracted data were handled in Microsoft Excel (version 2007). The data extraction process was completed by this author. Two investigators (Tracy Roberts and Emma Frew) verified the data extraction process by extracting data from 10% of the above-classified studies. Disagreements for all stages were resolved by discussion.

Table 5.1 Criteria for data extraction

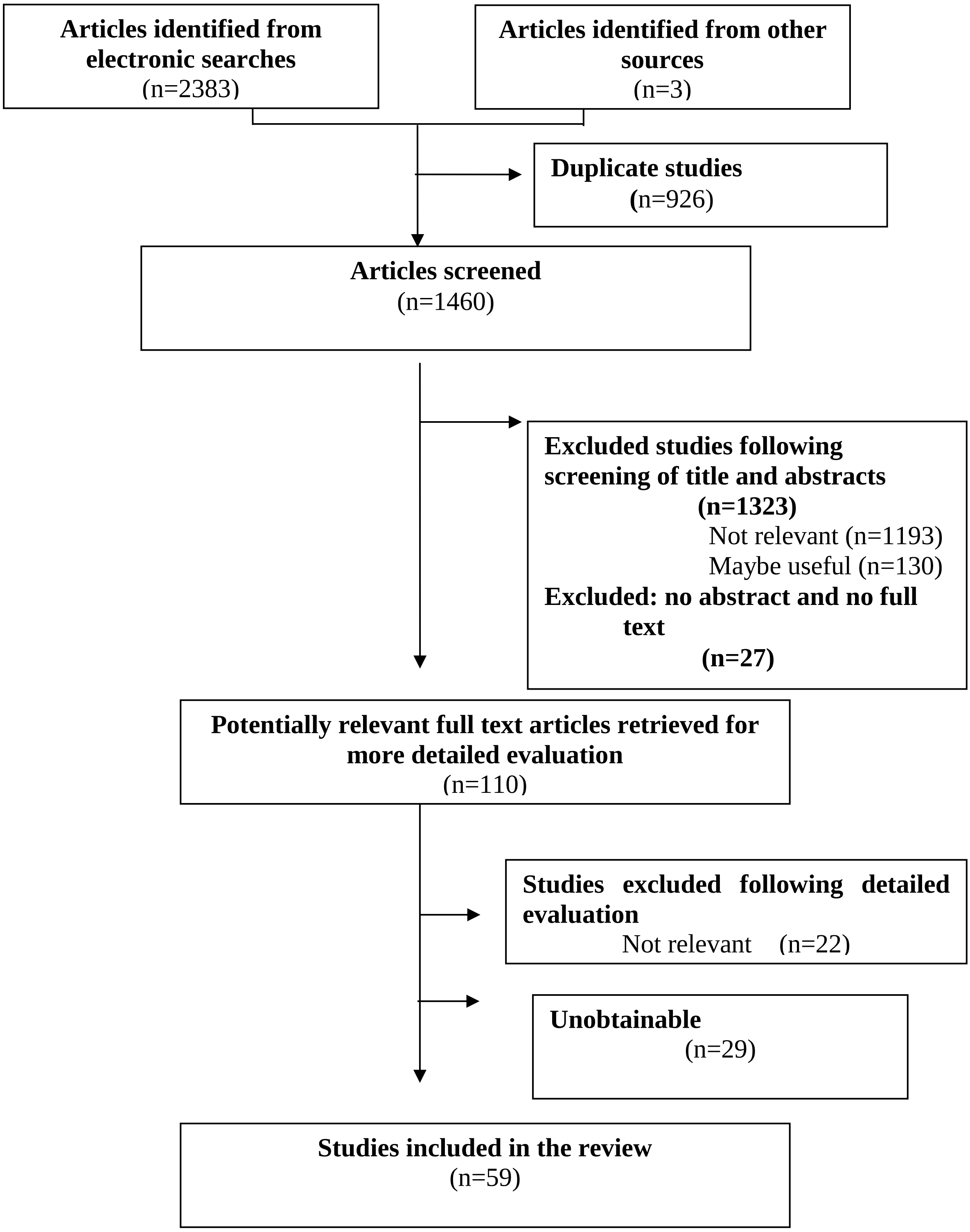
Criteria	*Justification*
Are details regarding the type of instrument(s) used provided? Is justification provided.	*Is the suitability of the instrument considered in detail? Or used just because it is typically used.*
Is the instrument used as secondary or primary outcome? Is justification provided for not using it as a primary outcome?	*It has long been recognised that QoL is most important and thus discussion about primary or secondary outcome would be valuable information.*
Validity	*Face validity: whether upon inspection the instrument measures what it claims to measure (Bannigan & Watson, 2009).* *Construct validity: determines whether an underlying relationship exists between questions in the instrument and an attribute that is measured (Streiner & Norman, 1995).* *Content validity: whether all relevant aspects of the condition are considered in the instrument.(Bannigan & Watson, 2009).*
Sensitivity/ responsiveness to change	*The extent to which the measure captures changes in comparison to the other measures used.*
Reliability	*Ability to provide consistent scores in the same respondent or across items in a test.*
Feasibility/ utility	*Having used the measures and assessed results the authors opinions on the instrument will provide valuable information on the utility. If the instrument is thought to be unacceptable then perhaps the measure should not be used.*
Is the completion rate stated or discussed? Are difficulties in completing the instrument discussed?	*This information will help to assess the appropriateness as a low completion rate may indicate difficulties with completion or highlight irrelevant information.*
Is a significant difference in QoL identified?	*If a significant difference in disease-specific measures is seen, is this reflected at all as a change in QoL?*
Are non-economic measures used? In terms of changes, is there a relationship between economic and non-economic measures?	*This will provide some insight into the suitability of economic measures. If several measures need to be used then it would suggest there is no consensus.*
Are health and non-health outcome measures used?	*Non-health outcomes are thought to be as the typically used health outcomes in this condition.*

QoL; quality of life

5.3 Results

A full break down of the number of studies identified and included in the review is provided in Figure 5.1. Fifty-nine papers were included in the review. Twelve papers assessed the psychometric properties of the outcome measures, twelve papers applied the measures in an economic evaluation or a cost study, eight and twenty-seven effectiveness studies used the measures as a primary and secondary outcome, respectively. The results are described according to the type of instrument assessed. Particular psychometric properties are not reported for all instruments if there were no relevant studies. Table 5.2 summarises the main findings and presents those studies that assessed psychometric properties and feasibility. Remaining studies are discussed in the text. (See Appendix A2.2, Table 1 for further details of the studies identified).

Figure 5.1 Flow of papers through study.



5.3.1 EQ-5D

Where EQ-5D was used as the single outcome measure, there were no studies reporting on the psychometric properties of the instrument. Some psychometric properties were reported when it was used alongside the menorrhagia multi-attribute utility scale (MMAS) instrument - the disease-specific measure which has been shown to be reliable and valid in menorrhagia (Pattison et al., 2011; Shaw et al., 1998). With respect to the construct validity, one case was identified where the overall score for EQ-5D had a poor correlation with MMAS (Pattison et al., 2011) which led to the authors suggesting that EQ-5D is unsuitable for patients with menorrhagia. A second study compared the sensitivity of EQ-5D and MMAS to changes in satisfaction post treatment, and found MMAS to be statistically associated with satisfaction whilst EQ-5D was not (Clark & Gupta, 2004). This result suggests that EQ-5D is not sensitive to important changes to women, such as their satisfaction with treatment outcome. However, in this particular study it is important to note that unconventionally, median QoL values were assessed and not mean QoL. The lack of sensitivity of the instrument to changes in QoL was also reported in another economic evaluation of treatment for menorrhagia (Kilonzo et al., 2010). With respect to rate of use of the instrument, EQ-5D was used in four of the economic evaluations identified and was the only economic measure used in two of the effectiveness studies (Dickersin et al., 2007; Frick et al., 2009; Hurskainen et al., 2001; Kennedy et al., 2003; Kilonzo et al., 2010; Sambrook et al., 2009b).

5.3.2 SF-36

Overall, more evidence was found on the psychometric properties of the SF-36 instrument compared to the EQ-5D. Three studies found SF-36 to be unreliable for individual decision-making (Jenkinson et al., 1996; Ruta et al., 1994) but reliable for group decision-making (Garratt et al., 1993; Ruta et al., 1994). In two of these studies (Garratt et al., 1993; Ruta et al., 1994), the patient group included not only patients with menorrhagia but also those with lower back pain, varicose veins and peptic ulcers.

The lack of face validity of SF-36 was identified in one study (Jenkinson et al., 1996) where, in interviews, women expressed difficulty completing the general health and mental health questions because the time frame is unspecified or inappropriate (refers to the past month). The authors argued that the ambiguity associated with these questions is due to the *cyclical episodic* nature of menorrhagia symptoms. This is evident in one participants response where she considered her health, which was 'fine', to be separate from her periods, which affected her greatly (Jenkinson et al., 1996).

The relevance of all of the SF-36 scales was confirmed by Garratt et al (1993) in a patient group consisting of four common conditions, including menorrhagia, lower back pain, varicose veins and peptic ulcers. Physical functioning was found to be the most relevant attribute (Garratt et al., 1993). However, when the subgroup results of the different conditions were assessed for menorrhagia, physical functioning was the only attribute that was not significantly different to the SF-36 scores of the general population, indicating that it may not be the most relevant attribute for the condition (Garratt et al., 1993). This finding was also observed in other studies identified in this review, as physical functioning is the least responsive domain in patients with menorrhagia and this finding has been observed in five studies (Cooper et al., 1997a; Cooper et al., 2005; Garratt et al., 1993; Hurskainen et al., 2004; Sambrook et al., 2009a). Indeed, Pattison et al (2011) found a lower correlation between the disease-specific MMAS and the physical functioning and general health subscales of the SF-36. The authors state that physical functioning is greatly influenced by mobility and self-care, which are not impacted by menorrhagia, and these findings suggest SF-36 may be unsuitable for measuring QoL in these women (Pattison et al., 2011). One Turkish study did show that all of the SF-36 domains in a population with menorrhagia were significantly different to a population without any specific condition, but the authors did adapt the SF-36 scale with very little explanation (Gokyildiz et al., 2013).

The SF-36 was shown to be sensitive to changes in the social impact score (Coulter et al., 1994), and the transition question which asks whether the condition has changed since the last point of measurement (Garratt et al., 1994). Discrepancies were seen between results, where SF-36 failed to reflect *large* changes that occurred in other questionnaires (Coulter et al., 1994). In Garratt et al (1994) when the response to the transition question was 'my health is somewhat worse', the SF-36 scores on the general health domain showed a moderate positive increase in general health rather than a negative change. One study also demonstrated that when compared to SF-36, MMAS is able to better predict the need for surgery (Habiba et al., 2010). This same study separated the Mental Composite Scale (MCS) and Physical Composite Scale (PCS) of the SF-36 to determine the impact of menorrhagia on specific aspects of life and concluded that the scoring procedure for these composite scales is inadequate and does not reflect changes seen on individual domains of the SF-36 (Habiba et al., 2010). The SF-36 MCS has been found to be more sensitive than the PCS as it was significantly associated with patient satisfaction, whilst the PCS was not (Hehenkamp et al., 2008).

Finally Habiba et al (2010) assessed the appropriateness of the use of SF-36 by examining score ranges, ceiling effects and standard deviations to identify whether the range of disability in the sample is similar to the range of disability covered in the measurement scale. At baseline, three domains, social functioning, role physical and role emotional, had a high ceiling effect indicating that the range of disability measured in the scale is less than the range of disability in the study sample, implying SF-36 had limited ability to distinguish between subjects.

SF-36 and SF-12 were the most commonly used outcome measures in effectiveness studies and the only outcome measure used in two of the identified economic evaluations but a lack of sensitivity in SF-36 is typically reported. Brown et al (2006) and Bongers et al (2005) feel that SF-36 fails to capture the concerns and experiences of patients. Brown et al (2006) observed small differences in SF-36 scores, but larger differences between treatments for women which fail and they therefore argue that the change in SF-36 does not correlate with failing treatment. The authors also highlight that sexual functioning is not addressed by SF-36 (an important factor identified in (de Souza et al., 2010)) and that the pain and discomfort attributes are too broad to detect changes in menorrhagia-related pain. Bongers et al (2005) assert that SF-36 is difficult to answer for patients with menorrhagia and Brown et al (2006) suggest that WTP should be investigated.

5.3.3 Willingness-to-Pay (WTP)

The WTP measure was evaluated for reliability in one study that assessed the consistency of responses from women with menorrhagia. WTP was elicited directly using a payment card in a questionnaire based format to compare hysterectomy against conservative treatment from an ex-ante perspective. Women's preference for treatments and WTP for treatments were elicited (Ryan & San Miguel, 2000) and in 30% (44/146 women) of cases, the greatest WTP and preferred treatment did not correspond, indicating that WTP lacks external reliability. The authors felt that the respondents were providing responses based on their knowledge of the cost of treatments and not their strength of preference for treatment (Ryan & San Miguel, 2000). The implications are argued to be particularly important in menorrhagia because treatment costs differ drastically thus the WTP values do not reflect maximum WTP, and WTP for a cheaper treatment would be underestimated (Ryan & San Miguel, 2000).

5.3.4 General Experience of Feasibility/Utility of Measures

Several authors (Bongers et al., 2005; Habiba et al., 2010; Kilonzo et al., 2010; Sculpher, 1998) commented on the difficulties in answering questions in *any* generic instrument, particularly questions on general health and mental health, and obtaining utility values in a chronic non-life threatening cyclical condition which has acute episodes of symptoms. Specifically, Sculpher (1998) explains that it is difficult to value health states of a chronic condition, where symptoms are episodic (Sculpher, 1998). As the EQ-5D instrument refers to health today, a high utility value may be assigned on the day of completion but a low utility thereafter. The generic QoL results are influenced by the timing of assessment and as it is difficult to identify the ideal timing, it may be difficult to elicit an accurate response from patients regardless of the generic instrument used (Sculpher, 1998).

Shaw et al (1998) and Jenkinson et al (1996) demonstrate that typically women do not consider menorrhagia to strictly be a health-related condition and they show that the condition could be considered separate to health. SF-36 and EQ-5D do not refer to non-health attributes which has resulted in their limited validity. Unlike SF-36 and EQ-5D the questions in the disease-specific MMAS mostly refer to the woman's state during her cycle, which is the first day of one period to the first day of the next period. This may overcome the problems of timing of assessment associated with SF-36 and EQ-5D (Pattison et al 2011). In the case of WTP, there is insufficient evidence available that has discussed the feasibility or utility of this measure in menorrhagia. Brown et al (2006) suggest WTP may be more suitable than extra-welfarist measures as it may be more sensitive to changes in utility following treatment and WTP captures more than general health related QoL. Only one study (Ryan & San Miguel, 2000) assessed WTP in menorrhagia.

5.3.5 Economic Measures in an Economic Evaluation/ Cost study

Twelve papers collected QoL data using economic measures in a study that either used primary data to conduct an economic evaluation or a cost study that had all of the relevant information to conduct an economic evaluation, but did not do so. Two papers were follow-up studies (Hurskainen et al., 2004; Sculpher et al., 1996). Time trade off (TTO), EQ-5D and 15D were the only measures (economic or non-economic) used alone in three studies (Kilonzo et al., 2010; Sculpher, 1998; Taipale et al., 2009). In addition to these studies, three studies that did use non-economic measures only used one economic measure, either SF-36 or the EQ-VAS (Brown et al., 2006; Sculpher, 1993; Van der Wilt et al., 2005). The remaining studies combined multiple economic measures as well as using non-utility based disease-specific measures. These non-utility based disease-specific measures include; menstrual loss, satisfaction, anxiety, Health and Depression Scale and effect on sexual activity. In seven of the twelve studies that used economic measures in an economic evaluation, justification for the measure used (EQ-5D, TTO, SF-36) is provided with the most common reason being due to the measure being validated and/or universally accepted. See Appendix A2.2, Table 2 for details of the studies identified.

5.3.6 Economic Measures in Effectiveness Studies

Eight papers collected QoL data to determine the effectiveness of an intervention using a utility-based instrument as a primary outcome. Twenty-seven studies used QoL instruments as secondary outcomes, and seven of these were follow-up articles. The details of these studies can be found in Appendix 2.2 (see Tables 3 and 4). The large number of studies using QoL instruments as a *secondary* outcome could indicate that the importance of using QoL in menorrhagia is either not yet accepted or is difficult to implement with the currently available instruments. Where QoL instruments are used as secondary measures, the primary measure is typically the degree of satisfaction and impact on sexual activity. Other than effect on sexual activity, non-health outcomes were considered in two studies (Dickersin et al., 2007; Malak & Shawki, 2006). These non-health outcomes included daily life, effect on employment, leisure activities and housework.

SF-36/ SF-12 was the most commonly used instrument (19/27 studies). It was also the most common economic measure to be used alone (13/27 studies). Several economic measures were used in six out of twenty-seven studies and in five of these studies three or more measures were used. Given that the economic measures were secondary outcomes, they were combined with non-utility based disease-specific measures. The use of multiple economic measures may reflect that the most appropriate measure to use is unclear and the use of many non-utility based measures may be due to the inability of the economic measures to completely capture patient's concerns. It was argued by one author that economic measures could complement disease-specific measures and the two should be combined to overcome their aforementioned problems (Bongers et al., 2005).

Table 5.2 Judgement on properties of instruments

Instrument	n	Validity			Reliability	Sensitivity	Feasibility/ utility
		Content	Face	Construct			
EQ-5D	3	None	None	Pattison et al 2011	None	(Clark & Gupta, 2004)	Kilonzo et al 2010
Judgement		*Insufficient evidence*	*Poor*[a]	*Poor*	*Insufficient evidence*	*Insufficient evidence*	*Mixed*
SF-36	9	None	Jenkinson et al 1996	Garratt et al 1993	Jenkinson et al 1996; Garratt et al 1993; Ruta et al 1994	Coulter et al 1994;Garratt et al 1994; Habiba et al 2010; Hehenkamp et al 2008	Brown et al 2006; Bongers et al 2005
Judgement		*Insufficient evidence*	*Poor*	*Mixed*	*Mixed*	*Mixed*	*Mixed*
WTP	1	None	None	None	Ryan et al 2000	None	Ryan et al 2000
Judgement		*Insufficient evidence*	*Insufficient evidence*	*N/A*	*Poor*	*Insufficient evidence*	*Insufficient evidence*

[a] Comments on the face validity of SF-36 can be applied to EQ-5D

5.4 Discussion

The findings suggest that there is no consensus on the most appropriate economic measure to use when valuing outcomes in menorrhagia. The fundamental problem with using the generic measures SF-36 and EQ-5D to value QoL in menorrhagia is poor face validity due to the cyclical nature of the condition. As these have a standard recall component (i.e how is your health today (EQ-5D); during the past 4 weeks (SF-36)), any results achieved from using them are critically affected by the timing of assessment. The evidence suggests that women do not consider menorrhagia to be solely a health-related condition, as practical difficulties related to carrying extra sanitary protection and clothes, in addition to impact on social life and daily routine are found to be important (Shaw et al., 1998). This begs the question whether health-related QoL measures are suitable for measuring outcomes associated with this condition. Furthermore, the findings from the economic evaluation suggest that researchers in this field do not feel they can rely on a single economic measure and have used several available instruments to allow an estimation of QALYs, in addition to disease-specific measures in order to strengthen their evidence.

The psychometric properties of EQ-5D and WTP in menorrhagia are under researched. Little evidence was found for the use of EQ-5D. Only one study assessed the reliability of WTP. Although the MMAS appears to be the most suitable measure due to the condition-specific nature, it cannot produce QALYs. One alternative would be to use mapping methodology whereby the outcome measured using the MMAS scale is mapped onto a 0-1 utility scale. However, this is not a perfect resolution as it requires a mapping algorithm to be generated based on the relationship between EQ-5D and MMAS. The limitations of mapping are described in section 3.3.2 under the subheading 'Mapping'.

5.4.1 Strengths and Limitations

Based on a systematic search of the available evidence, this is the first systematic review that has sought to identify and review 1) the economic instruments that have been assessed in the menorrhagia patient population, 2) the instruments that have been used in economic evaluations and 3) those that have been used in effectiveness studies. The search methodology was robust and in line with others in the field (Blumenthal et al., 2011; Clark et al., 2002). The review has identified a limited set of studies that have assessed the psychometric properties of economic outcome measures within menorrhagia.

5.4.2 Comparison with Other Studies

Two previous reviews have aimed to assess the quality of QoL instruments used in menorrhagia and to review economic and health related QoL outcomes data associated with Levonorgestrel-releasing intrauterine system (LNG-IUS) (Blumenthal et al., 2011; Clark et al., 2002). The present review differs to previous reviews as it assesses economic instruments used in *any* menorrhagia treatments and in studies of any design. Unlike Blumenthal et al (2011), who used only studies related to LNG-IUS, this review included any study using an economic measure in menorrhagia. The review reported here also incorporated both studies that used economic measures in clinical trials and observational studies to compare treatments, in addition to identifying evidence on the psychometric properties and feasibility of these measures in menorrhagia. Additionally, several more databases were searched to ensure all studies have been identified. In comparison to Clark et al (2002), the current review considers economic measures in terms of their appropriateness or use in economic evaluations, rather than assessing the performance of *any* non-utility based QoL instruments according to the psychometric measurement properties. Furthermore, the search terms used in the review by Clark et al (2002) were adapted in order to ensure all relevant articles were identified. In particular, the key additional search term included in the review reported in this chapter was ‘heavy menstrual bleeding’, as this term and menorrhagia are commonly and increasingly used interchangeably in clinical practice and in the literature. Hence the review reported in this chapter comprehensively assesses the relative success of any economic measure used in any study in menorrhagia, which has not been done before.

5.4.3 Further Research

It has been demonstrated that there is a case for combining QoL instruments in menorrhagia. However, the limitations of the instruments outlined in this review, which primarily relate to the cyclical episodic nature of the condition, may not be unique to menorrhagia. The findings may be relevant to other conditions where symptoms occur in episodes such as migraines, exacerbations of chronic obstructive pulmonary disease and asthma amongst others. Nevertheless, the implications of our findings are particularly important for menorrhagia because treatment is driven by the women's perception of the impact of the condition on their QoL.

Mixed results in support of using SF-36 and EQ-5D to value health states in menorrhagia have been identified. Both the validity and reliability of these instruments are affected by the cyclical nature of the condition. WTP considers a broader range of QoL and overcomes the issue of timing of assessment but there is currently insufficient evidence for its use. Therefore the use of this outcome measure should be further explored. Many studies have successfully elicited WTP values in other disease areas (Frew et al., 2001; Haefeli et al., 2008), so perhaps WTP could prove to be the most suitable economic outcome measure in menorrhagia.

5.5 Conclusion

In this chapter a systematic review of economic outcome measures used in menorrhagia is reported. It is found that there is no consensus on the most appropriate measure to use in menorrhagia. The currently used extra-welfarist measures are shown to be limited due to their focus on health related QoL and the standard recall periods used meaning that the results are affected by the timing of assessment. The welfarist WTP, which enables a broader assessment of wellbeing and may overcome the issue of timing of assessment, has not been thoroughly explored in menorrhagia. Therefore there is scope for assessing the feasibility of WTP in menorrhagia. In the next chapter, the economic evaluation alongside the ECLIPSE trial, comparing LNG-IUS and usual medical treatment is reported. A comparison of the cost-effectiveness results provided by the two extra-welfarist measures, EQ-5D and SF-6D, will be drawn to determine the extent to which the findings from these measures differ.

CHAPTER 6. ECONOMIC EVALUATION ALONGSIDE ECLIPSE TRIAL

6.1 Introduction

In this chapter the economic evaluation alongside the NIHR HTA funded ECLIPSE randomised controlled trial is reported. The objective of the ECLIPSE trial is to compare the clinical effectiveness and cost-effectiveness of the levonorgestrel-releasing intrauterine system (LNG-IUS) versus usual medical treatment. As discussed in Chapter 2, following the recommendations of the National Institute for Health and Care Excellence (NICE), that surgical intervention should be the *last* course of treatment for patients with menorrhagia, the clinical and cost-effectiveness of non-surgical interventions has become the primary focus of clinicians and decision-makers alike.

As discussed in Chapter 4, in 2007 NICE introduced guidelines (NICE, 2007) for the LNG-IUS to be used for treatment of menorrhagia based on limited evidence of cost-effectiveness (Stewart et al., 2001). Nine other small trials compared LNG-IUS to non-hormonal and hormonal treatments, showing reduction in menstrual blood loss but they did not consider cost-effectiveness (Endrikat et al., 2012; Shapley et al., 2002). The findings of the systematic review, in Chapter 4, on economic evaluations, further demonstrated that the existing economic evaluations have predominantly compared non-surgical interventions (*either* LNG-IUS *or* usual medical treatment) to surgery, and non-surgical interventions have rarely been directly compared.

The aim of this chapter, therefore, is to conduct an economic evaluation alongside the ECLIPSE trial to provide robust evidence on the cost-effectiveness of LNG-IUS compared to usual medical treatment. Further, as impact on quality of life (QoL) is the primary outcome measure in menorrhagia, the outcome was measured using a number of different instruments including EQ-5D, SF-6D and the menorrhagia multi-attribute scale (MMAS). While MMAS is disease-specific, EQ-5D and SF-6D are both instruments that represent the extra-welfarist approach to measure outcomes. The influence on the cost-effectiveness results of using these extra-welfarist measures is explored. As the ECLIPSE trial has been described in detail in Chapter 2, a brief overview of the trial is provided here as part of the methods. The remaining section of the methods describes the model, the derivation of transition probabilities and utility values using the trial data, and resource use. The analysis and results then follow. The results are separated into two parts; Part 1 refers to EQ-5D and Part 2 to SF-6D. Finally, the discussion and implications for further research are reported. The reporting of this economic evaluation follows CHEERS guidelines and the checklist is presented in Appendix 3.1.

6.2 Methods

A model-based economic evaluation in the form of a cost-utility analysis (CUA) based on an outcome of cost per quality adjusted life year (QALY) was carried out alongside the ECLIPSE trial (Gupta et al., 2013). The analysis took a UK National Health Service (NHS) perspective in a primary care setting and provides an assessment of the difference in costs and QALYs between interventions over a 24 month time horizon. As the economic evaluation was carried out alongside a trial, an intention to treat analysis was adopted. A second analysis will be carried out at the 5 year time point, but falls outside the remit of this thesis. A societal perspective to include private costs to women was considered but deemed not to be feasible given the resource constraints for data collection.

6.2.1 Participants and Trial Design

The ECLIPSE trial, which found LNG-IUS to be more effective than usual treatment, is reported in detail elsewhere (Gupta et al., 2013). Briefly, 571 women with menorrhagia from 63 UK centres were randomised between February 2005 and July 2009. Women between 25 and 50 years of age presenting to their general practitioner (GP) with menorrhagia, occurring over at least three consecutive cycles, provided written informed consent to participate. Women were randomised to having a LNG-IUS fitted, or usual medical treatment, chosen by the GP and the woman based on contraceptive needs or desire to avoid hormonal treatment.

Usual medical treatment options included mefenamic acid, tranexamic acid, norethisterone, a combined oestrogen/progestogen or progestogen only oral contraceptive pill (any formulation), or methoxyprogesterone acetate injection (NCCWCH, 2007; RCOG, 1998). The particular medical treatment was specified prior to randomisation. Treatment reviews by GPs were carried out at 6 weeks and 3 months. Subsequently, treatments could be changed or discontinued due to perceived lack of benefit, side effects, change in contraception requirements, referral for endometrial ablation or hysterectomy as per usual practice (NCCWCH, 2007; RCOG, 1998). Treatment changes reported by patients were confirmed with the GP.

The study was approved by the South West England multicentre research ethics committee and all relevant local ethics committees.

6.2.2 Model

A model-based analysis was used as a vehicle for the within trial economic analysis to comprehensively account for the changes in QoL that occurred whilst the women were taking these treatments. The model was completed alongside the trial.

The trial data showed that, due to its non-curative nature, women were changing their treatment to identify the best method for managing menorrhagia, and this process had an influence on their QoL. The most suitable method to capture changes in QoL, occurring throughout the trial, and provide a robust CUA, was to represent these experiences as health states in a decision model, which follows the process of management of menorrhagia used in the ECLIPSE trial. The analysis does not lend itself to a regression framework because patients change between different health states on a monthly basis and QoL was not measured at that frequency. Therefore, the optimal method is to attach a QoL value to the health states. As patient level QALYs are not available on a monthly basis, it would be inappropriate to infer them from QoL scores at the time they happen to be taken in the trial. Hence, a measure of change in utility from baseline and the endpoint of the trial, as in a typical trial-based analysis which does not use a decision model, would not accurately capture the health states that women had experienced throughout the time span of the trial. Furthermore, a trial-based analysis, without a decision model, would not capture the time spent in health states or the associated repetitive costs and resource use. In this case, a decision model based on trial data provides a more realistic explanation of the utility pathway, providing information that can be synthesised with other data and projected forward.

A state transition (Markov) model was developed using Microsoft Excel. A Markov model was used because the treatment of menorrhagia has a complex prognosis that cannot be captured in a decision tree, again in part due to the repeated events. First, each time cycle of the menorrhagia model requires recurring events, which cannot be modelled in decision trees. Secondly, the elapse of time, i.e. monthly cycle, would also need to be made explicit which is only possible in Markov models (Briggs et al., 2006). Finally, as there are several possible consequences, or pathways, from each health state a decision tree would become overly complicated and may lead to the exclusion of such data, which are likely to have resource implications and may affect cost-effectiveness recommendations.

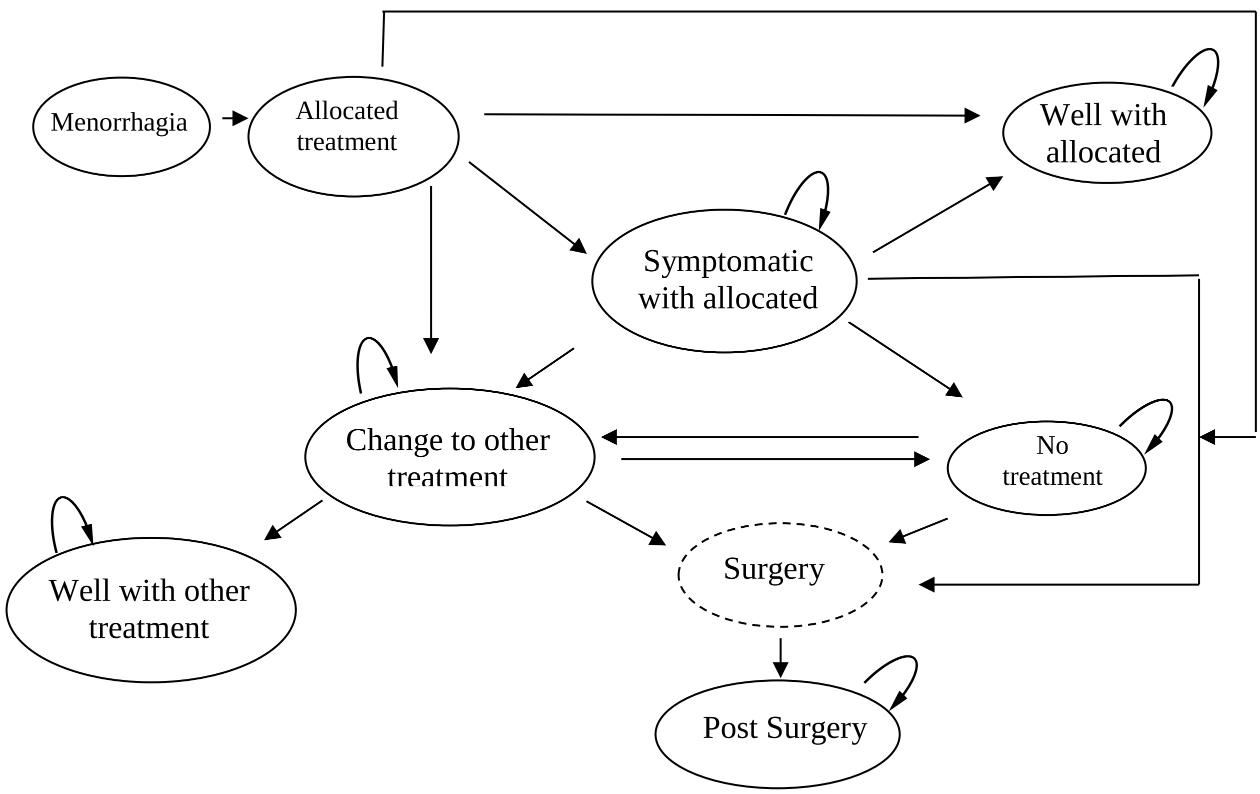
Model Structure

Existing model structures from published studies, which were identified in Chapter 4, were not deemed relevant to this analysis. The findings of the systematic review on economic evaluations in menorrhagia, presented in Chapter 4, revealed that previous decision models have predominantly compared surgical interventions to LNG-IUS (Brown et al., 2006; Clegg et al., 2007; Roberts et al., 2011). When women have surgical interventions, it is unlikely that they will cross-over to a non-surgical treatment, thus these previous model pathways do not allow for cross-over, which is required for this analysis, and were therefore not used to inform the model structure. Furthermore, the previous models are not sufficiently comprehensive to include all changes in treatments and health status that are relevant here. Since previous model structures were not used, the structure was developed using information from several sources which are outlined below.

To begin constructing the model, a map of the patient clinical pathway was drawn. The structure was informed by clinical input and by reviewing the ECLIPSE trial data to identify the pathway that women followed in the trial. Whilst the structure was not informed by literature, the health states for the model were developed using the pathways in the trial data and by adapting those used previously (NICE; 2007; Roberts et al., 2011). Several iterations of the structure were made as the model was continuously modified to ensure that it most closely reflected the trial data. The trial team, which included clinical experts, were extensively involved in the process of developing the model structure. Whilst the inclusion of menopause as a health state was considered, the 2 year time horizon of the analysis and the age of the patients, between 25-50, meant very few, if any, women would be reaching menopause within 2 years. For similar reasons, ‘death’ was not included as a health state. That is, the interventions do not cause death, and the age of the women meant that the likelihood of death due to natural causes was minimal. As models are a simplification of reality, it is only necessary to include transitions between health states that commonly occur for the majority of patients, and this was not the case for ‘death’ and ‘menopause’.

Figure 6.1 presents the clinical pathways and the progress of the two cohorts of women in the ECLIPSE trial who were randomised to either LNG-IUS or usual medical treatment. It can be seen that the pathway for each treatment can be represented in the same structure.

Figure 6.1 Clinical pathway for LNG-IUS and usual medical treatment



Definition of Health States for LNG-IUS Pathway

- **Menorrhagia**: Every woman in the ECLIPSE trial cohort suffers from excessive menstrual bleeding

- **LNG-IUS**: All women have LNG-IUS inserted

- **Well with LNG-IUS**: following the insertion of ‘LNG-IUS’ women are satisfied with treatment and remain in this health state

- **Symptomatic with LNG-IUS**: following the insertion of ‘LNG-IUS’, adverse effects may arise and women may feel unsettled with their current treatment. From this state women may remain in the state or change to an alternative treatment i.e. usual treatment, surgery, or no treatment

- **Surgery**: Women may choose to have surgery, which includes endometrial ablation or hysterectomy, which is determined from the data. Women can move to surgery from ‘symptomatic with LNG-IUS’, ‘change to usual medical treatment’, or ‘no treatment’

- **Post-Surgery**: following ‘surgery’, women will then remain in the ‘post-surgery’ state

- **Change to usual medical treatment**: if women have chosen ‘no treatment’ or are ‘symptomatic with LNG-IUS’ they may choose to begin usual medical treatment. Women may immediately become ‘well with usual medical treatment’ or may experience adverse effects and feel unsettled with usual medical treatment and then eventually become ‘well with usual medical treatment’. Alternatively, women may choose to have ‘surgery’ or ‘no treatment’. If women are in this state at 3 months the cost of a GP review will be incurred. Women can change to usual medical treatment from LNG-IUS in the first cycle. (See model assumptions for further details)

- **Well with usual medical treatment**: following ‘change to usual medical treatment’ women are satisfied with this treatment and remain in this health state

- **No treatment**: Following LNG-IUS, ‘symptomatic with LNG-IUS’ and ‘change to usual medical treatment’, women may choose to discontinue treatment altogether. Following this, women may decide to ‘change to usual medical treatment’ or have ‘Surgery’.

Definition of Health States for Usual Medical Treatment Pathway

- **Menorrhagia**: Every woman in the ECLIPSE trial cohort suffers from excessive menstrual bleeding

- **Usual medical treatment**: All women have usual medical treatment

- **Well with usual medical treatment**: following ‘usual medical treatment’ women are satisfied and remain in this health state

- **Symptomatic**: following usual medical treatment, adverse effects may arise and women may feel unsettled. From this state women may have an alternative treatment i.e. ‘change to LNG-IUS’,’ surgery’, or ‘no treatment’

- **Surgery**: Women may choose to have surgery, which includes endometrial ablation or hysterectomy, which is determined from the data. Women can move to surgery from ‘symptomatic with usual medical treatment’, ‘change to LNG-IUS’, or ‘no treatment’

- **Post-Surgery**: following ‘surgery’, women will then remain in the ‘post-surgery’ state

- **Change to LNG-IUS**: if women have chosen ‘no treatment’ or are ‘symptomatic with usual medical treatment’ they may choose to begin treatment with LNG-IUS. Women may immediately become ‘well with LNG-IUS’ or may experience adverse effects and feel unsettled with LNG-IUS and then eventually become ‘well with LNG-IUS’. Alternatively women may choose to have ‘surgery’ or ‘no treatment’. Women can change to LNG-IUS from usual medical treatment in the first cycle. (See model assumptions for further details)

- **Well with LNG-IUS**: following ‘change to LNG-IUS’ women are satisfied with treatment and remain in that health state

- **No treatment**: Following ‘usual medical treatment’, ‘symptomatic with usual medical treatment’ and ‘change to LNG-IUS’ women may choose to discontinue treatment altogether. Following this, women may decide to ‘change to LNG-IUS’ or have ‘Surgery’.

Time Cycle

A monthly time cycle was used in the analysis as this represented the clinically meaningful changes observed in treatment and resource use. The results of the systematic review in Chapter 4, on economic evaluations in menorrhagia, showed that there is no agreement on which time cycle should be used. A time cycle of either 1 year (You et al., 2006), 3 months (NICE, 2007; Ganz et al., 2013), 6 months (Lete et al., 2011) or 1 month was used previously (Clegg et al., 2007; Roberts et al., 2011). The NICE guidelines reported a 3 month time cycle, because it was constrained by available evidence (NICE, 2007). The ECLIPSE data showed that in many cases treatments had been altered and women had visited their GP within a month of treatment administration. Thus treatment changes were occurring more frequently than every 3 months. A monthly time cycle was therefore required to reflect the data and incorporate the associated costs and treatment changes into the CUA. A half-cycle correction was not used to allow for transitions to occur in the middle of the cycle (instead of the beginning or the end of the cycle) because the time cycle is small relative to the time horizon. On this basis a correction would not have made a substantial difference to the costs and outcomes (Briggs et al., 2006). The following assumptions were developed with clinical expertise from the ECLIPSE trial, which included a GP and a gynaecologist.

Model Assumptions

- A woman is ‘well’ with the allocated treatment if she does not change or stop treatment. Some of these women may not be ‘well’ but are coping with treatment, and the utility values for the ‘well’ with allocated treatment state reflect this
- A woman who is ‘well with LNG-IUS’ or ‘well with usual medical treatment’ cannot spontaneously become ‘symptomatic’
- Based on the data, if in the first cycle, women move from the allocated treatment to an alternative state other than ‘well’, it is assumed they either move to the ‘change to alternative treatment’ or ‘no treatment’ state. It is assumed that they do not move to the ‘symptomatic’ state in the first cycle because insufficient time has elapsed for this to be established and so it is assumed they changed for other reasons

- From the second cycle onwards, if women change from their allocated treatment they do not go to 'well' but to the 'symptomatic' state and move on from there

- For the transition to 'surgery', data were collected on whether a woman had ablation or hysterectomy, but not the precise technique (e.g. thermal balloon endometrial ablation or microwave endometrial ablation). Data on the weighted likelihood of surgery undertaken were taken from a previous study (Roberts et al., 2011). It is assumed that if a woman in the trial has endometrial ablation, then it will be for her first ablation and we apply the cost for first line endometrial ablation techniques

- Once a woman has changed from the allocated treatment, it is not possible for the woman to move back to the allocated treatment

- It is assumed that if a woman changes to the other treatment', she must spend at least one cycle in 'change to other treatment' before she can move to 'well with other treatment'. This is assumed as it will take at least one menstrual cycle for any effect to become apparent.

Calculating Transition Probabilities

To use the trial data in the model, it was necessary first to generate the 'Markov trace' by hand to identify the distribution of women in the states at any time and the transition probabilities between health states (Briggs et al., 2006).

First, for every patient, the time in months from randomisation to every event was calculated. For example, the number of months from randomisation to 'change to usual medical treatment' was calculated, and similarly for any other changes to states.

Using Microsoft EXCEL, two spreadsheets were constructed, one for LNG-IUS and the other for usual medical treatment. In each spreadsheet, all of the possible transitions from one state to another were outlined in the columns of a table and the rows of the table represented each patient. For every patient, the treatment allocated was identified and the time in months at which the patient moved from each state was entered into the appropriate column of the relevant spreadsheet, as shown in Appendix 3.2.

For each column that represents a change in state, the number of women who made the transition was summed according to the month at which the move was made. For example, all the women who stopped treatment in cycle 1 were summed, and those who stopped in cycle 2 were summed and so on. This was repeated for every cycle and every change in state.

To generate the Markov trace and obtain probabilities, the movements of all patients leaving and entering all of the states in each cycle of the model were tracked. The usual medical treatment arm is used to provide an example of how the Markov trace was conducted, but the method also applies to LNG-IUS.

In the model, it is assumed that patients who move to an alternative treatment ('surgery', 'no treatment', 'change to LNG-IUS') should not move to 'well with usual medical treatment' but should be moved to the 'symptomatic' state. Therefore the probability of moving from **'usual medical treatment' to 'well with usual medical treatment'** was the total number of women who made no moves divided by the total number of women who were randomised to usual medical treatment.

From **'usual medical treatment' to 'symptomatic with usual medical treatment'** the number of women who did not move to symptomatic was counted and deducted from the total number of women randomised to usual medical treatment. To obtain the transition probability, this number of women entering symptomatic was divided by the total number of women randomised to usual medical treatment.

In cases where women made moves that were not possible in the model, for example, returning to allocated treatment after changing to another state, these women were censored at the last point a move was possible. This change only occurred in 4 cases and is advised against clinically. Therefore this transition was not considered to be a typical pathway and does not represent the population. As the point in the cycle at which the women moved back to 'usual medical treatment' is unknown, the women were censored at one cycle prior to the return to randomised treatment.

From **'usual medical treatment' to 'change to LNG-IUS'** to be consistent and prevent biasing any data, the two patients that changed to LNG-IUS within the first month of treatment allocation were moved directly to this state without spending any time in symptomatic as reflected in the data. These women were identified by referring to the number of women who move in cycle 1. The total number who made the transition was divided by the total number of women who were randomised to usual medical treatment. The same method was used for movements from **'usual medical treatment' to 'no treatment'** in the first cycle. In the majority of the cases, the usual medical treatment was never taken and so considering these patients as symptomatic with usual medical treatment would not be accurate.

For the transition of **remaining in symptomatic,** it was necessary to calculate the total number of occasions that a woman starts in the symptomatic state to obtain the denominator for the probability of remaining in and leaving the symptomatic state. To calculate the number of occasions that a patient starts in the symptomatic state, the total number who moved into symptomatic in the first cycle was deducted from the number of women who move out of symptomatic i.e. to ‘no treatment’, ‘change to LNG-IUS’ or ‘surgery’, in cycle 1. This value was zero for the first cycle as all symptomatic women move into this state in the first cycle. For the next cycle, the number of women leaving symptomatic was deducted from the number of women remaining in symptomatic from cycle 1, as previously. This calculation was continued for every cycle. The number of occasions that women start in the symptomatic state was then summed for every cycle to give the total number of occasions someone starts in the state. The total number of times someone *remains* in the symptomatic state was then calculated by subtracting the total number of times a woman leaves the symptomatic state from the total number of occasions a woman starts in the symptomatic state. The probability of remaining in symptomatic was then this value divided by the total number of occasions someone starts in symptomatic.

The method is similar for **remaining in ‘change to LNG-IUS’** and **remaining in ‘no treatment’,** to that of ‘remaining in symptomatic’. The difference between these transitions and ‘remaining in symptomatic’ is that women enter ‘change to LNG-IUS’ and ‘no treatment’ from more than one state, and at any given cycle. Therefore when counting the number of women who start in the state and the number of women who leave in each cycle, the number of women entering from all states in each cycle must also be accounted for.

Transition probabilities for **‘symptomatic with usual medical treatment’ to ‘change to LNG-IUS’** or **‘symptomatic’ to ‘no treatment’** or ‘**symptomatic’ to ‘surgery’** were calculated by dividing the total number of women that leave symptomatic for the respective states, by the total number of occasions a woman starts in the symptomatic state.

For the explanation of the transition probability for **‘change to LNG-IUS’ to ‘well with LNG-IUS’,** one of the model assumptions set out above should again be highlighted. That is, women must spend at least one cycle in ‘change to LNG-IUS’ to then enter the ‘well with

LNG-IUS' state. When a patient moves to 'change to LNG-IUS' and would theoretically be classed as 'well with LNG-IUS' because she did not make any subsequent moves, she is placed into 'well with LNG-IUS' in the next cycle. Therefore if a woman 'changes to LNG-IUS' in cycle 4, she would be entered into the 'well with LNG-IUS' at cycle 5. To obtain the transition probability, the total number of women that move from 'change to LNG-IUS' to 'well with LNG-IUS' was summed and divided by the total number in 'change to LNG-IUS'.

The probability of moving from **'no treatment' to 'surgery'** or **'change to LNG-IUS' to 'surgery'** was calculated by dividing the total number of women who enter surgery, from each of the respective states, by the total number of women in the prior respective states. As women cannot stay in surgery, it is constructed as a transition state and therefore 100% of the women in **'surgery' move to 'post-surgery'** after one cycle.

Remaining in 'post-surgery' and **remaining in 'well with usual medical treatment'** are both absorbing states. This means women cannot leave the state once they have entered and therefore the probability of remaining in these states is 100%.

The calculation of the utility values assigned to each health state will be described in the next section following a description of the outcome measures collected in the trial.

6.2.3 Outcome Measures

Outcome measures were collected using both EuroQol-5D (EQ-5D) and Short-Form-36 (SF-36) at baseline prior to randomisation, then by post at 6 months, 1 year and 2 years post-randomisation. The booklet questionnaire, given to women in the trial, contained the generic EQ-5D-3L questionnaire which measures the impact of treatment on broader aspects of health related QoL (Brooks, 1996). SF-36 was converted into SF-6D using the algorithm by

Brazier et al (2002). [Provided by John Brazier's team at Sheffield University Personal Communication]. A complete case analysis is presented due to the nature of the analysis (further discussion is provided in section 6.5.1 of the discussion).

Calculating Utility Values

EQ-5D is reported as this is used for the base case; SF-6D is also reported, but in parentheses. The utility values for the health states were collated separately for each treatment arm. That is, the utilities for health states in the LNG-IUS treatment arm are taken only from women that were randomised to LNG-IUS. Utility values for the individual states were calculated by averaging the EQ-5D (SF-6D) values obtained by each woman in the given state at any given time. For example if a woman is randomised to LNG-IUS and then does not change treatment she is considered to be in the 'well' state for the remainder of the analysis, as outlined in the model assumptions previously. Therefore all of the woman's utility values collected at 6 months, 1 year and 2 years will be assigned to the 'well' health state in the model. Similarly if a woman is initially 'symptomatic with LNG-IUS' and then moves to 'no treatment' at 2 years, the utility values for 6 months and 1 year will be assigned to 'symptomatic' and the utilities for 2 years assigned to the 'no treatment' health state. This method was used to derive the utility values because the utility for the state is important, not the values associated with the individual woman's journey, as decision models are a reflection of the typical population.

The time from randomisation to the date of completion of EQ-5D (SF-6D) for each time point was also calculated and converted into the monthly time cycle. This was necessary to ensure that the correct EQ-5D (SF-6D) value was taken for the correct health state. The patient's state was then checked against the date when EQ-5D (SF-6D) was completed, and the EQ-5D (SF-6D) value was assigned correspondingly for all relevant time points. Where a patient was shown to have changed treatment on the same cycle that the EQ-5D (SF-6D) was completed, the exact EQ-5D (SF-6D) completion date and treatment change date was checked and dealt with accordingly. If the completion date and treatment change date were identical, the EQ-5D (SF-6D) value was assigned to the treatment change. This assumption will be tested in the sensitivity analysis, where the EQ-5D (SF-6D) value will belong to the state prior to the change.

Patients' EQ-5D (SF-6D) values were classed as 'symptomatic' up until a treatment change occurred. Where no change was made, the patient did not enter 'symptomatic', and the EQ-5D (SF-6D) values were all classed as 'Well'. Where the EQ-5D (SF-6D) was completed beyond 24 months for the 2 year follow-up, the date of completion and patient data were checked. If the patient had not made any subsequent moves between states and remained in the last known state the '2 year' EQ-5D (SF-6D) data were included. If a patient moved after 2 years and the EQ-5D (SF-6D) was completed prior to the move, but beyond 24 months, the EQ-5D (SF-6D) value was included. However, the EQ-5D (SF-6D) value would not be included if EQ-5D (SF-6D) was completed after the move, which occurred beyond 24 months. EQ-5D (SF-6D) data from patients who were censored from the follow-up, because they made a move in the model that was not possible, were excluded after the last point a move was possible.

For the utility values for patients who crossed over to the other arm of the trial and became well, the usual medical treatment arm will be used to provide the example explanation of how the utility values were derived. In the usual medical treatment arm, for the 'change to LNG-IUS' and 'well with LNG-IUS' states, the EQ-5D (SF-6D) value was classed as 'change to LNG-IUS' if EQ-5D (SF-6D) was completed in the same cycle as the 'change to LNG-IUS'. If the patient subsequently moved to 'well with LNG-IUS' in the next cycle, the next EQ-5D

(SF-6D) value and subsequent values were classed as ‘well with LNG-IUS’. Where a patient ‘changes to LNG-IUS’ and then did not move to ‘well with LNG-IUS’, but moved to another state, the EQ-5D (SF-6D) values in between the ‘change to LNG-IUS’ and the next move were classed as ‘change to LNG-IUS’. A similar method was used for changes from the LNG-IUS arm to the usual medical treatment arm.

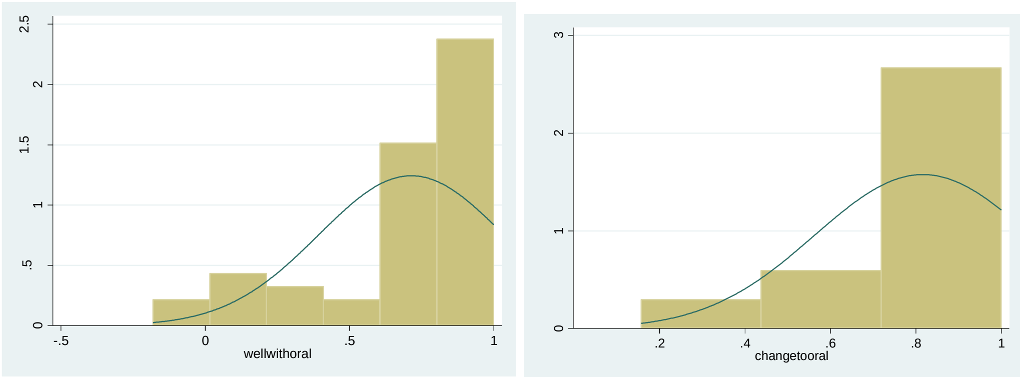
In the case of the utility values for surgery and post-surgery, as utility data were collected at 6 months, 1 year and 2 years the utility value obtained would be for any time point after surgery. The utility values given could not be considered the same as the utility value at the time of surgery because data on the exact date of surgery were not available. Therefore, the utility value collected after surgery was assigned to the ‘post-surgery’ state and the ‘surgery’ utility value was ¾ of the post-surgery utility. This was assigned because the patient was likely to be most severely affected by surgery during the first week of the month after surgery.

Significance tests for counterintuitive EQ-5D utility values

A series of significance tests were conducted to identify whether a statistically significant difference was observed between states where the utility values were counterintuitive.

In the LNG-IUS arm of the trial the EQ-5D utility values for the **‘change to usual medical treatment’** state and **‘well with usual medical treatment’** states were counterintuitive. This was because the utility value for ‘change to usual medical treatment’ was higher than the value for ‘well with usual medical treatment’. The histograms for the utility data for each health state illustrate that the data were found to be negatively skewed (Figure 6.2). To account for the non-normality, a Wilcoxon rank sum test was conducted using STATA (v11.0). The null hypothesis for the test was that there is no statistically significant difference in utility values between the states.

Figure 6.2 Histograms for EQ-5D utility data



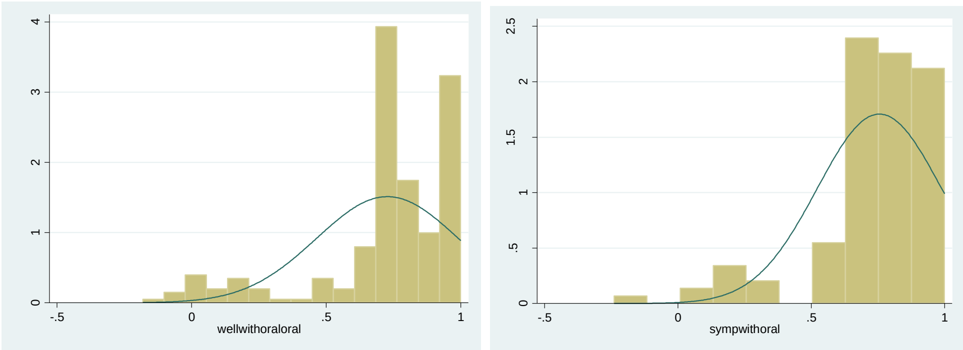

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

        group |      obs    rank sum    expected
--------------+---------------------------------
            0 |       12       407.5         360
            1 |       47      1362.5        1410
--------------+---------------------------------
     combined |       59        1770        1770

unadjusted variance     2820.00
adjustment for ties     -149.08
                       ----------
adjusted variance       2670.92

Ho: chang~ll(group==0) = chang~ll(group==1)
             z =   0.919
    Prob > |z| =   0.3580
```

0; change to usual medical treatment, 1; well with usual medical treatment

The p-value of 0.3580 is greater than 0.05 and therefore the null hypothesis cannot be rejected as there is no statistically significant difference between the utility values for ‘change to usual medical treatment’ and ‘well with usual medical treatment’ in the LNG-IUS arm.

In the usual medical treatment arm of the trial, the utility value for **‘well with usual medical treatment’** is unexpectedly lower than the utility value for **‘symptomatic with usual medical treatment’**. The utility data were found to be negatively skewed for both states (Figure 6.3) and a Wilcoxon rank sum test was conducted with a null hypothesis that there is no statistically significant difference in utility values across states.

Figure 6.3 Histograms for EQ-5D utility data (2)




```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

    grouporal |      obs    rank sum    expected
--------------+---------------------------------
            0 |      255       46880       47685
            1 |      118       22871       22066
--------------+---------------------------------
     combined |      373       69751       69751

unadjusted variance    937805.00
adjustment for ties    -23305.57
                      ----------
adjusted variance      914499.43

Ho: sympwell(groupo~l==0) = sympwell(groupo~l==1)
             z =  -0.842
    Prob > |z| =   0.3999
```

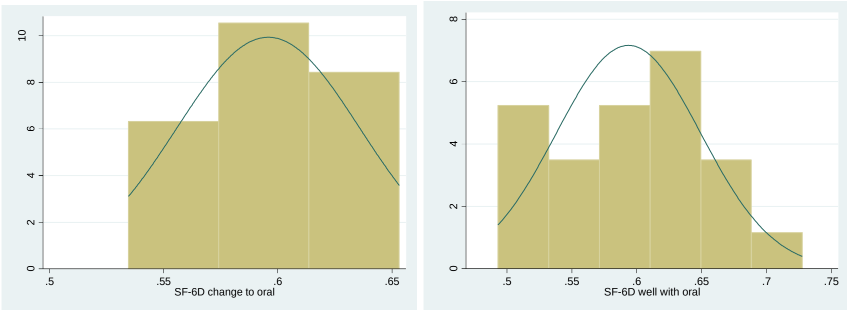
0; well with usual medical treatment, 1; symptomatic with usual medical treatment

From the test it can be seen that the p-value is 0.3999 which is greater than 0.05. Therefore the null hypothesis cannot be rejected as there is no statistically significant difference between the utility values for ‘well with usual medical treatment’ and ‘symptomatic with usual medical treatment’ in the usual medical treatment arm.

<u>Significance tests for counterintuitive SF-6D utility values</u>

In the LNG-IUS arm of the trial, when SF-6D is used, the utility value for **‘change to usual medical treatment**’ is greater than the utility value for **‘well with usual medical treatment’**. The utility data for ‘change to usual medical treatment’ were found to be slightly negatively skewed (Figure 6.4). So, a Wilcoxon rank sum test was conducted with a null hypothesis that there is no statistically significant difference in utility values across states.

Figure 6.4 Histograms for SF-6D utility data



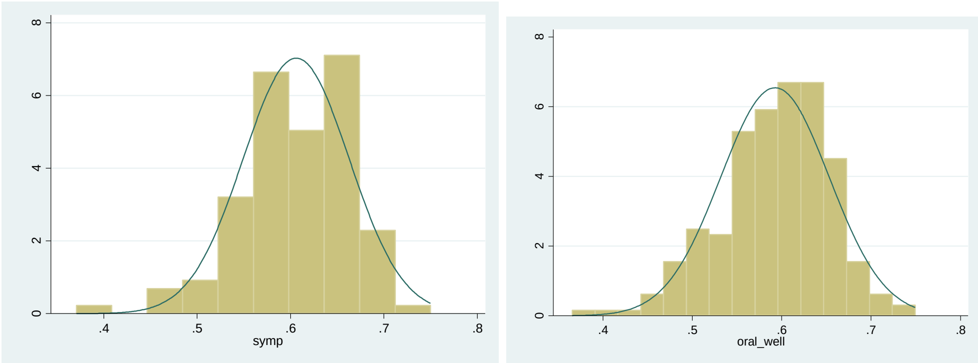

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

grpchangeo~a |      obs    rank sum    expected
-------------+---------------------------------
           0 |       12       343.5         342
           1 |       44      1252.5        1254
-------------+---------------------------------
    combined |       56        1596        1596

unadjusted variance     2508.00
adjustment for ties       -2.31
                     ----------
adjusted variance       2505.69

Ho: chang~al(grpcha~a==0) = chang~al(grpcha~a==1)
             z =   0.030
    Prob > |z| =   0.9761
```

0; change to usual medical treatment, 1; well with usual medical treatment

The p-value is 0.9761 which is greater than 0.05 therefore the null hypothesis cannot be rejected and it can be said that the difference between the states is not statistically significant.

In the usual medical treatment arm of the trial, the utility value for **'symptomatic with usual medical treatment'** is greater than the utility value for **'well with usual medical treatment'.** The utility data for 'symptomatic' were found to be slightly negatively skewed (Figure 6.5), so a Wilcoxon rank sum test was conducted with a null hypothesis that there is no statistically significant difference in utility values across states.

Figure 6.5 Histograms for SF-6D utility data (2)



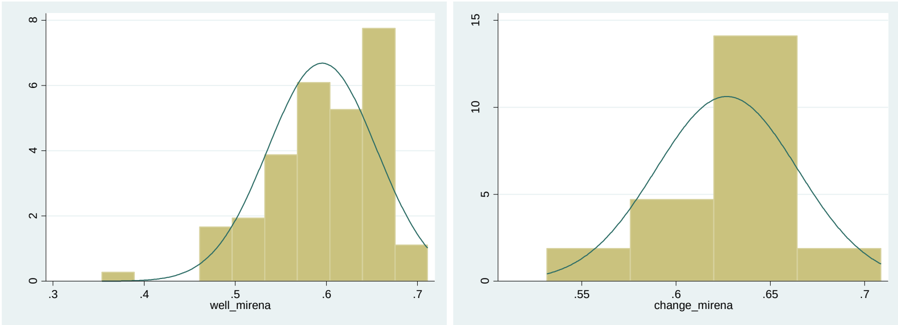

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

groupsympw~l |      obs    rank sum    expected
-------------+---------------------------------
           0 |      251     44185.5     46058.5
           1 |      115     22975.5     21102.5
-------------+---------------------------------
    combined |      366       67161       67161

unadjusted variance    882787.92
adjustment for ties       -60.28
                      ----------
adjusted variance      882727.63

Ho: sympwell(groups~l==0) = sympwell(groups~l==1)
             z =  -1.994
    Prob > |z| =   0.0462
```

0; well with usual medical treatment, 1; symptomatic with usual medical treatment

As the p-value of 0.0462 is very close to the arbitrary value of 0.05, it could be argued that it is likely that this statistical significance has occurred by chance due to the number of statistical significance tests conducted.

Also in the usual medical treatment arm, the utility value for **'change to LNG-IUS'** is greater than the utility value for **'well with LNG-IUS'.** The utility data for 'well with LNG-IUS' were found to be negatively skewed (Figure 6.6), so a Wilcoxon rank sum test was conducted with a null hypothesis that there is no statistically significant difference in utility values across states.

Figure 6.6 Histograms for SF-6D utility data (3)




```
. ranksum  changemirenawell, by ( group)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

       group |      obs    rank sum    expected
-------------+---------------------------------
           0 |      101      6002.5        6363
           1 |       24      1872.5        1512
-------------+---------------------------------
    combined |      125        7875        7875

unadjusted variance    25452.00
adjustment for ties      -20.88
                      ----------
adjusted variance      25431.12

Ho: change~l(group==0) = change~l(group==1)
             z =  -2.261
    Prob > |z| =   0.0238
```

0; well with LNG-IUS, 1; change to LNG-IUS

The p-value is 0.0238 which suggests that there is a statistically significant difference between the two states. It should be noted that this difference has arisen when SF-6D utility values are used and not EQ-5D. Perhaps this difference reflects that the measures are detecting different aspects of QoL and LNG-IUS may be impacting on areas that are captured by EQ-5D. Alternatively, the utility values may reflect that women were hopeful in the 'change to LNG-IUS' state that, after having tried usual medical treatment, the effect of LNG-IUS would be greater than it was.

6.2.4 Costs and Resource Use

Costs were collected from a UK NHS perspective. Data on healthcare resource use, including GP or gynaecologist consultations, were collected from women alongside other outcome measures. Similar to the utility values, data on costs and resource use from the trial were collated and analysed according to the randomised treatment arm. The general healthcare costs for both groups included healthcare staff costs and the cost of the interventions. An LNG-IUS fitting was estimated to take 20 minutes (informed by clinical experts within trial team), require both a GP and nurse to be present and also require disposable consumables. Treatment review by the GP was assumed to last 10 minutes (informed by clinical experts within trial team). Staff costs were calculated using nationally recognised reference costs (Curtis, 2011). The costs of standard medical treatment and LNG-IUS were estimated from the British National Formulary (BNF, 2011). Cost data on surgical interventions were taken from a previously published study and inflated to 2011 prices (Curtis, 2011; Roberts et al., 2011). As recommended by NICE a discount rate of 3.5% was applied to both costs and utilities as the model time horizon is beyond 1 year (NICE, 2008). All costs are reported in 2011 prices in UK (£) sterling using the UK hospital and community health services index (Curtis, 2011). Tables 6.1-6.3 present the data used in the analysis.

In cases where women were prescribed a combination of usual medical treatments, a weighted average of the cost was taken. When a patient was randomised to usual medical treatment, precise data on the type of usual medical treatment administered were not available. As these data were available in the change to usual medical treatment state in the LNG-IUS arm, the weighted average cost of the usual medical treatment in the LNG-IUS arm of the trial was used as a proxy for the usual medical treatment arm. Similarly, repeat prescription costs were calculated based on the average weighted cost of repeat prescriptions in the ‘change to usual medical treatment’ state of the LNG-IUS arm. As the most commonly prescribed usual medical treatments involved GP review for effectiveness at 3 months, it was assumed that GP review occurs at 3 months.

Where one-off costs were incurred and were not repetitive monthly costs, such as the cost of LNG-IUS removal, these costs were assigned to the transition arrow rather than the state as is typically done when costs are not repetitive.

Table 6.1 Cost data used in the analysis

	Unit cost	Source
LNG-IUS		
Consultation (GP 10 mins)	£26.67	Curtis 2011/ expert opinion
Insertion		
GP (20 mins)	£53.33	Curtis 2011/ expert opinion
Practice nurse (20 mins)	£17.00	Curtis 2011/ expert opinion
Device cost	£88.00	BNF 62
Sterile pack (insertion)	£21.63	NICE (inflated to 2011)
Discontinuation		
GP (10 mins)	£26.67	Curtis 2011/ expert opinion
Practice nurse (10mins)	£8.50	Curtis 2011/ expert opinion
Sterile pack (removal)	£3.77	NICE (Inflated to 2011)
Follow-up		
6 week review: (GP 10 mins)	£26.67	Curtis 2011/ expert opinion
3 month: (GP 10 mins)	£26.67	Curtis 2011/ expert opinion
Usual medical treatment		
Progestogen (Cerazette)	£8.68	BNF 62
Tranexamic acid (cyclokapron)	£14.30	BNF 62
Mefenamic acid (Ponstan)	£15.72	BNF 62
Norethisterone	£2.18	BNF 62
Combined oral contraceptive (microgynon)	£2.82	BNF 62
Methoxyprogesterone acetate injections (Depo-provera)	£6.01	BNF 62
Consultation: (GP 10 mins)	£26.67	Curtis 2011/ expert opinion
3 or 6 month review (GP 10 mins)	£26.67	Curtis 2011/ expert opinion
Discontinuation (GP 10 mins)	£26.67	Curtis 2011/ expert opinion
Surgery	£1720.18	Weighted cost from trial

All costs are presented in 2011 (£) sterling

Table 6.2 Health state utility data used in the model

Health State	EQ-5D value	PSA Distribution (EQ-5D)	SF-6D Value[#]	PSA Distribution (SF-6D)[#]	Source
LNG-IUS					
LNG-IUS	0.756	Beta (653, 211)	0.597	Beta (10204, 6883)	ECLIPSE trial
Well with LNG-IUS	0.98	Beta (1169, 297)	0.598	Beta (17912, 12061)	ECLIPSE trial
Symptomatic with LNG-IUS	0.744	Beta (130, 45)	0.589	Beta (3464, 2418)	ECLIPSE trial
Change to usual medical treatment	0.817	Beta (20, 5)	0.596	Beta (1066, 723)	ECLIPSE trial
Well with usual medical treatment	0.714	Beta (66, 26)	0.594	Beta (2032, 1390)	ECLIPSE trial
No treatment	0.785	Beta (70, 19)	0.604	Beta (2108, 1380)	ECLIPSE trial
Surgery	0.620	Linked to post surgery	0.430	Linked to post surgery	ECLIPSE trial
Post-surgery	0.827	Beta (59, 12)	0.574	Beta (330, 245)	ECLIPSE trial
Usual medical treatment					
Usual medical treatment	0.714	Beta (514, 206)	0.603	Beta (9892, 6519)	ECLIPSE trial
Well with usual medical treatment	0.728	Beta (528, 197)	0.592	Beta (9664, 6647)	ECLIPSE trial
Symptomatic with usual medical treatment	0.756	Beta (311, 100)	0.606	Beta (5168, 3359)	ECLIPSE trial
Change to LNG-IUS	0.694	Beta (49, 21)	0.627	Beta (2494, 1484)	ECLIPSE trial
Well with LNG-IUS	0.801	Beta (282, 70)	0.595	Beta (4069, 2766)	ECLIPSE trial
No treatment	0.766	Beta (223, 68)	0.586	Beta (3548, 2509)	ECLIPSE trial
Surgery	0.619	Linked to post-surgery	0.454	Linked to post surgery	ECLIPSE trial
Post-surgery	0.825	Beta (64, 14)	0.606	Beta (2136, 1391)	ECLIPSE trial

Utility values are rounded to 3 decimal places. α and β values for the PSA distribution are rounded to the nearest whole number. LNG-IUS; levonorgestrel-releasing intrauterine system, PSA; probabilistic sensitivity analysis.

[#]Values used in sensitivity analysis 4.

Table 6.3 Probability parameters used in the analysis

Probability Parameters	Probability	PSA distribution
LNG-IUS		
LNG-IUS to well with LNG-IUS	0.639	(182, 103)
LNG-IUS to symptomatic with LNG-IUS	0.253	(72, 213)
LNG-IUS to change to usual medical treatment	0.067	(19, 266) Dirichlet
LNG-IUS to no treatment	0.042	(12, 73)
Remain Well with LNG-IUS	1	Fixed
Symptomatic with LNG-IUS to well with LNG-IUS	0	Fixed
Remain symptomatic with LNG-IUS	0.907	(700, 72)
Symptomatic with LNG-IUS to change to usual medical treatment	0.035	(27, 745)
Symptomatic with LNG-IUS to no treatment	0.041	(32, 740) Dirichlet
Symptomatic with LNG-IUS to surgery	0.017	(13, 759)
Remain change to usual medical treatment	0.708	(109, 45)
Change to usual medical treatment to well with usual medical treatment	0.208	(32, 122)
Change to usual medical treatment to no treatment	0.045	(7, 147) Dirichlet
Change to usual medical treatment to surgery	0.039	(6, 148)
Remain well with usual medical treatment	1	Fixed
No treatment to change to usual medical treatment	0	(1, 547)
Remain no treatment	0.984	(540, 8) Dirichlet
No treatment to surgery	0.016	(10, 538)
Surgery to post surgery	1	Fixed
Remain post surgery	1	Fixed
Usual medical treatment		
Usual medical treatment to well with usual medical treatment	0.402	(115, 171)
Usual medical treatment to symptomatic with usual medical treatment	0.566	(162, 124)
Usual medical treatment to change to LNG-IUS	0.007	(2, 284) Dirichlet
Usual medical treatment to no treatment	0.024	(7, 279)
Remain Well with usual medical treatment	1	Fixed
Symptomatic with usual medical treatment to well with usual medical treatment	0	Fixed
Remain symptomatic with usual medical treatment	0.901	(1474, 162)
Symptomatic with usual medical treatment to change to LNG-IUS	0.049	(80, 1556)
Symptomatic to no treatment	0.040	(65, 1571)Dirichlet
Symptomatic to surgery	0.010	(17, 1619)
Remain change to LNG-IUS	0.603	(120, 79)
Change to LNG-IUS to well with LNG-IUS	0.312	(62, 137)
Change to LNG-IUS to no treatment	0.045	(9, 190) Dirichlet
Change to LNG-IUS to surgery	0.040	(8, 191)
Remain well with LNG-IUS	1	Fixed
No treatment to change to LNG-IUS	0.001	(1, 852)
Remain no treatment	0.992	(846, 7) Dirichlet
No treatment to surgery	0.007	(6, 847)
Surgery to post surgery	1	Fixed
Post surgery to post surgery	1	Fixed

α and β values for the PSA distribution are rounded to the nearest whole number. LNG-IUS; levonorgestrel-releasing intrauterine system, PSA; probabilistic sensitivity analysis.

6.3 Analysis

An incremental CUA which provides information on the difference in costs and QALYs between LNG-IUS and usual medical treatment is reported as an incremental cost-effectiveness ratio (ICER), in terms of cost per QALY gained. If a treatment is less costly and generates a greater number of QALYs, dominance is said to occur. Analysis was by intention-to-treat to provide a pragmatic estimate of ICERs. The base case analysis and sensitivity analyses will be first carried out using EQ-5D and then using SF-6D.

Uncertainty in the model was assessed by conducting both deterministic and probabilistic sensitivity analysis (PSA). Population heterogeneity was not considered by assessing the cost-effectiveness according to population subgroups because the randomised nature of the trial should mean that there are no systematic differences between women in each treatment arm.

Deterministic sensitivity analyses were carried out where three individual model parameters were changed as set out below.

1. In the first, the mean utility values for each state were replaced by the median utility value. Previously some published studies used the median and not the mean value, which greatly impacts the cost-effectiveness results and is argued to be inappropriate (Roberts et al., 2011). The current analysis assesses the impact of using such values when primary data are collected.

2. In sensitivity analysis 2, the assumptions used in the UK national guidelines costing template which include a practice nurse for only 10 minutes for the initial consultation and insertion, treatment review by a nurse at 6 weeks only for those with a LNG-IUS fitted and annual follow-up for both treatment groups thereafter were incorporated (NICE, 2007).

3. Finally, in the base case, it was assumed that when an EQ-5D (SF-6D) completion date and notification of change of treatment coincide, the EQ-5D (SF-6D) value will belong to the subsequent state. In sensitivity analysis 3, the EQ-5D (SF-6D) value to the state prior to the change was assigned.

The PSA simultaneously changes all relevant parameters in the model. For each parameter, a distribution is assigned and a value for each parameter is randomly drawn from the assigned distribution. This is repeated 1000 times and the range of incremental cost and QALY results for LNG-IUS and usual medical treatment are presented on the cost-effectiveness plane. These 1000 values were used to construct a cost-effectiveness acceptability curve (CEAC) to illustrate the probability of LNG-IUS being more cost-effective than usual medical treatment, across a range of monetary values that decision-makers may be willing to pay for an additional QALY.

6.3.1 Distributions for the Probabilistic Sensitivity Analysis

A Dirichlet distribution was used for the transition probabilities because this distribution allows for a possible movement to more than two states. Where a transition between states is possible but did not occur in the trial data, a value of one is added to all such transitions to enable the Dirichlet distribution to be assigned and to compensate for not observing the transition in the data (Briggs et al., 2003).

In Table 6.3 it can be seen that some distributions are fixed at the given value. This is because these transitions are not dependent on the data and logically it would not be appropriate for the transition to be made probabilistic. For example, the transitions for remaining in well states should remain at 100% because these are absorbing states and women cannot leave them. Fixed distributions are also assigned for the transition to 'symptomatic' to 'well' because given the assumptions of this model, that once a woman changes treatment and moves elsewhere she is classed as symptomatic, this transition to well from symptomatic is not possible.

Costs are not assigned distributions because they are primarily unit costs, which are variable but not uncertain. Where costs are uncertain, the PSA on utilities and transitions will be sufficient to account for this uncertainty. Furthermore, as distributions for these costs, and transitions, would be particularly complex, the small number of uncertain costs and very small cost values are unlikely to affect the results in a manner sufficient to warrant using a separate distribution.

Utility parameters are assigned a beta distribution using the method of moments. This is thought to be most suitable because event data are not used. In this case average values are used for each state and the method of moments is most suitable (Briggs et al., 2006). As the utility value for surgery is ¾ of the value for post-surgery, the distribution for surgery is linked to that of post-surgery. It would not be appropriate to sample for an individual distribution for 'surgery' because the value was taken from post-surgery and so this state does not have individual values to sample from.

6.4 Results

The results of the base case analysis and sensitivity analyses are presented in two parts. Part 1 depicts the results using EQ-5D and Part 2, the results using SF-6D.

6.4.1 Part 1: EQ-5D

At the 2-year time point 93 women had withdrawn from trial or were lost to follow-up (55 in the usual treatment arm, with 16 lost to follow-up and 38 in LNG-IUS, with 13 lost to follow-up). In line with the trial findings, missing follow-up questionnaires were assumed to be missing at random. Complete EQ-5D data were available at all time points for 367 patients (64%). LNG-IUS had a slightly higher completion rate (66%) than the usual medical treatment arm (63%). However, as a decision model was employed and utility data at any time point were used for the health states, it was not necessary to have complete EQ-5D data at all three time-points to include it in the analysis. Therefore, 75% complete EQ-5D data were available at 6 months (76% in LNG-IUS and 73% in usual treatment), 77% at 1 year (77% in LNG-IUS and 77% in usual treatment) and 84% at 2 years (87% in LNG-IUS and 81% in usual treatment). A complete case analysis is used and is discussed further in section 6.5.1.

Table 6.4 presents a summary of the base case and the deterministic sensitivity analysis results using EQ-5D.

The base case results show that LNG-IUS costs £100 more than usual medical treatment, as it costs £430 whilst usual medical treatment costs £330. However, LNG-IUS also generated 0.067 more QALYs than usual medical treatment as LNG-IUS generated 1.580 QALYs and usual medical treatment 1.513 QALYs. The ICER for the base case analysis shows that LNG-IUS generates £1600 per additional QALY when compared to usual medical treatment.

In all three of the deterministic sensitivity analyses, the findings supported the base case results. However, the ICER in each analysis did differ. Sensitivity analysis 1 and 2 had a slightly less favourable effect, increasing the ICER to £2030 and £1640 per QALY gained respectively, whilst sensitivity analysis 3 resulted in a more favourable effect on the ICER with a reduction to £1560 per additional QALY.

Table 6.4 Base case and deterministic sensitivity analysis results using EQ-5D

	Total mean costs per intervention (£)	**Total mean QALYs per intervention**	**Incremental cost effectiveness ratio (ICER) (v usual medical treatment)**
Summary of base case deterministic results			
Usual medical treatment	330	1.513	1600
LNG-IUS	430	1.580	
Mean Difference	100	0.067	
Deterministic sensitivity analysis 1*			
Usual medical treatment	330	1.590	2030
LNG-IUS	430	1.643	
Mean Difference	100	0.053	
Deterministic sensitivity analysis 2#			
Usual medical treatment	340	1.513	1640
LNG-IUS	450	1.580	
Mean Difference	110	0.067	
Deterministic sensitivity analysis 3°			
Usual medical treatment	330	1.514	1560
LNG-IUS	430	1.582	
Mean Difference	100	0.068	

Cost are rounded to nearest 10. QALYs are rounded to 3 decimal places QALYS; quality adjusted life year, LNG-IUS; levonorgestrel-releasing intrauterine system, ICER; incremental cost-effectiveness ratio.
*Deterministic sensitivity analysis 1 = Use median utility values
#Deterministic sensitivity analysis 2 = Use NICE assumptions
°Deterministic sensitivity analysis 3 = Assigning EQ-5D completion date utility for change treatment, if change treatment date is the same as EQ-5D completion date.

The results of the PSA (Figure 6.7) illustrate the distribution of the incremental costs and effects from 1000 Monte Carlo simulations. It is shown that the majority of the incremental costs and effects lie in the upper right hand quadrant of the cost-effectiveness plane. This indicates that LNG-IUS is both more costly and more effective than usual medical treatment. The uncertainty is then summarised in relation to the changes in the decision-makers threshold for considering an intervention cost-effective in Figure 6.8. It depicts the CEAC, which shows that from £2000 per QALY, LNG-IUS has a greater probability of being the more cost-effective intervention. This probability increases to over 90% at approximately £4000 per QALY.

Figure 6.7 Results of the probabilistic sensitivity analysis (EQ-5D)

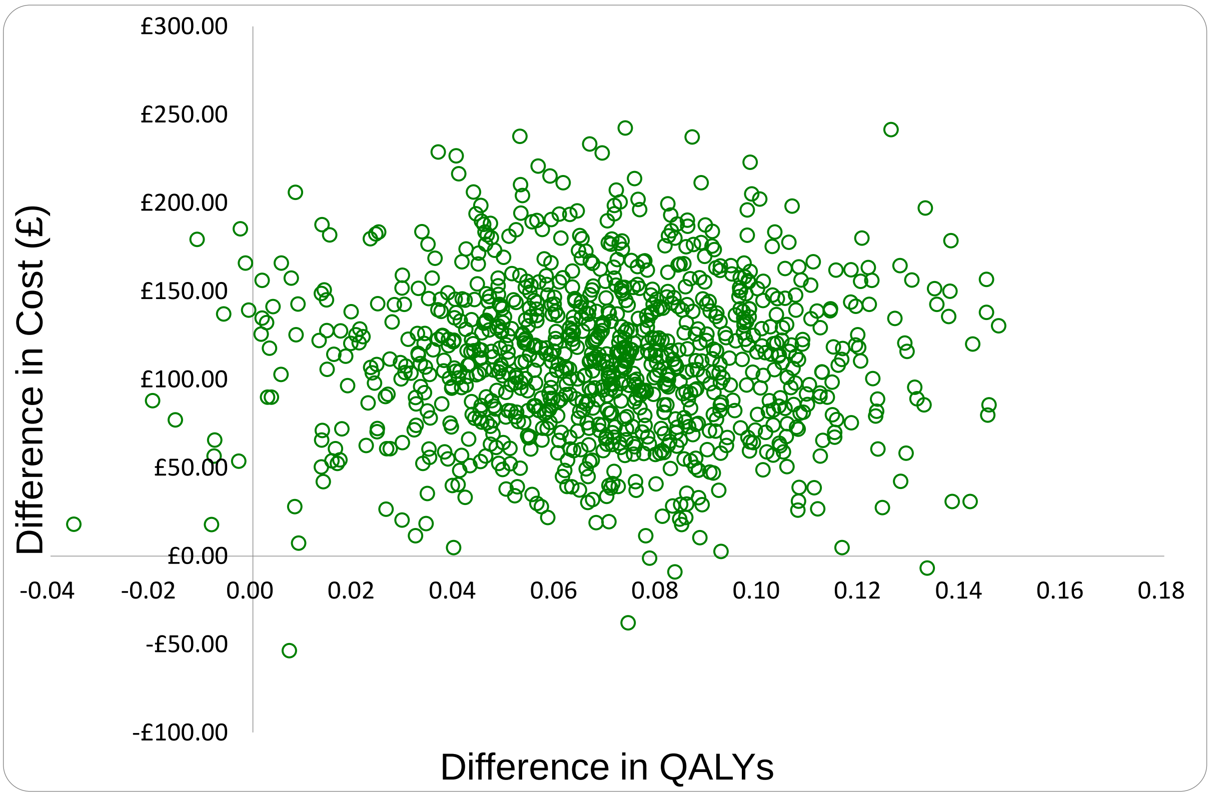
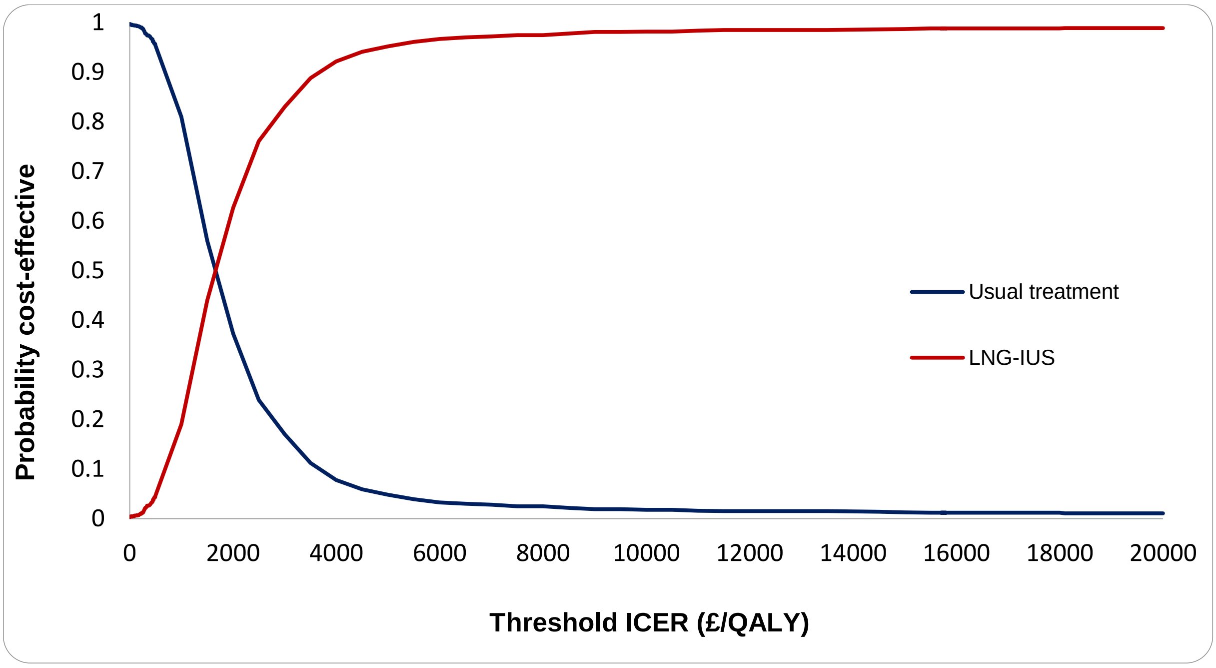


Figure 6.8 Cost-effectiveness acceptability curves for usual medical treatment and LNG-IUS using EQ-5D



6.4.2 Part 2: SF-6D

As mentioned in section 6.4.1 by the 2-year time point 93 women had withdrawn from trial or were lost to follow-up (55 in the usual treatment arm, with 16 lost to follow-up and 38 in LNG-IUS, with 13 lost to follow-up). Complete SF-6D data were available at all time points for 355 patients (62%). LNG-IUS had a slightly higher completion rate (64%) than the usual medical treatment arm (60%). However, as a decision model was employed it was not necessary to have complete SF-6D data at all three time-points to include it in the analysis. Therefore, 73% complete SF-6D data were available at 6 months (76% in LNG-IUS and 71% in usual treatment), 76% at 1 year (77% in LNG-IUS and 76% in usual treatment) and 82% at 2 years (85% in LNG-IUS and 80% in usual treatment).

Table 6.5 presents the base case and deterministic sensitivity analysis results for SF-6D.

The base case results show that LNG-IUS costs £100 more than usual medical treatment as it costs £430 whilst usual medical treatment costs £330. However, when SF-6D is used, usual medical treatment is shown to generate 0.002 more QALYs than LNG-IUS, as usual medical treatment generated 1.200 QALYs and LNG-IUS 1.198 QALYs. Therefore, it is shown that usual medical treatment dominates LNG-IUS.

In two of the three deterministic sensitivity analyses, the findings supported the base case results. In sensitivity analysis 1 and 2 usual medical treatment dominated LNG-IUS. Whilst in sensitivity analysis 3, usual medical treatment does not dominate LNG-IUS. LNG-IUS was shown to be more effective than usual medical treatment and more expensive, generating an ICER of £112,340.

The results of the PSA in Figure 6.9 are presented on the cost-effectiveness plane, as previously, and show the distribution of the incremental costs and effects from the 1000 Monte Carlo simulations. In this case when SF-6D is used to generate the effectiveness outcome, it can be seen that the incremental costs and effects are spread between the upper right hand and upper left hand quadrant of the cost-effectiveness plane, indicating that LNG-IUS is either more costly and more effective, or more costly and less effective. This is then summarised in the CEAC in Figure 6.10 to show that for any threshold willingness-to-pay per QALY, usual medical treatment has the greater probability of being the more cost-effective intervention. This probability is 100% at £0 per QALY and decreases to 90% at approximately £20,000 per QALY.

Table 6.5 Base case and deterministic sensitivity analysis results using SF-6D

	Total mean costs per intervention (£)	Total mean QALYs per intervention	Incremental cost effectiveness ratio (ICER) (v usual medical treatment)
Summary of base case deterministic results			
Usual medical treatment	330	1.200	Dominates
LNG-IUS	430	1.198	
Mean Difference	100	-0.002	
Deterministic sensitivity analysis 1*			
Usual medical treatment	330	1.215	Dominates
LNG-IUS	430	1.215	
Mean Difference	100	0	
Deterministic sensitivity analysis 2#			
Usual medical treatment	340	1.200	Dominates
LNG-IUS	450	1.198	
Mean Difference	110	-0.002	
Deterministic sensitivity analysis 3°			
Usual medical treatment	330	1.198	112,340
LNG-IUS	430	1.199	
Mean Difference	100	0.001	

Cost are rounded to nearest 10. QALYs are rounded to 3 decimal places. QALYS; quality adjusted life year, LNG-IUS; levonorgestrel-releasing intrauterine system, ICER; incremental cost-effectiveness ratio.
*Deterministic sensitivity analysis 1 = Use median utility values
#Deterministic sensitivity analysis 2 = Use NICE assumptions
°Deterministic sensitivity analysis 3 = Assigning SF-6D completion date utility for change treatment if change treatment date is the same as SF-6D completion date.

Figure 6.9 Results of the probabilistic sensitivity analysis (SF-6D)

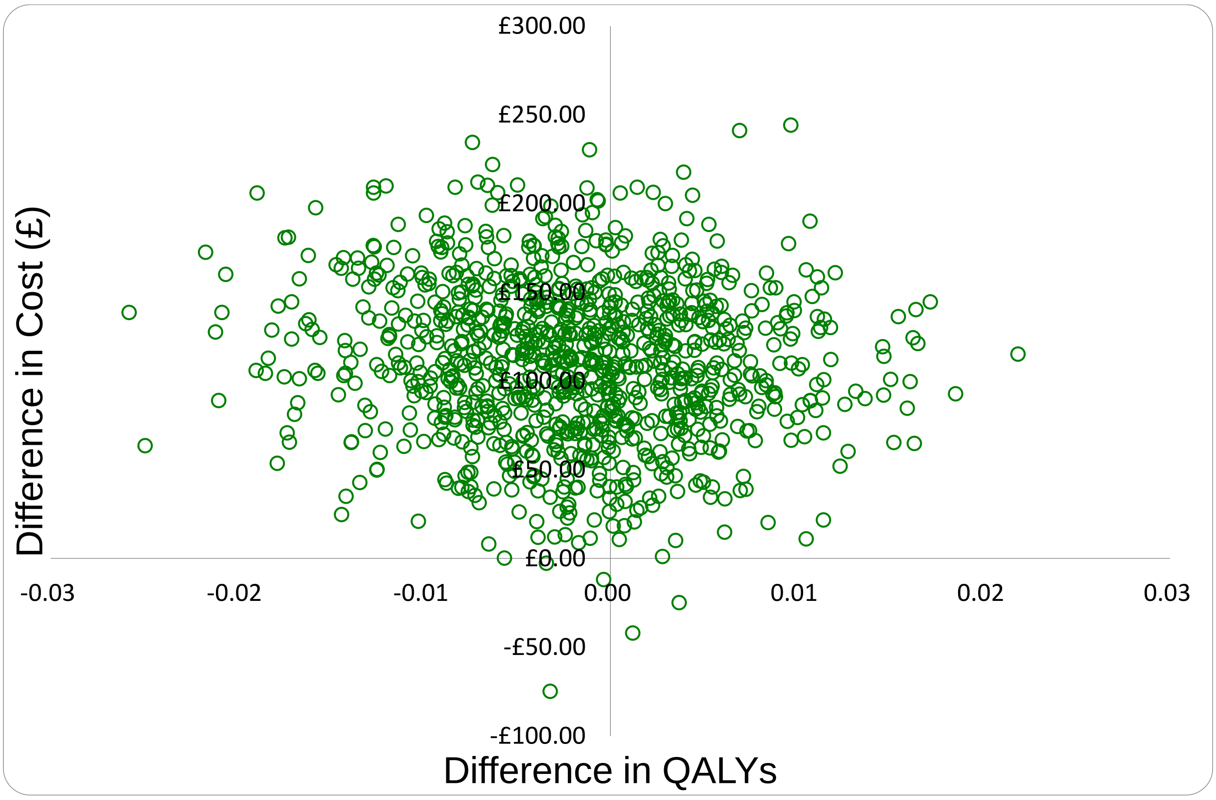
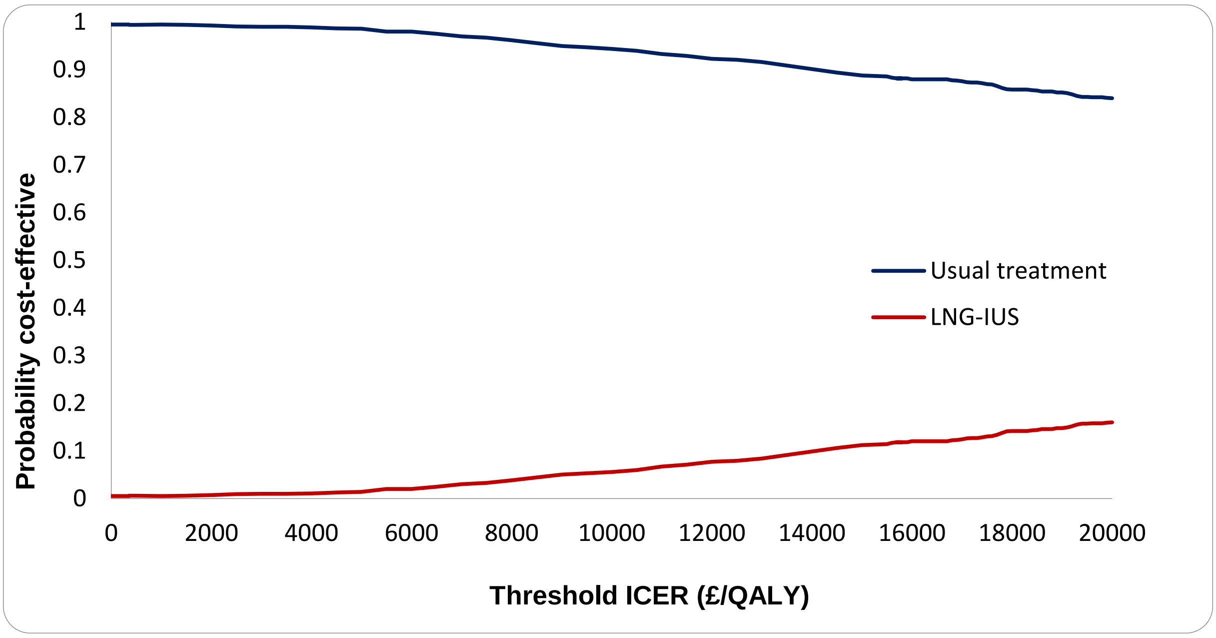


Figure 6.10 Cost-effectiveness acceptability curves for usual medical treatment and LNG-IUS using SF-6D



6.5 Discussion

In this chapter, the economic evaluation comparing LNG-IUS against usual medical treatment has been presented. The economic evaluation was first conducted using the EQ-5D to generate QALYs and then re-estimated using SF-6D to assess the impact of the alternative extra-welfarist measures on the cost-effectiveness results.

Using EQ-5D, treating menorrhagia using LNG-IUS costs more but is also more effective than usual medical treatment. The relative cost-effectiveness of LNG-IUS compared to usual medical treatment is £1600 per QALY. This means every additional QALY costs an extra £1600. The deterministic sensitivity analyses showed the uncertainty in this ICER to be in the range of £1,560-£2,030 for an additional QALY gained. As the NICE guidelines recommend new interventions into practice if the ICER is below £20,000 per QALY, LNG-IUS would be considered cost-effective and recommended as the primary choice for patients who: require treatment, have no preference against contraception or intrauterine insertion, and no contraindication to LNG-IUS insertion.

However, the importance of selecting the most appropriate QoL instrument is highlighted when the measure used to assess QoL is changed from EQ-5D to SF-6D. In the second analysis, Part 2, where utility values from SF-6D are used rather than EQ-5D, the cost-effectiveness results differ. In contrast to the findings using EQ-5D, usual medical treatment is the more cost-effective intervention. Usual medical treatment was found to dominate LNG-IUS in the base case and the two deterministic sensitivity analyses. In the third deterministic analysis, the ICER for LNG-IUS was £112,340 per QALY which is evidently greater than the £20,000 decision-maker willingness-to-pay threshold currently set by NICE, therefore usual medical treatment would be recommended for implementation in clinical practice.

The difference in the cost-effectiveness results derived by using the alternative extra-welfarist measures to value utilities has a considerable impact on the cost-effectiveness decision. The different measures did not just change the strength of cost-effectiveness of the same treatment, but the most cost-effective treatment itself changed. Therefore the recommendation to decision-makers would differ depending on the QoL instrument used.

6.5.1 Strengths and Limitations

The main strength of this CUA is that it is based on data from the largest multi-centre randomised trial undertaken for menorrhagia. Since the treatment is aimed at managing the condition, the changes in both QoL and costs throughout the woman's treatment journey are critical to the analysis and these are most appropriately captured by using the trial data to populate a model. The model structure was developed based on the patient pathway data from the trial and supported by the advice of expert clinicians. All assumptions were agreed by the team in the model development stage prior to analysis.

There are a number of limitations based on some of the assumptions. For instance it was assumed that women are well if they do not change treatment. However, if women enter the 'well' state in the model but are not actually well it will be reflected in the overall utility value for 'well'.

A further potential limitation is that baseline differences in QoL data at the outset of the trial were not adjusted to be the same. The initial EQ-5D score in the LNG-IUS group was 0.042 higher at the outset of the trial than in usual medical treatment and the difference is significant ($p<0.05$). These data are based on individuals who have been randomised, so it is assumed that this difference occurred by chance and it does not follow that this initial difference between groups would be sustained over the 2 year time horizon in the absence of treatment. Adjustment for baseline therefore risks imposing a difference at every point in time over the time period of analysis.

If a full adjustment was made for the difference in baseline, LNG-IUS would be shown to be less effective and potentially less cost-effective than usual medical treatment, but in so doing, the difference would be maintained over the time horizon and the probability of regression to the mean over time would be ignored. Therefore, it is acknowledged that the base case results might be over optimistic, but it is not clear how much of an adjustment, if any, would be appropriate.

A final limitation is the inability to handle missing data using conventional methods such as multiple imputation due to the nature of the analysis, as a result a complete case analysis was used. Missing values can be considered as a potential source of bias, particularly if the complete data differ from the missing data and if there is a chance that the reason for missingness is due to the intervention (as is the case for this evaluation as data are considered to be missing at random). Whilst it is not possible to establish the relationship between the missingness and the unobserved values with certainty it is suggested that approaches for handling missing data be considered (Carpenter & Kenward, 2007). Therefore, as missing cases are deleted in a complete case analysis potentially revealing data are excluded and the analysis could be subject to bias. The use of a complete case analysis also means that conclusions are drawn from a smaller sample size than initially stated which can potentially limit the generalisability of the findings to the wider population with menorrhagia (Briggs et al, 2003). Therefore the most appropriate method to handling missing data for this study would have typically been multiple imputation, where the observed values are used to generate predicted values for the missing data in order to create a full dataset, this imputation minimises the potential for bias. However, as the trial data are used in a decision model, it would be necessary to impute missing data at the individual level rather than overall at the QALY level for each time point. As this has not been attempted before and is not yet a recommended method it did not seem appropriate. Despite this limitation, the results are not unduly affected. The PSA (described in section 6.3) produces a range of iterations using the distributions assigned for the utilities to handle any uncertainty around the results and will to an extent compensate for not handling missing data.

6.5.2 Comparison with Other Studies

To current knowledge, this is the first study to conduct a CUA using prospectively collected primary data from a trial to compare LNG-IUS and usual medical treatment for menorrhagia. Whilst another primary study has shown LNG-IUS to be cost-effective (Hurskainen et al., 2004), the comparator was hysterectomy, and the study is considered to have methodological flaws (Roberts et al., 2011). As shown in Chapter 4, other studies have compared these alternative treatments using model based analyses and secondary data from reviews (Blumenthal et al., 2006; NICE, 2007; You et al., 2006). But the studies typically compared various *surgical* techniques, LNG-IUS and various oral treatments against one another. Two of these showed LNG-IUS to be the most cost-effective intervention (Blumenthal et al., 2006; NICE, 2007). The remaining study suggested that hysterectomy was the optimal intervention but the authors acknowledged that insufficient published data on the effectiveness of LNG-IUS were available at the time (You et al., 2006). Thus, the results reported in this chapter are new and based on the largest randomised controlled trial comparing LNG-IUS and usual medical treatment. When using the decision-maker recommended EQ-5D measure, the results concur with some of the studies which are based on secondary evidence.

As EQ-5D is recommended by NICE for use in economic evaluation, this study has determined that LNG-IUS is a cost-effective treatment at 2 years of follow up but as the condition is chronic and continues until menopause, further economic evaluation with a longer term follow-up is needed, and planned to include all relevant costs and outcomes (Blumenthal et al., 2006; NICE, 2007; You et al., 2006).

6.5.3 Implications and Further Research

The results, based on the decision-maker recommended EQ-5D, provide clear evidence in support of the NICE guidelines that recommend LNG-IUS be considered the primary treatment for menorrhagia. As the findings presented here are based on data from the largest multi-centre randomised controlled trial conducted for menorrhagia, this study provides robust evidence in favour of LNG-IUS over usual medical treatment. As the study was carried out to provide evidence to decision-makers on the cost-effectiveness of these two treatments for menorrhagia the primary results which are taken from this study are based on EQ-5D. However, as SF-6D data were also collected, the use of this measure was deemed worthy of exploration.

In contrast to the EQ-5D finding, when SF-6D is used to generate QALYs usual medical treatment is the more cost-effective intervention. As the results are based on data from a large randomised controlled trial they indicate that the measures may be capturing different aspects of QoL, which clearly has a great impact on the results. The reasons for the differences in results will be explored elsewhere as it is beyond the remit of the thesis and has been observed in several other disease areas (Davis et al., 2012; Sach et al., 2009). The main difference between the instruments has already been explained by Brazier et al (2004) and Whitehurst and Bryan (2011). Since QoL is the key indicator of success in menorrhagia and the woman's own assessment of this determines treatment choice, it must be determined which measure is most suitable given that the instrument used has such a significant impact on the results. Furthermore, as the results of the systematic review on QoL instruments in Chapter 4 showed that these generic extra-welfarist measures might not be appropriate, further research should explore the use of alternative measures, such as the welfarist willingness-to-pay (WTP), to estimate the appropriateness of alternative instruments and alternative ways of capturing and measuring this outcome.

6.6 Conclusion

In this chapter an economic evaluation alongside the ECLIPSE trial, comparing LNG-IUS and usual medical treatment is reported. The economic evaluation was first estimated using the decision-maker recommended EQ-5D-3L and then re-estimated using SF-6D. The findings from the economic evaluation using EQ-5D showed that LNG-IUS is the most cost-effective treatment. Therefore, the recommendation to decision-makers would be that LNG-IUS be used as the first line treatment for menorrhagia in clinical practice. However, when SF-6D was used to generate QALYs, usual medical treatment was shown to be the most cost-effective intervention. These conflicting findings show that the recommendation to decision-makers would differ depending on the measure used. This illustrates the importance of ensuring that the most suitable measure is used to value outcomes in menorrhagia. Given the conflicting findings presented in this chapter, and that the systematic review on economic outcome measures illustrated that there are concerns with the use of extra-welfarist measures, in the next chapter the use of WTP will be explored. Specifically, the methodology used for the WTP empirical work is reported next.

CHAPTER 7. WILLINGNESS TO PAY STUDY – METHODOLOGY

7.1 Introduction

In this chapter the methodology associated with the willingness-to-pay (WTP) research is described. It has been suggested in Chapters 5 and 6 that the routinely used health related quality of life (QoL) measures (SF-36 and EQ-5D) may not be entirely suitable for use in menorrhagia. It is important to ensure that an appropriate measure of benefit is used because impact on QoL is the sole measure of benefit in this condition. Thus the aim of this empirical research is to explore the suitability of using the WTP measure, which enables the respondent to take into consideration both health and non-health outcomes and may overcome the issue of timing of assessment. In this feasibility study, WTP is elicited from both an ex-ante and an ex-post perspective. In this chapter, the methodology for the study design and data collection are outlined in the following order: (1) the WTP study from an ex-ante perspective, comparing LNG-IUS against usual medical treatment; (2) the WTP study from the ex-post perspective, comparing LNG-IUS against usual medical treatment; and (3) the interviews with a sub-group of the respondents who completed the WTP questionnaire from the ex-ante perspective. These sections are then followed by details of the data analysis and a discussion of the methodology used. In the next chapter the results of each of the analyses are presented, in turn, along with conclusions on the implications of the findings and the prospects for future work.

7.2 Method

A substantial amendment to the ECLIPSE trial ethics was required, because the exploratory research, related to WTP, was a sub-study of the overall funded project. A self-complete questionnaire for the ex-ante and ex-post perspective was created based on previously successfully completed questionnaires (Donaldson et al., 1997b; Frew et al., 2001). The content of each questionnaire was reviewed and commented on by several experts, including a consultant who specialises in the area, a general practitioner (GP), psychologists, social scientists and external health economists. The questionnaires were subsequently revised according to their suggestions.

The study protocol was then submitted and received favourable approval from the ethics committee and the research and development department of the hospital.

7.2.1 Methodology of Ex-ante Perspective

As explained in Chapter 2 in the discussion of the ex-ante perspective, the WTP value reflects the level of expected change in utility from having the treatment available. The assumption is that treatment will improve utility therefore a WTP value, and not a willingness-to-accept (WTA) value, was elicited. This section details the methods used to elicit WTP from an ex-ante perspective using respondents from the general population.

1) Design of Questionnaire for Ex-ante WTP Elicitation

A booklet questionnaire was designed for completion without supervision to capture data on WTP, patient socio-demographic details and EQ-5D. First, in the questionnaire a description of menorrhagia and its treatments was required, since respondents may not have had any experience of either menorrhagia or the treatment. Therefore data from the ECLIPSE trial, (described in Chapter 6), were used to develop the scenarios for the ex-ante perspective as the trial data are related to women with experience of menorrhagia and its treatments.

1a) Scenario description development

First, a scenario description of the impact of menorrhagia on sufferer's lives was outlined. The description of menorrhagia was based on the domains of the menorrhagia multi-attribute scale (MMAS) QoL instrument because this measure incorporates both the health and non-health outcomes associated with menorrhagia, which are known to be affected by the condition. Generic extra-welfarist measures (EQ-5D and SF-6D) were not used due to their narrower focus on health-related QoL. The scenario was based on the average value for each MMAS domain that was observed at baseline in the ECLIPSE trial. In other words, the trial results were used to generate a description of the expected average 'outcome' for menorrhagia, and this outcome contained both health and non-health variables. Data on MMAS from the ECLIPSE trial were used because this was thought to be the most reliable and practical source of data on how menorrhagia and its treatments impact women's lives.

The 'practical difficulties' attribute of the MMAS is presented below to provide an example of the calculation. Table 7.1 below shows the number of responses for each of the four levels of the practical difficulties attribute at baseline. Box 7.1 shows the weighting for the levels of the practical difficulties attribute.

Table 7.1 Frequency of responses for the practical difficulties attribute

Practical Difficulties	Freq.	Percent	Cum.
a - No practical difficulties	13	2.37	2.37
b - Carry extra sanitary protection	186	33.88	36.25
c - Carry clothes	184	33.52	69.76
d - Severe problems	166	30.24	100.00
Total	549	100.00	

Box 7.1 Weightings for MMAS attribute

1. Practical difficulties	**Weighting**
a. I have no practical difficulties, bleed no more than I expect and take no extra precautions.	14.0
b. I have to carry extra sanitary protection with me but take no other precautions.	9.4
c. I have to carry extra sanitary protection and clothes because of the risk of flooding.	3.1
d. I have severe problems with flooding, soil the bedding and need to be close to a toilet.	0

The following formula, which was deemed the most intuitive, was applied to calculate the average value for each attribute of the MMAS.

$$[(f_a \times w_a) + (f_b \times w_b) + (f_c \times w_c) + (f_d \times w_d)] / Tf \qquad [1]$$

For each level of the attribute w is the weighting, f is the frequency, and Tf is the total frequency for the attribute. The subscript letters (a, b, c, d) represent the different levels of the attribute, as notated in Table 7.1. Shaw et al (1998) had previously determined the weighting for the levels of all the attributes, which has been described in section 2.5.1 of Chapter 2 (see Appendix 4.1 for MMAS questionnaire and weightings).

Using the formula, the average value for the ‘practical difficulties’ attribute was calculated as follows: the frequency of (a)- no practical difficulties (13) was multiplied by the weighting for the level (14) which equals 182. This calculation was then carried out for every level of the ‘practical difficulties’ attribute. To obtain the average value for the attribute, the total for each level was then summed and divided by the total frequency (549). In this case the average value is 2500.8 divided by 549 which gives 4.56. From Box 7.1 it can be seen that the nearest level weighting to 4.56 is 3.1. Therefore the average baseline value for practical difficulties is (c). This method was then repeated for every attribute of the MMAS at baseline to present a full scenario description of the impact of menorrhagia on women's lives.

The scenario description of menorrhagia was then followed by a description of the outcomes for the treatments, LNG-IUS (termed 'Mirena' in the questionnaire) and usual medical treatment (termed Oral treatment in the questionnaire). The outcomes associated with Mirena and Oral treatment were similarly based on the MMAS data from the ECLIPSE trial to capture changes in health and non-health outcomes associated with menorrhagia. However, in this case the 6 month data on the average MMAS score for each treatment were used. The 6 month follow-up was the next data collection time point in the ECLIPSE trial after baseline, as after 6 months it is more than likely that the benefits of both treatments will begin to be seen. For each treatment, the average outcomes (expressed as health and non-health outcomes) were described using the domains within the disease-specific MMAS questionnaire and the results for each treatment arm within the ECLIPSE trial. At 6 months, for each treatment, the same process and formula [1] for the menorrhagia description was used to calculate the average outcome for each attribute. Similarly, the closest level weighting to the value derived for each attribute was then chosen as the average scenario for that attribute. This meant that the scenario descriptions were kept consistent to make it easier for respondents to identify potential benefits of treatment in terms of the impact on health and non-health outcomes and process utility.

1b) Scenario description presentation

Bullet points were used to present the scenario for menorrhagia, as is typically done in health economics for EQ-5D and SF-6D to generate a quality adjusted life year (QALY) (Smith, 2008). For the scenario description of Mirena and Oral treatment, the outcomes associated with treatments are presented as bullet points and the process utility derived from the intervention is presented using a concise narrative (Smith, 2006; Smith, 2007). The process utility is based on the method of administration of the treatment, its burden on the patient, the most likely side effects and the length of time it takes for the treatment to begin working successfully. Information on the process utility of the two treatments was primarily obtained from the literature (CKS, 2012) and was checked against data from the ECLIPSE trial and by relevant professionals (Consultant gynaecologist, Professor Janesh Gupta, and GP, Professor Joe Kai).

The health and non-health outcomes (MMAS) achieved from the intervention in addition to the process utility derived from the intervention itself were included to provide a comprehensive description of the impact of the treatments on menorrhagia and in turn on QoL. This approach was used because the impact of treatment on QoL is the primary measure of success in menorrhagia. The use of outcomes as well as the process associated with each intervention ensured that the WTP measure was not constrained to the time-point of measurement and that the measure could be applied for any treatment of menorrhagia.

At the end of the scenarios for both Mirena and Oral treatment, three tick boxes were presented to determine the respondents' preference for treatment: Mirena, Oral treatment or no preference.

1c) Choice of elicitation method used

Next, a maximum monthly out of pocket WTP value up from the current point in time until menopause was elicited for both Oral treatment and Mirena. The time frame of payment of ‘up until menopause’ was explicitly stated to ensure that: WTP values were not overestimated, individuals were aware of their budget constraint and recognised that spending money on the commodities measured would mean money could not be spent elsewhere (Smith, 2003). The time frame was selected intuitively given the nature of the condition, as menorrhagia would cease when menopause begins.

The payment scale elicitation format was used to elicit WTP values because it has a higher completion rate than other self-completion methods that can be used in a postal questionnaire (Whynes et al., 2003). Although the payment scale format can be prone to range bias where respondents WTP values may be influenced by the range presented, this issue can be minimised by piloting the questionnaire to determine the most suitable range. Alternatively, the use of several ranges of payment scales could be used and randomly allocated to respondents. However, the latter approach would require a large sample size (McIntosh et al., 2010).

A payment scale range of £0-£500 was considered to be most suitable, given that the questionnaire asked respondents to provide a monthly WTP value and this estimate is line with likely levels of ability to pay. For example, it may be unrealistic to expect women to pay values greater than £500 on a monthly basis, but an option to pay more than £500 was presented in an open-ended question. The scale began with £0-£20 in increasing increments of £2, from £20-£50 in increments of £5 and £50-£100 increments of £10. The payment scale was immediately followed by an open-ended question to state a WTP value if greater than £500. Two vertical payment scales were presented alongside each other for each treatment. The questionnaire asked respondents to circle a maximum monthly WTP for Mirena and then Oral treatment. This payment scale structure was chosen as it is similar to previously published scales (Frew et al., 2001). The questionnaire included a reminder to respondents to consider the amount that they can afford to pay to ensure that the responses obtained were realistic and within the respondents means (Smith, 2006). The monthly payment time frame was used because women generally pay monthly (or every three months) for prescriptions for menorrhagia and for sanitary protection. The out of pocket payment vehicle was deemed to be most appropriate for this context because there is some form of private payment for prescriptions for the Oral treatment. Although this private payment does not exist for Mirena it is argued that the existence of private payment within this context would minimise the issue of hypothetical bias.

Following the WTP question, an open-ended question was presented to the respondent to outline reasons for their WTP values. The next question was a yes or no tick box, which asked the respondent to indicate whether they found the WTP question difficult to answer. This tick box was then followed by an open-ended question which asked for the reason they found the question difficult. The format used follows that of a previously published study (Donaldson et al., 1997b). As the aim of the research was to assess the suitability of the measure in menorrhagia it was necessary to identify if respondents found the question difficult to answer. The subsequent questions were related to socio-demographic details such as age, marital status, employment status, household income (eight income ranges, starting from less than 10,000 to more than 70,000), whether they have experienced heavy periods, reason for visiting the clinic, satisfaction with life overall (five tick boxes starting from 'very unsatisfied' to 'very satisfied'), EQ-5D and an invitation for a follow-up interview within 1 to 2 months.

If the women agreed to the interview, consent was taken for the patient to be contacted for interview and for the interviews to be recorded. The women either gave their email or home address details. The full ex-ante questionnaire is presented in Appendix 4.2.

2) Population Sampled

Women visiting the fertility and general outpatient clinics at the Birmingham Women's Hospital were invited to complete the questionnaire. This convenience sample of women attending the Birmingham Women's Hospital was recruited for practical reasons. Firstly, a group of women who are menstruating were easily accessible through this route and, secondly, the use of women visiting a clinic would ensure a higher response rate than a cold mail out. The women were approached prior to their appointment, which may have been routine or one-off, and given a patient information sheet to read, which explained the study and what would be required. The consenting women were asked to complete the questionnaire in the clinic either prior to or after their appointment. This author then collected the completed questionnaires from the women. Where women requested to complete the questionnaire at home a pre-paid stamped addressed envelope was provided.

Questionnaires were administered over a 2 month period (December 2012-January 2013). Due to the nature of the study the aim was to recruit between 50-100 respondents. It was the opinion of the research team that this size would be sufficient and pragmatic for an exploratory WTP study given the time constraints of the research.

The next section outlines the methodology behind the questionnaire design, elicitation method and population for the ex-post perspective.

7.2.2 Methodology of Ex-post Perspective

As explained in Chapter 3 from the ex-post perspective WTP, the WTP value reflects the level of actual change in utility from having the treatment available. This section details the methods used to elicit WTP from an ex-post perspective, thus WTP was elicited from respondents who were suffering with the condition and have experience of treatment.

1) Design of Questionnaire for Ex-post WTP Elicitation

A booklet questionnaire was designed for self-completion without supervision to capture data on disease-specific QoL (MMAS), WTP, patient socio-demographic details and EQ-5D (See Appendix A4.3). The design of the questionnaire was based on the standard questionnaire booklet sent to the ECLIPSE trial women as part of the trial follow-up. As the ex-post perspective requires respondents who are experiencing menorrhagia and its treatments it was not necessary to present scenario descriptions of menorrhagia and the treatments.

1a) Choice of elicitation method used

To assess the disease-specific QoL, the MMAS was first presented in the questionnaire. The MMAS questionnaire was followed by questions to identify which treatment was currently taken and which have previously been taken as part of the ECLIPSE trial. Following these questions the WTP value for the respondent's current treatment was elicited.

Similar to the ex-ante study, a maximum monthly out of pocket WTP value up until menopause was elicited, the time frame of payment of 'up until menopause' was explicitly stated and the payment scale elicitation format with the same range and increments as outlined in 7.2.1 was used for reasons explained previously. However, due to the nature of the ex-post study the questionnaire only required respondents to circle their maximum monthly WTP for their *current* treatment. Thus only one vertical payment scale was presented in the questionnaire. This was immediately followed by an open-ended question which asked the respondent to state the WTP value if it was greater than £500, as in the ex-ante study.

To be consistent with the ex-ante study, following the WTP question an open-ended question was presented to the respondent to outline reasons for their WTP values. The next question was a yes or no tick box, which asked the respondent to indicate whether they found the WTP question difficult to answer. This tick box was then followed by an open-ended question which asked for the reason why the respondent found the question difficult or not. The subsequent questions were related to socio-demographic details such as age, marital status, employment status, household income (eight income ranges, starting from less than 10,000 to more than 70,000), length of time experienced menorrhagia, impact of bleeding on work, daily activities, pain scale using VAS (0 no pain at all to 10 worst imaginable pain), regularity of cycle, perceived impact of bleeding on QoL (four tick boxes starting from 'not at all affected' to 'extremely affected'), satisfaction with life overall (five tick boxes starting from 'very unsatisfied' to 'very satisfied'), EQ-5D, amended EQ-5D (with the order randomly changed) and number of visits to the GP or hospital as a result of the condition since an ECLIPSE trial form was last completed.

The amended EQ-5D asked women to complete the questionnaire with reference to their health during their cycle. The phrase 'during your cycle' was used as an alternative to 'health today' because this wording is used in the validated disease-specific MMAS questionnaire. As the results of the systematic review on QoL measures in Chapter 5 showed the recall period to be an issue for EQ-5D, it was amended, with approval from EuroQoL, to assess whether this change altered the sensitivity of the measure.

2) Population sampled

Women who were currently enrolled in the ECLIPSE trial and were receiving either Mirena or Oral treatment were recruited for this WTP ex-post study. As women who were currently using either Mirena or Oral treatment were required, the ECLIPSE trial patients seemed most suitable. These women are part of an established trial and were already routinely posting back follow-up questionnaires for the trial analysis, which would presumably result in a high completion rate. Questionnaires were posted out in August 2012 to all ECLIPSE patients, with the exception of those women who had asked to be withdrawn from the trial.

By post, women received the following information: (1) a letter from the principal investigator of the ECLIPSE trial explaining that the woman's assistance in the study would be appreciated and that the WTP questionnaire booklet was different to the ECLIPSE trial follow-up; (2) a patient information sheet outlining the purpose of the work and what is required; (3) an ex-post questionnaire; and (4) a pre-paid stamped addressed envelope to return the completed questionnaire.

Similar to the ex-ante perspective, due to the nature of the study, the aim was to recruit between 50-100 respondents, as this was believed, by the current research team, to be sufficient for an exploratory WTP study.

3) Further Considerations for the Ex-post Perspective

In the case of the ex-post perspective additional data from the ECLIPSE trial were available. The ECLIPSE trial has been described in Chapter 6. Before the analysis of the WTP study is outlined it is necessary to provide details of the data that have already been collected as part of the ECLIPSE trial, as the ex-post perspective results which use this ECLIPSE data are presented in the next chapter.

Women were recruited to the ECLIPSE trial over a 3 year time period. At the time the exploratory work was conducted, women were at various time points in trial follow-up (either 2 years or 5 years). Thus data on MMAS, EQ-5D and resource use are available at baseline, 6 months, 1 year, 2 years, and 5 years where appropriate. This additional data allowed the association between the WTP for the change in outcome (from baseline to current time point) as measured by MMAS and EQ-5D to be assessed.

In the next section the methods for the interviews using women from the ex-ante perspective study are described.

7.2.3 Interview Methodology

The interviews were designed to consider the women's understanding of the question, to delve deeper into their WTP answers and to explore whether these women were completing the questionnaire according to WTP theory, as required by WTP analysts. The aim was to conduct 5-10 interviews. It is accepted that interview sample sizes are small and that they are typically conducted until saturation is observed, i.e. where little new evidence is observed (Ritchie & Lewis, 2003). As the interviews were designed to supplement the questionnaire data, in addition to the time constraints and the resource intensive nature of the interviews it was considered to be feasible to conduct between 5-10 interviews, which is likely to demonstrate saturation.

1a) Population sampled

As the ex-ante perspective is recommended by decision-makers for extra-welfarist measures and women from the ex-ante perspective were asked to imagine the scenarios and apply WTP values, they were thought to be a suitable group to assess the WTP questionnaire. Therefore the women who completed the ex-ante questionnaire were sought for interview.

The 30 women who were eligible and agreed to be interviewed were contacted according to their preferred contact method. Eligibility was assessed according to whether the patient met the strict definition of ex-ante, i.e. could potentially be in the menorrhagia state. Thus women who completed the ex-ante questionnaire and were going through the menopause or surgery, such as hysterectomy, which would mean they could not experience menorrhagia, were not selected first for interview.

1b) Population recruitment process

In late January 2013, women were initially contacted by email to attend an interview at the Birmingham Women's Hospital at a date and time that suited them in February or March 2013. If they did not respond within 1 week they then received the interview invitation by post.

1c) Interview details

On the interview day, women were met in the main reception of the Women's hospital and taken to a designated interview room in the hospital. The interviews were semi-structured, lasted up to 30 minutes and were recorded. The topic guide can be found in the Appendix 4.4.

In the next section, the methods of data analysis for all three studies (ex-ante, ex-post and interviews) are briefly described.

7.3 Analysis

The results of the analysis are presented in three parts. Part 1 refers to the ex-ante perspective, Part 2 refers to the ex-post perspective and Part 3 relates to the interviews. For both the ex-ante and the ex-post studies (Part 1 and Part 2) the mean WTP for the treatments are reported.

7.3.1 Data Preparation

For Part 1 and 2 all data from the questionnaires were entered into STATA (v 11.0) by this author. All of the data were checked twice on different occasions to avoid double entry. This author carried out all data analysis in STATA (v11.0) and Microsoft Excel. In places where a respondent ticked two boxes, as in the case of the 'level of satisfaction with life' variable and EQ-5D, two variables were created named variableBEST and variableWORST. As the name suggests, in one variable the highest score was used and in the other variable the lowest score was used. These were not coded as missing data, as these data were not missing and it was believed to be the most suitable method to ensure as much data as possible were kept in the analysis.

Appropriate variables were re-coded as dummy variables and re-categorised where necessary, i.e. income had eight categories and was reduced to five where each of the five income variables formed a dummy variable; Marital status was re-coded to a dummy variable as married or not and employment was re-coded to a dummy variable as employed or not. In each case when conducting the regression models, discussed next, the most common category formed the reference case for the analysis.

There were four types of responses for the WTP question:

1. Positive WTP values for either treatment
2. Zero values for either treatment, which is not a protest.
3. Non-response
4. Protest response: zero value or non-response

Categories 1 and 2 were the primary WTP values used in the base case analyses for Parts 1 and 2. Category 4 refers to protest responses, which relate to the respondent misunderstanding the exercise and refusing to provide a WTP value. It is generally accepted in the literature that protest zeros should not be included in the primary analysis, however some authors do express concerns regarding the exclusion of such data (Halstead et al., 1992; Jorgensen et al., 1999; Mitchell & Carson, 1989; Whitehead et al., 1993) so the effect of including category 4 (protest responses) responses on the mean WTP were considered in a sensitivity analysis. Where provided, the qualitative reasons behind category 3 and category 4 were also explored.

7.3.2 Econometric Analysis

The maximum mean WTP values were reported with 95% bootstrapped confidence intervals. The bootstrapped confidence intervals were generated in STATA (v11.0) and enable an assessment of uncertainty in the mean value. 1000 bootstrapped datasets were generated by randomly sampling values, with replacement from the observed values. Descriptive statistics were carried out to understand the distribution of the data. WTP data were found to be non-normal and were log transformed to normalise the data. Each WTP value had a constant of 1 added to it to ensure that zero values were incorporated when the data were log transformed, as applied in other WTP studies (Shackley & Donaldson, 2002).

A paired t-test (Part 1 – ex-ante) or two-sample t-test (Part 2 – ex-post) was carried out on the log transformed data to determine the statistical significance of the difference between the two WTP values for each treatment. A test of statistical significance for the difference observed between the socio-demographic details of those who preferred (Part 1), or gave a WTP value (Part 2) for Mirena and those who preferred (Part 1), or gave a WTP value (Part 2) for Oral treatment was carried out using the Chi-squared test for categorical variables and either the two-sample/ paired t-test or the Wilcoxon Rank Sum test for continuous variables, depending on the perspective and the normality of the distribution.

In Parts 1 and 2, an econometric analysis including the collected exploratory variables was carried out to identify predictors of WTP for Mirena and Oral treatment. The sample included zero and positive WTP values, but not protest answers. The regression analyses were backward selection stepwise analyses where all appropriate independent variables were entered into the model and rejected or selected at a $p<0.1$ (Draper & Smith, 1998).

The type of regression model used is dependent on the type of variable that the WTP value is considered to be. Payment scale data can be considered to be discrete in that the respondent's true WTP value lies within the intervals provided in the payment scale. Alternatively, the payment scale data could also be considered as continuous and the circled WTP value would be the maximum WTP and is not considered to lie within an interval. As there appears to be no consensus on the most suitable model to use and in practice a range of regression models are estimated, the two most theoretically suitable approaches that were appropriate for the data were applied. In the two models explored, the WTP variable was treated as either a discrete variable or a continuous variable according to the model specification.

It was expected that a two-part model would be used where the positive WTP values and zero values are assessed separately in two stages. Where the first part is used to identify predictors of (positive) WTP for Mirena and Oral treatment using ordinary least squares (OLS) linear regression and the second part of the model is a logistic regression model used to identify the characteristics of respondents who provide zero WTP values. However the proportion of respondents who provided zero values was very small so it was not appropriate to conduct the two-part model, due to limited variation in the sample. Two-part models are only recommended when the proportion of zero values is high (Donaldson et al., 1998).

In Parts 1 and 2 the base case analysis therefore took the form of a linear OLS model which included both positive WTP values and zero values. The link test and Ramsey RESET were used to assess the model specification and the model form. Where appropriate, to explore any potential relationships between those that provided zero values and those that did not, a univariable analysis with descriptive statistics and significance tests was carried out to explore the relationship between each individual exploratory variable and the WTP outcome.

The sensitivity analysis took the form of an interval regression or grouped data regression to analyse the predictors for WTP. An interval regression analysis has also been recommended for use when payment scale data are available (Donaldson et al., 1998). The log likelihood ratios and the significance of the model compared to the constant only model were assessed to determine the model fit.

Further as few respondents gave a protest or non-response in both Parts 1 and 2, it was not possible to conduct an analysis to identify the characteristics of the women who either gave a protest answer or did not respond to the WTP question.

7.3.3 Analysis of WTP and the Alternative Perspectives

Women's understanding of the WTP elicitation technique, the reasons given for WTP values, and whether they found the question difficult to answer were compared across the ex-ante and ex-post perspective. The findings of the qualitative responses were also used to assess overall whether WTP can be considered for use in menorrhagia. The number of protest responses and non-response to the WTP question was also compared, and whether the predictors of WTP, as measured in the econometric analysis, differ across perspectives.

In the ex-post perspective (Part 2) associations between WTP, EQ-5D and MMAS measures were assessed by Spearman's correlation analysis. The change in health state from baseline to current value (as measured by MMAS) was assessed against WTP for the change in health state and change in EQ-5D. The correlation coefficient was then compared to that of change in EQ-5D and change in MMAS. As discussed in Chapter 2, MMAS is scored on a 0-100 scale, where each level of each question dimension is given a weighting (Shaw et al., 1998). These weightings are then summed, according to the levels ticked by the respondent, to provide an overall score between 0-100. This was calculated for the baseline MMAS score, obtained from the ECLIPSE trial, and the current MMAS score. Similarly, the EQ-5D tariff was calculated using the relevant algorithm (Brooks, 1996). The change in scores from baseline to current value for both EQ-5D and MMAS was calculated by subtracting the current score from the baseline score.

7.3.4 Interview Analysis

In Part 3 the analysis of the interview findings is explored. The interviews were transcribed and analysed according to thematic analysis using Nvivo. The way in which the themes were derived follows the commonly used framework analysis method presented by Ritchie and Lewis (2003). First, descriptive themes that emerged in each interview were determined and used to generate the coding framework. Second, these multiple themes were placed into more general themes to generate an index. Each main theme and subtheme was denoted a number and these numbers were assigned to all of the text in the interview transcripts to indicate the theme discussed. In other words the interview transcripts were coded. The themes were collated and presented as descriptive accounts in the results chapter next.

7.3.5 The Development of Coding Reasons

In both the ex-ante and the ex-post questionnaire there were two qualitative questions. The first asked the respondent for reasons behind their WTP value and the second was related to why they did or did not find the WTP question difficult to answer. In both cases the data were analysed using content analysis (Bryman, 2012). For the first qualitative question, the qualitative reasons for the WTP value for both the ex-ante and the ex-post perspective were analysed by this author and categories generated based on a previous published WTP study (McIntosh et al., 2010). These codes were thought to be appropriate to apply to both the ex-ante and the ex-post perspective and the decision was verified by two other researchers (Tracy Roberts and Emma Frew).

The codes for the second qualitative question, regarding difficulty, were derived by a two-stage process. First, all three study researchers were given all of the ex-post qualitative responses for the question and were asked to independently identify themes. Second, the three sets of themes were brought together and discussed to produce one coherent list of themes. These themes were then applied, by all three researchers, to code the difficulty question within both the ex-ante and ex-post study, with only one additional theme within the ex-post study.

In the next section, a discussion regarding the reasons behind the methods chosen is provided.

7.4 Discussion of Methodology Used

This chapter reports the methodology behind the elicitation of WTP for the from the ex-ante and the ex-post perspective. As the most appropriate method to use to elicit WTP is not entirely clear, this section focuses on the discussion and justification of the methodology used in this research.

Overall, an out of pocket maximum monthly WTP for treatment was elicited in a questionnaire format using the payment scale. The ex-ante perspective recruited women who were attending any appointment at any Birmingham Women's Hospital outpatient clinic. The ex-post perspective involved women who were already a part of the ECLIPSE trial and had experience of menorrhagia and its treatments. The questionnaire was designed to be self-completed. Women from the ex-ante perspective were then asked if they would be willing to take part in an interview to gain a more in-depth understanding of their WTP answers.

7.4.1 Strengths and Limitations

The strength of the methodology used is that WTP is elicited from both the ex-ante and ex-post perspective to enable comparisons of the WTP results to be drawn across the two perspectives. As reported in Chapter 3, the ex-ante perspective is arguably preferred theoretically, and the ex-post perspective, is the most commonly used. Therefore, a comparison of the findings will determine the degree of difference in the results depending on the perspective. Also, the questionnaires were reviewed by experts in several relevant health related fields to determine their suitability for the respondents. Further, the inclusion of the qualitative questions regarding reasons behind WTP and whether the respondent had any difficulty with the WTP question does not only provide additional data to enable comparisons to be drawn between the two perspectives, but also provides an assessment of the validity of the WTP responses. Hence evidence is provided in favour or against the use of WTP and the questionnaire format in general in menorrhagia, and also provides evidence as to whether this format is more suitable for one perspective over the other, for valuing outcomes associated with menorrhagia and its treatments.

As this research is aimed at exploring whether WTP can be readily applied to value the interventions used for the treatment of menorrhagia, and a convenience sample of the ex-ante group could be obtained from the Birmingham Women's Hospital, it was decided that only women should be approached for WTP values. It would be difficult to recruit a sufficient number of men from this source. However, difficulties would also be expected when attempting to capture WTP values from men of the general public due to the nature of the condition, which may arguably be difficult for men to understand. Furthermore, the appropriateness of eliciting men's WTP values for such a sensitive area could be questioned as men are not at risk of the condition or at risk of consuming treatment, which are both requirements for eliciting ex-ante WTP values. There may be some cases where men are affected by partners with the condition. WTP values could have been elicited based on caring externalities, but this did not seem appropriate for this work. It could be argued that caring externalities have been accounted for to some degree as women that were approached knew family members and friends with the condition. Also, it may have been beneficial to pilot the questionnaire to identify the most suitable range of bid levels for the payment scale to minimise bias as the respondents WTP values may be influenced by the scale presented, but given the time constraints and the focus of the research to assess the feasibility of eliciting WTP the payment scale used was considered to be most suitable as the range is in line with likely levels of ability to pay.

Further, in the case of the ex-ante perspective, as menorrhagia is a subjective condition and women were recruited from the Birmingham Women's Hospital clinic, it was likely that women who: may have experienced menorrhagia in the past, have experienced the treatments but not menorrhagia, or have had a hysterectomy, were recruited into the sample, which does not strictly meet the ex-ante perspective definition as outlined in Chapter 3. However it is more representative of the female general population.

For the analysis of the ex-post perspective, it is recognised that due to the nature of the condition and its non-curative treatments that women in the ECLIPSE trial may have crossed-over and experienced both Mirena and Oral treatment, the numbers of which are reported in the results. It is recognised that there are limitations to the approach used of eliciting WTP for current treatment but this was deemed to be the most suitable method. Alternative methods such as eliciting WTP for the originally randomised treatment were considered. However, this would not be intuitive as women would be asked to reveal WTP for a less preferred treatment. It is important to take women who have crossed-over into account to ensure that WTP values do not solely reflect women who are happy with the treatment.

Finally, although calculations of sample size are typically conducted in studies assessing the effectiveness of treatments, i.e. in clinical trials, this is not always suitable in health economics. A method of obtaining the largest sample possible is often adopted, but much research in this area is currently being conducted. It is believed that, for the research carried out here, it is not readily possible to calculate a sample size as a-priori data on the distribution of WTP values for the Mirena or Oral treatment are not available (Donaldson et al., 1997b). That is, there are no data available to base the power calculation upon. It is necessary to identify the number and range of responses given to determine how many respondents are required to detect a certain difference in WTP across treatments. Furthermore, assumptions around the possible WTP range would also be unfeasible given the paucity of evidence on WTP effect sizes. Whilst information on what WTP value constitutes a meaningful difference in improvements from baseline or between treatments was not available for this research, this study could be used for future studies to calculate the required sample size.

7.4.2 Comparison with Other Research

While an interview environment is argued to be the recommended method to elicit WTP (Arrow et al., 1993), current recommendations for using interviews are taken from environmental economics where scenario descriptions are thought to require further clarification because respondents are not as knowledgeable of the goods. However, it has been argued that within healthcare the public may be more familiar with the goods and do not need such a rich context of the programme and the market to elicit meaningful WTP values as private payment is also taken for some interventions (Smith, 2003; Smith, 2006). Thus, as the healthcare interventions (Mirena and Oral treatment) and health outcomes are not complex and can be easily explained in a written scenario, a questionnaire was thought to be most suitable. There are also issues of the interviewing situation influencing the respondent, resulting in real WTP values not being elicited due to interviewer bias (McIntosh et al., 2010). Furthermore, interviews were deemed to be too resource intensive and costly, and difficult to recommend on a wider scale if WTP was found to be a suitable measure for menorrhagia.

Although the best method for eliciting WTP has not yet been established, the payment scale method was used because it has a higher completion rate than other methods that can be used in questionnaires, such as the open-ended questions (Whynes et al., 2003). Further the closed-ended question requires a-priori distribution of WTP and a large sample size (Frew et al., 2003). Research has shown that in the open-ended format protest responses and zero responses are more likely (Donaldson., 1997b; Frew et al., 2003; Reaves et al., 1999). Respondents are also more likely to try to estimate the cost of treatment when providing a WTP value because the format provides little guidance (Arrow et al., 1993; Donaldson, 1997b). The closed-ended format is prone to yea-saying and requires a-priori distribution of WTP in addition to a large sample size to obtain a good distribution of WTP answers (Diener et al., 1998).

7.5 Conclusion

In this chapter the methodology of the WTP feasibility study is reported. WTP was elicited in the form of a payment scale, presented in a questionnaire, from the ex-post and ex-ante perspective. WTP was elicited as an out of pocket monthly cost for LNG-IUS (Mirena) and usual medical treatment (Oral treatment). The methodology of the interviews using the ex-ante perspective respondents is also outlined. In the next chapter, the results of the three parts of research, the ex-ante perspective, the ex-post perspective, and the interviews are presented.

CHAPTER 8. WILLINGNESS TO PAY STUDY – RESULTS

8.1 Introduction

In this chapter the findings for the willingness-to-pay (WTP) empirical research, described in the previous chapter, are reported. The results are separated into three parts. Parts 1 and 2 refer to the study perspective, that is Part 1 (8.2) refers to the WTP feasibility study from the ex-ante perspective and Part 2 (8.3), the WTP feasibility study from the ex-post perspective. Part 3 (8.4), refers to the interviews which were carried out to identify patients understanding of WTP and its ease of use. Each part is followed by a discussion related to the more technical and specific issues related to the research presented. Parts 1 and 2 largely follow the same structure and report the following; questionnaire response, WTP for treatment, qualitative findings, the econometric analysis and demand curves.

8.2 Part 1: The Ex-ante Perspective

The results in this section refer to the ex-ante perspective and the analysis explained in the previous chapter (subsection 7.2.1). First the questionnaire response is reported, followed by the WTP for Mirena and Oral treatment and the qualitative findings related to reasons for the WTP values and whether the respondent found the questionnaire difficult to answer. The predictors of WTP for each treatment are then explored in an econometric analysis, the demand curves presented for WTP for each treatment, the intensity of respondents' preferences for treatment are assessed according to WTP values, and a summary of WTP with values using different subgroups is presented. Finally, a discussion of the findings from the ex-ante perspective is provided.

8.2.1 Questionnaire Response

As outlined in the methodology chapter (Chapter 7), in the ex-ante study, the WTP question was included in a self-complete questionnaire which was administered to women visiting the Birmingham Women's Hospital clinics. In this section, the response to the questionnaire, and specifically the WTP questions, is detailed. The questionnaire was administered over a 2 month period (Dec 2012 – Jan 2013). Overall 110 women completed and returned the questionnaire. The overall response rate for the ex-ante perspective was difficult to measure as every woman visiting the clinics, which were situated in different areas, was approached. Multiple women were given a patient information sheet followed by a questionnaire at the same time, by this author. Some women left the clinic after their appointment without completing the questionnaire or taking a stamped addressed return envelope. Approximately 142 questionnaires were given out to women who agreed to take part and 110 were returned, giving a 77% return rate. The number of women approached and given a patient information sheet can be calculated but with limited accuracy. Approximately 174 patient information sheets were handed out resulting in an estimated response rate of (110/174) 63%. However, it should be noted that this is an overestimation of the response rate as some women returned the patient information sheet when refusing to take part in the study and that information sheet was likely to have then been given to another potential respondent.

WTP values and corresponding qualitative answers were checked to identify non-response to the WTP question and protest answers (a breakdown of the qualitative information is provided in section 8.2.3). Out of the 110 women who completed the questionnaire, 3 (3%) women did not provide a WTP value for either Mirena or Oral treatment. Two additional women provided a WTP for one of the treatments only. One woman did not provide a WTP value for Mirena and another did not provide a WTP value for Oral treatment. Thus both Mirena and Oral treatment received the same number of non-response (4 in each). 7 protest answers, which relate to the individual misunderstanding the exercise and refusing to provide a WTP value, were identified from the qualitative explanations offered for the WTP value. These 7 protest answers and 4 non-responses to WTP questions were removed from the analysis. After these exclusions 99 respondents provided a WTP for each treatment, and in total 100 respondents provided a WTP for at least one treatment. A sensitivity analysis was carried out which included the protest answers.

Participants

The characteristics of the sample, excluding those who gave a protest answer or non-response, are presented in Table 8.1. The average expected age of menopause was slightly lower than the typically suggested average age of menopause for the population. Many of the women (80%) stated that they had experience of heavy menstrual bleeding. In some cases a respondent may have ticked that she experienced heavy menstrual bleeding but stated in her qualitative explanations that she had "no experience of heavy periods". The reason for this may be that at some time in their lives they may have had heavy periods but these periods may not have been consistently heavy over consecutive cycles. Whilst in other cases, where the respondent ticked that she did have experience of heavy periods, the respondent may have perceived that her bleeding was heavy through consecutive cycles and that it interfered with her quality of life (QoL), which would reflect experience of menorrhagia (heavy menstrual bleeding). The responses to the income categories showed that one half of the sample lie at the extremes of the income categories. One quarter have a household income of less than £20,000 and the other quarter have a household income greater than £50,000. The average household income in the UK lies between £20,000-£30,000, at £26,572 and approximately 65% of households have an income below the national average (Cribb et al., 2012). In this study reported here, a higher overall proportion (60%) of women have a household income of less than £30,000, which would suggest that the sample is fairly representative of the UK population in this respect.

Table 8.1 Sample characteristic (ex-ante)

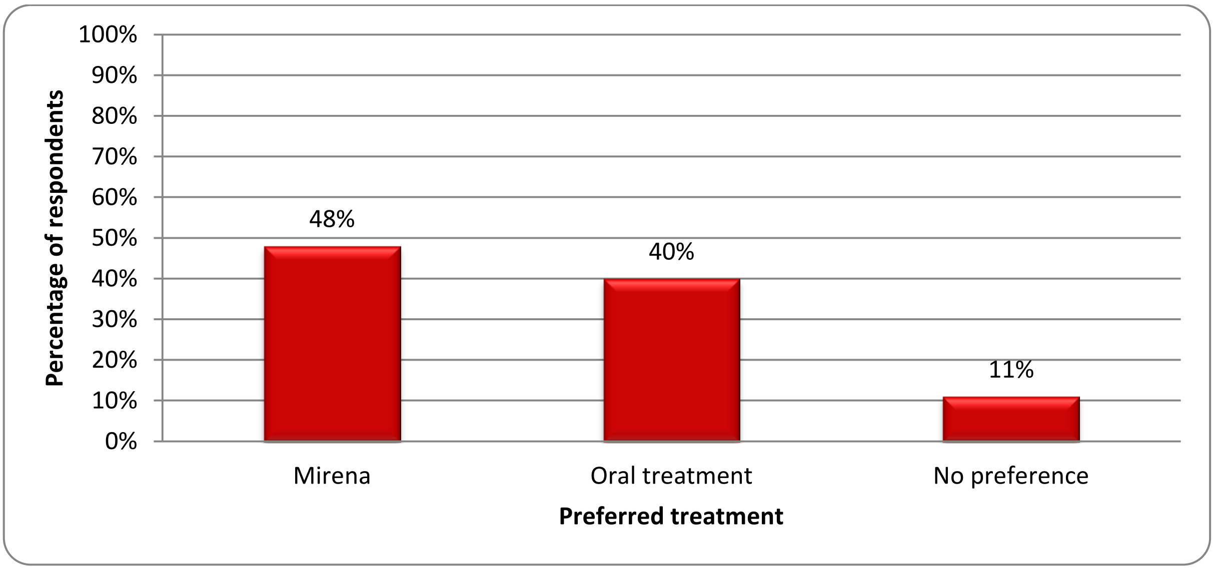
Variable	Sample analysed (n=100)
Expected age of menopause (yrs) [SD]	52.1 [5.53]
Age [SD]	36.6 [9.29]
Marital status	
Married or living with partner	63
Not	37
Wanted children in future?)	
Yes	55
No	34
Employment status	
Employed (FT)/(PT)	74
Not	26
Household income	
Less than 20,000	24
20,001-30,000	22
30,001-40,000	14
40,001-50,000	12
More than 50,000	23
Main earner	
Yes	45
No	53
Satisfaction with life overall [best (worst)]	
Very unsatisfied	13 /13
Slightly unsatisfied	15 /16
Neither satisfied nor dissatisfied	10 /11
Slightly satisfied	22 /21
Very satisfied	39 /38
Experience of menorrhagia	
Yes	80
No	20
EQ-5D best [SD]	0.766 [0.257]
EQ-5D worst [SD]	0.743 [0.283]

FT; full time, PT; part time

Preferred treatment group

Out of the 100 women who provided a WTP value for at least one treatment, Mirena was most preferred (47), followed by Oral treatment (39) and then no preference (11). Three women did not answer this question. The percentage of women preferring each treatment is presented in Figure 8.1.

Figure 8.1 Percentage of respondents preferring each treatment



3 women did not state whether they had a preference for treatment.

The socio-demographic characteristics of the study respondents were assessed to identify if there was a statistically significant difference between those who preferred Mirena to those who preferred Oral treatment. The Chi-squared test was used for categorical variables and either the two sample t-test or the Wilcoxon Rank Sum test for continuous variables, depending on the normality of the distribution. The results are presented in Table 8.2 and demonstrate that the only variable that does show a statistically significant difference between preference groups is ‘children in future’ ($p<0.05$). More people who preferred Oral treatment wanted children in the future. Oral treatment may have been preferred by women who wanted children in the future because the cessation of Oral treatment does not require a general practitioner (GP) appointment for removal. Unlike in the case of Mirena, women can just simply stop taking the Oral treatment when they please. However, as several significance tests were conducted, the possibility that this result may have occurred by chance cannot be excluded.

Table 8.2 Descriptive statistics according to preferred treatment

Variable	Prefer Mirena (n=47)	Prefer Oral (n=39)	p-value
Expected age of menopause (yrs) [SD]	52.43 [4.63]	51.5 [6.10]	0.564
Age [SD]	38.3 [9.49]	35 [9.54]	0.130
Marital status (%)			0.075
Married /living with partner	34 (72%)	21 (54%)	
Not	13(28%)	18 (46%)	
Wanted children in future? (%)			
Yes	17 (45%)	29 (76%)	
No	21 (55%)	9 (24%)	
Difference across preferred treatment			0.005*
Employment status (%)			
Employed (FT)/(PT)	33 (70%)	30 (77%)	
Not	14 (30%)	9 (23%)	
Difference across preferred treatment			0.484
Household income (%)			
Less than 20,000	7 (16%)	12 (32%)	0.077
20,001-30,000	10 (23%)	8 (22%)	0.931
30,001-40,000	7 (16%)	5 (14%)	0.782
40,001-50,000	8 (18%)	4 (11%)	0.367
More than 50,000	12 (27%)	8 (22%)	0.583
Main earner (%)			
Yes	21(47%)	17 (44%)	
No	24 (53%)	22 (56%)	
Difference across preferred treatment			0.778
Satisfaction with life [best (worst)] (%)			0.783 [0.841]
Very unsatisfied	8 (17%)	4 (11%)	0.367 [0.637]
Slightly unsatisfied	9 (19%)	6 (16%)	0.647 [0.887]
Neither satisfied nor dissatisfied	3 (6%)	4 (11%)	0.513 [0.513]
Slightly satisfied	9 (19%)	6 (16%)	0.647 [0.647]
Very satisfied	18 (38%)	18 (47%)	0.462 [0.619]
Experience of menorrhagia (%)			
Yes	34 (72%)	32 (82%)	
No	13 (28%)	7 (18%)	
Difference across preferred treatment			0.289
EQ-5D best [SD]	0.745 [0.285]	0.771 [0.229]	0.821
EQ-5D worst [SD}	0.734 [0.293]	0.740 [0.270]	0.727

*significant difference between preference groups (p<0.05)

8.2.2 WTP Mirena and Oral Treatment

The mean WTP for each treatment was calculated from the respondents who provided a WTP for at least one treatment. The maximum average WTP for Mirena along with the bootstrapped confidence intervals (CI) was £15.11 (95% CI £10.22-20) and for Oral treatment was £15.38 (95% CI 10.92-19.85). The results, presented in Table 8.3, show the descriptive statistics for WTP for Mirena and Oral treatment. It can be seen that the minimum and maximum values are the same for both treatments. For Mirena, 19 people were not willing to pay anything and two people were willing to pay £150. Whilst for Oral treatment, seven people were not willing to pay for Oral and one person was willing to pay £150. In both treatments the WTP data were found to be skewed and thus the WTP data for both Mirena and Oral treatment were log transformed. As detailed in Chapter 7, each WTP value has a constant of 1 added to it to ensure that zero values would still be taken into consideration when the data are log transformed.

Table 8.3 Descriptive statistics for WTP (ex-ante)

	WTP Mirena (n=99)	WTP Oral treatment (n=99)
Mean	£15.11	£15.38
Standard deviation (SD)	£24.50	£22.40
Min-Max	£0-£150	£0-£150
Inter-quartile range (IQR)	£4-£20	£6-£16
Group difference (paired t-test)	p=0.1247	

WTP; willingness-to-pay

To test for a statistical significance between the two log transformed WTP values for each treatment, a paired t-test was conducted. The results show that there is no statistically significant difference in mean WTP (p=0.1247) between the two treatments. Table 8.4 shows the results of mean WTP according to income which is often recommended when analysing WTP data as a theoretical validity check (Donaldson, 1999). It can be seen that generally the mean WTP increases as household income increases. However, this relationship is not monotonic as the mean WTP does decrease at some points as income increases (£30,001-£40,000). Hence WTP is not necessarily related to ability to pay as is required to confirm theoretical validity.

Table 8.4 Mean WTP against household income (ex-ante)

Household Income	WTP Mirena (n)	WTP Oral treatment (n)
Less than £20,000	£12.42 (24)	£12.52 (23)
£20,001 - £30,000	£14.90 (21)	£17.59 (22)
£30,001 - £40,000	£12.36 (14)	£14.43 (14)
£40,001 - £50,000	£24.17 (12)	£20.17 (12)
More than £50,000	£10.43 (23)	£10.26 (23)

WTP; willingness-to-pay

In Table 8.5 the WTP values for each treatment are presented against the number of respondents who chose that WTP value to explore the influence of prominent numbers on the WTP value. Prominent numbers are those that are typically selected by respondents and include 0, 1, 2, 5, 10, 20, 50, 100, 200, 500 and so on (Whynes et al., 2007).

Table 8.5 Frequency of WTP value (ex-ante)

WTP	Mirena (n)	Oral treatment (n)
£0	19	7
£2	3	5
£4	3	11
£6	8	8
£8	8	11
£10	22	23
£12	1	3
£14	1	3
£16	3	4
£18	4	2
£20	18	9
£25	2	2
£30	3	5
£45	0	1
£50	0	1
£100	2	3
£150	2	1
Total	99	99

The underlined numbers are the prominent numbers that were present in the payment scale

It can be seen that in the case of Mirena, the majority of the respondents have selected prominent numbers (£10 and £20). However, for Oral treatment a WTP value of £10 was selected the most number of times but there is a greater spread of values amongst non-prominent numbers compared to Mirena. The reasons for this difference in findings may be related to the nature of the two treatments and is explored further in the discussion (section 8.2.9).

As the question order of WTP for treatment was not randomly changed, Mirena was always asked first followed by Oral treatment. The effect on the WTP values of not randomly changing the order of the treatment payment scale first presented to respondents was also investigated. Hence the effect on the mean WTP of those who preferred Oral treatment providing a WTP for Mirena first, their least preferred treatment, is explored. Table 8.6 shows the mean WTP values for Mirena and Oral treatment according to preferred treatment.

Table 8.6 Assessment of ordering effects

Preferred treatment	WTP Mirena (mean)	WTP Oral (mean)
Prefer Mirena (n=47)	£20.30	£15.83
Prefer Oral treatment (n=39)	£10.39	£17.26

It can be seen from Table 8.6 that those who prefer Mirena are willing to pay more for Mirena than Oral treatment and vice versa for those who prefer Oral treatment. It also shows that despite those who prefer Oral treatment being presented with their least preferred treatment option first, the mean WTP value for their preferred option (Oral treatment) is greater. However, a greater difference between WTP values for the two treatments is produced for those who prefer Oral treatment, compared to those who were asked the WTP value for their most preferred treatment (Mirena) first. This finding could be due to the lack of randomisation of the question order, or that those who prefer Oral treatment placed a greater value on it compared to the value placed on Mirena, by those that prefer Mirena.

8.2.3 Qualitative Findings

Following the WTP questions for Mirena and Oral treatment, the women were asked to answer two qualitative questions. The first, related to the reason behind their WTP value and the second, asked women to explain why they did, or did not, find the WTP question difficult to answer. As outlined in the methodology Chapter (7), categories or 'themes' were generated using content analysis.

Reason for WTP

Nine categories of reasons for a WTP value were generated from the qualitative information from the full sample of women, which included protests and non-response. In total, 107 respondents offered an explanation for the WTP value. Only one code was required to categorise the written explanation in 61 cases. In 46 cases, two or more categories of explanations were required. The categories of reasons for the full sample and for the sample used in the analysis are presented in Table 8.7.

Table 8.7 Explanation given for WTP value (ex-ante)

Category	Explanation	Full sample n (%)	One WTP n (%)
R1	Subject expressed difficulty estimating WTP owing to: - Difficult to answer - Cannot put a price on healthcare	3 (2%)	1 (0.7%)
R2	WTP based on nominal amount - Arbitrary sum/ guess/ out of thin air	5 (3%)	5 (3%)
R3	WTP reflects ability to pay (affordability) - Maximum affordable amount given current situation	39 (24%)	39 (26%)
R4	WTP reflects reasonable value - NHS should pay but this is a reasonable limit	11 (7%)	11 (7%)
R5	WTP reflects cost of treatment - Attempted to estimate cost - Used a comparator such as prescription costs	30 (19%)	30 (20%)
R6	WTP reflects effects of treatment - In terms of effectiveness outcomes - In terms of process utility	54 (34%)	54 (36%)
R7	Protest expressed at idea of payment - Paid National Insurance and taxes/ NHS should pay - Women cannot help the condition/ treatment should be free	6 (4%)	0 (0%)
R8	Related to cost of sanitary wear - Washing clothes/wipes/painkillers	9 (6%)	9 (6%)
R9	Misunderstood exercise but provided WTP value	3 (2%)	3 (2%)
Total		160	152

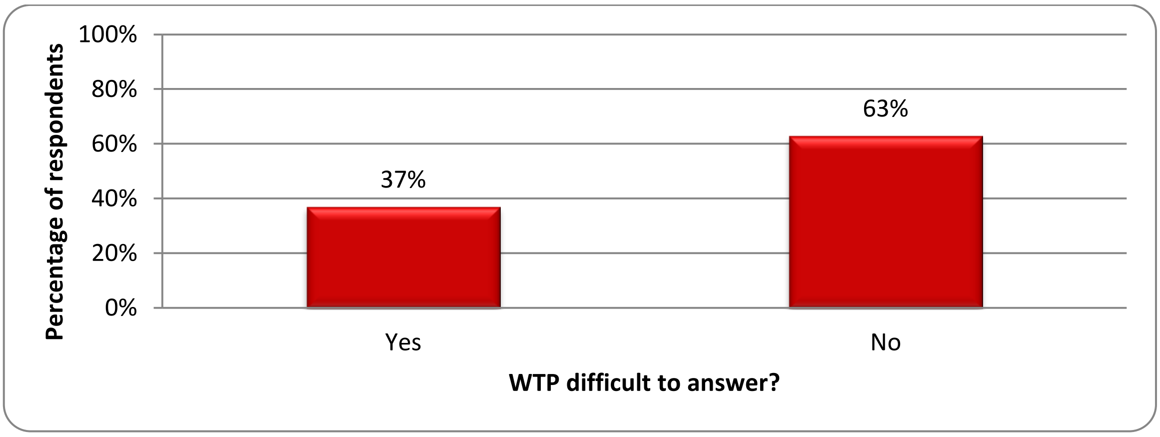
NHS; National Health Service, WTP; willingness-to-pay

Among the respondents that offered at least one WTP value, all but one offered a reason. The most commonly cited reason for a WTP value for both the full sample and the sample analysed was 'R6 where WTP reflects the effect of treatment', followed by 'R3: affordability'. Three respondents misunderstood (R9) the WTP question. As it could be argued that these WTP values are not valid, a sensitivity analysis was conducted and is reported in section 8.2.8 to estimate the mean WTP excluding these three respondents.

Difficulty with WTP explanation

Over 60% of women who completed at least one WTP question said that the question was not difficult to answer (Figure 8.2).

Figure 8.2 Proportion of women who did and did not find WTP difficult to answer



Respondents were then asked to provide a written explanation for their answer. One respondent did not complete the previous question on whether she found WTP difficult to answer but did provide an explanation for whether she found the question difficult to answer. Nine categories of explanation were generated from the qualitative information from the entire sample of women, including protests and non-response. Out of the 110 women, 92 provided a reason for their answer and 16 of these reasons were coded as one category. 76 provided an explanation that was coded into two or more categories. Among the 98 women who provided a WTP value for both Mirena and Oral treatment, or a WTP for at least one treatment, 86 provided an explanation as to why they did or did not find the question difficult to answer. The categories of explanations for the sample analysed (which excludes those who did not provide a protest answer) are presented according to whether the question was difficult to answer in Table 8.8 below.

Table 8.8 Explanation given for difficulty question (sample analysed, ex-ante)

Category	Explanation	Difficult n (%)	Not difficult n(%)	Total n (%)
D1	Found valuation difficult - Not used to it/ difficult to quantify/WTP out of thin air - Not aware of cost implications	19 (44%)*	1 (2%)	20 (20%)
D2	Used prescription costs as a proxy - "I know the cost of prescription..."	1 (2%)	5 (8%)	6 (6%)
D3	Benefits of treatment – reasonable amount for expected outcomes - Due to way menorrhagia impacts lives - Due to nature of treatment effects	8 (19%)	23 (39%)	31 (30%)
D4	Ability to pay dictated amount	4 (9%)	8 (14%)	12 (12%)
D5	Difficulty with hypothetical nature of WTP question - No personal experience of menorrhagia	7 (16%)	0 (0%)	7 (7%)
D6	Balance of impact on QoL and affordability - Explicitly explained WTP not difficult to answer	2 (5%)	20 (34%)	22 (22%)
D7	Protest – I would not pay	0 (0%)	0 (0%)*	0 (0%)
D8	Misunderstood exercise but provided WTP value	2 (5%)	1 (2%)	3 (3%)
D9	WTP reflects reasonable value - NHS should pay but this is a reasonable limit	0 (0%)	1 (2%)	1 (1%)
Total		43	59	102

*5 additional reasons were cited in the entire sample (protests and non-response were considered). Three protested again (D7) to the WTP question and considered the question not to be difficult, two were 'D1: not used to valuing healthcare' and said the question was difficult.

In the case where the respondents completed one WTP question it can be seen that for those who did find the question difficult to answer, the most common reason was related to 'D1: not being used to valuing healthcare'. For those who did not find the valuation difficult the most commonly cited reason was 'D3: a reasonable amount to pay for the expected benefits'. There were two cases where inconsistent responses were observed between the answer provided for the WTP question and the explanation given. That is, one respondent stated that the question was not difficult to answer and then in the difficulty reason stated that it was difficult. The other respondent ticked that it was difficult and explicitly stated that it was not difficult in the qualitative explanation. This inconsistency in responses will be discussed further in section 8.2.10.

8.2.4 Econometric Analysis

It was outlined in the analysis section of the methodology chapter that a two-part model could not be used to assess WTP as the proportion of respondents who provided zero values was quite small (details next). To explore any potential relationships between those that provided zero values and those that did not, a univariable analysis was carried out. Further, as the number of respondents who gave a protest (n=7) or non-response (n=4 in each arm) was too few, it was also not appropriate to conduct an analysis to identify the characteristics of the women who either gave a protest answer or did not respond to the WTP question.

In the following section, the results of the analysis on the zero values is reported first and is followed by the OLS linear regression on the sample, which includes the zero values, but not protests and non-response. The exclusion of these respondents is further explored in a sensitivity analysis. As part of a sensitivity analysis, an interval regression was also run on Mirena and Oral treatment (See Appendix A5.1).

Analysis of Zero WTP values

As there is insufficient variation in the sample, a regression analysis of the 19% and 7% of respondents who provided a zero WTP value for both Mirena and Oral treatment respectively cannot be conducted to identify predictors for providing a zero value. Only 3 respondents gave a zero value for Oral treatment alone, and 15 respondents gave a zero value for Mirena alone (4 gave zero values for both treatments). Table 8.9 presents the descriptive statistics of the respondents who provided a zero value for each treatment. Significance tests using Chi-squared test for categorical variables and Wilcoxon Rank Sum for continuous variables were carried out to identify if there were any significant differences between those who provided zero values for either treatment and those that did not. The significance tests showed that respondents who provided zero values are significantly less likely to have experience of menorrhagia (p=0.040), earn £20,000-30,000 (p=0.015), earn £40,000-50,000 (p=0.050) and a lower EQ-5Dbest score (p=0.050). There are insufficient numbers in each treatment group to obtain any robust significance test results between those who provided zero values for each treatment, and those that did not.

Table 8.9 Descriptive statistics for zero WTP values (ex-ante)

Variable	Mirena Zero (n=19)	Oral Zero (n=7)
Expected age of menopause (yrs)[SD]	53.3 [3.56]	54 [2.94]
Age (%)	35.4 [6.98]	33.3 [8.81]
Marital status (%)		
Married or living with partner	13 (68%)	5 (71%)
Not	6 (32%)	2 (29%)
Wanted children in future? (%)		
Yes	13 (72%)	5 (83%)
No	5 (28%)	1 (17%)
Employment status (%)		
Employed (FT)/(PT)	13 (68%)	5 (71%)
Not	6 (32%)	2 (29%)
Household income (%)		
Less than 20,000	7 (39%)	2 (33%)
20,001-30,000	6 (33%)	3 (50%)
30,001-40,000	2 (11%)	0 (0%)
40,001-50,000	0 (0%)	0 (0%)
More than 50,000	3 (17%)	1 (17%)
Main earner (%)		
Yes	9 (47%)	2 (29%)
No	10 (53%)	5 (71%)
Satisfaction with life overall [best (worst)] (%)		
Very unsatisfied	4 (21%) / 4 (21%)	1 (14%) / 1 (14%)
Slightly unsatisfied	3 (16%) / 4 (21%)	2 (29%) / 2 (29%)
Neither satisfied nor dissatisfied	4 (21%) / 4 (21%)	1 (14%) / 1 (14%)
Slightly satisfied	4 (21%) / 4 (21%)	1 (14%) / 1 (14%)
Very satisfied	4 (21%) / 3 (16%)	2 (29%) / 2 (29%)
Experience of menorrhagia (%)		
Yes	19 (100%)	6 (86%)
No	0 (0%)	1 (14%)
EQ-5D best [SD]	0.693 [0.236]	0.642 [0.317]
EQ-5D worst [SD]	0.665 [0.274]	0.642 [0.317]

SD; standard deviation

Figure 8.3 presents the preference groups against the percentage of zero values. The figure shows that in the case of zero values for Mirena, intuitively, the respondents who were most likely to provide a zero value for Mirena preferred the Oral treatment (78%). However, for the women who preferred Mirena, they were not more likely to provide a zero value for Oral treatment, as one would expect. It was observed that it was still the Oral treatment preference group that provided the most zero values for Oral treatment (43%).

Figure 8.3 Zero values provided against preferred treatment

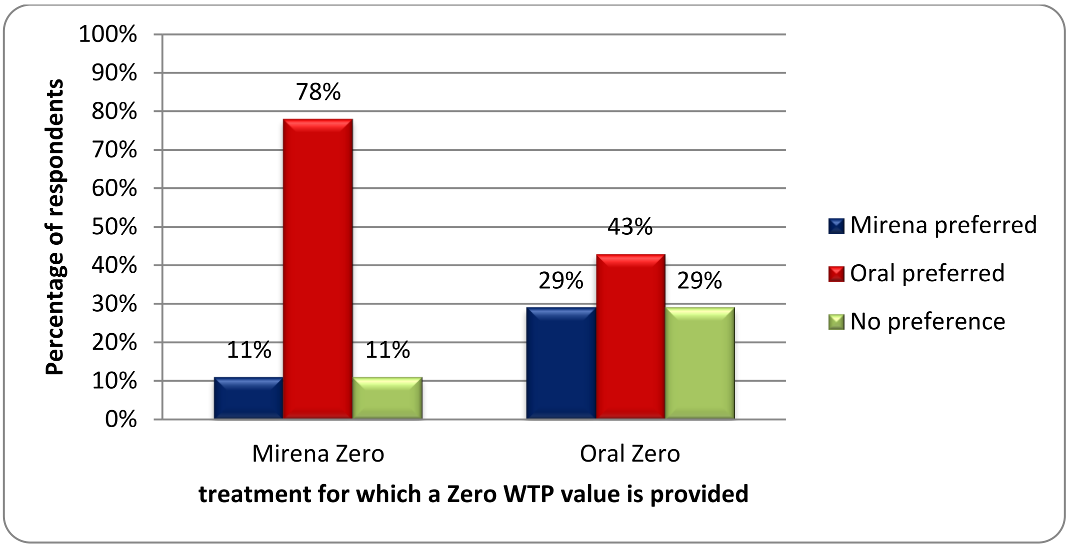
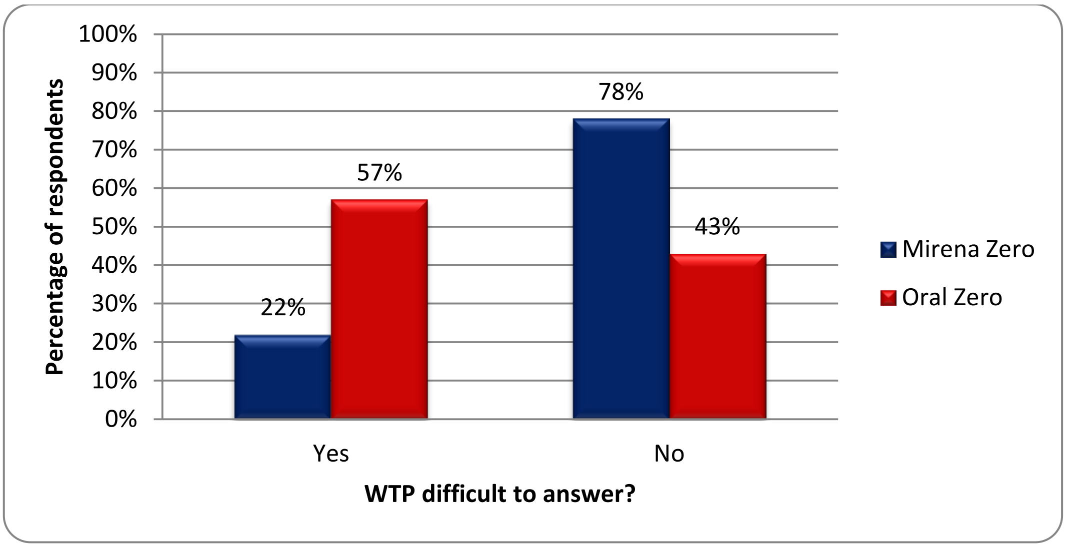


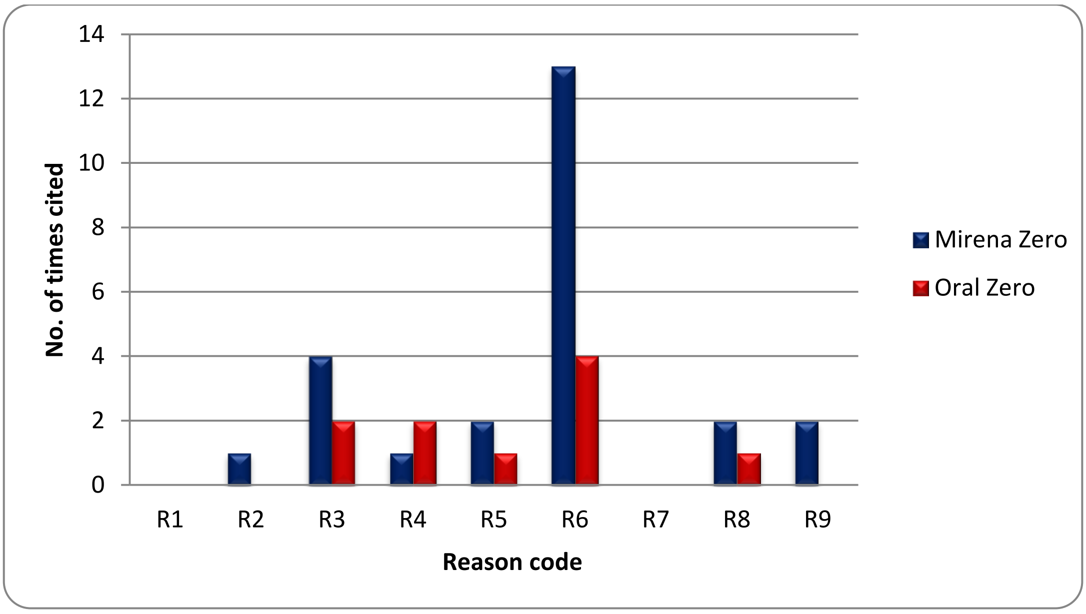
Figure 8.4 illustrates that out of those who found the WTP question difficult to answer, a higher percentage of respondents provided a zero value for Oral treatment. Conversely out of those who <u>did not</u> find the question difficult to answer, a higher percentage of respondents provided a zero value for Mirena.

Figure 8.4 Difficulty with WTP according to zero value



To explore the reason for these zero values Figure 8.5 illustrates the reason codes cited by those who provided a zero WTP value for Mirena or Oral treatment.

Figure 8.5 Explanation given for zero value



R1: difficulty ; R2: nominal amount; R3: ability to pay; R4: reasonable value; R5: cost fo treatment; R6: effects of treatment; R7: protest; R8: cost of sanitary wear; R9: misunderstood

It can be seen that for those who provided a zero value for Mirena, the most commonly cited reason was ‘R6: the effect of the treatment’. As all 19 women who gave a zero value for Mirena had experience of heavy periods (Table 8.9), it is likely that they have experience of Mirena and therefore that these women did not think that the effect of Mirena was worth any monetary amount. The second most common reason for providing a zero value for Mirena was ‘R3: affordability’. For those that provided a zero value for Oral treatment, similarly to Mirena, ‘R6: effect of treatment’ was the most commonly cited reason for providing a zero value followed by both ‘R3: affordability’ and ‘R4: A reasonable value but the NHS should pay’. Similarly, as all but one of the respondents who provided a zero value for Oral treatment had experience of menorrhagia and by extension was likely to have experience of Oral treatment, they may not have believed that Oral treatment was worth any monetary amount.

OLS linear regression

The results of the stepwise OLS linear regression on the log transformed data are presented in Table 8.10. Two variables were excluded - one was excluded to preserve the sample size, the variable ‘children in future’ was dropped from the analysis as this variable contained 10% missing data. Due to the nature of the stepwise analysis to keep the sample size constant throughout the analysis even if ‘children in future’ was not significant in the final model, the sample size would remain at the reduced level. Thus a univariable analysis was conducted using ‘children in future’ for both treatment options and in both cases it was found that this variable was not significant at the 30% level. A flexible value of $p<0.3$ was used at the univariable stage to ensure that the variable was not excluded unnecessarily. The ‘satisfaction with life’ variables were dropped to prevent collinearity. ‘Satisfaction’ was also assessed in a univariable analysis and was not found to be significant at $p<0.3$.

It can be seen that in the case of Mirena that WTP is positively influenced by income. Whilst the WTP responses from the lower income categories are not significant, the ranges of the confidence intervals show that there is a tendency for an increasing positive association with WTP as income increases. However, this relationship is not monotonic as there is a reduction in the association with WTP at the highest income, but this also did not reach significance. The only significant income category was £40,001-£50,000 thus compared to those in the lowest household income bracket, those earning £40,001-£50,000 are willing to pay significantly more for Mirena ($p=0.005$; $p<0.05$). It can be seen in Table 8.10 that ‘R9:

misunderstanding the exercise' ($p=0.020$), 'preferring Oral treatment' ($p=0.002$), having 'experience of heavy menstrual bleeding' ($p=0.019$) and being 'married' ($p=0.008$) all have a significantly negative impact on WTP ($p<0.05$). It could be intuitively expected that preferring Oral treatment would lead to a negative impact on WTP for Mirena. It is also possible that women who suffer from menorrhagia may have already experienced Mirena and would not pay for the treatment effects. Further as Mirena causes temporary negative effects on women for the first 6 months, as described in the treatment scenario, it could be that women who are married would prefer a more immediate effect to minimise the strain on the relationship. In contrast 'R3: affordability' ($p=0.010$), 'R5: the cost of treatment' ($p=0.023$), and current 'EQ-5Dbest' ($p=0.004$) have a significantly positive effect on WTP for Mirena ($p<0.05$). The reason code 'R1: finding the valuation difficult' had a tendency towards significance ($p=0.063$) but the confidence intervals suggest there could be either a positive or negative relationship with WTP.

The link test, used to assess the model form and model specification, revealed that the model is correctly specified ($p=0.095$). Thus the null hypothesis, that an important variable has not been excluded, cannot be rejected. Despite the significant p-value observed in the Ramsey RESET test ($p=0.0483$), indicating that variables were omitted, as the p-value is only bordering on significance and the link test revealed that the model is correctly specified it is possible that this result has occurred by chance.

Table 8.10 OLS linear regression for Mirena WTP values (ex-ante)

Variable	Coefficient	95% CI	P-value
Expected age of menopause	-0.0626	-0.100, -0.026	0.001
R3: affordability	0.598	0.148, 1.047	0.010
R5: cost of treatment	0.563	0.081, 1.044	0.023
R9: Misunderstanding exercise	-1.381	-2.536, -0.226	0.020
Experience of menorrhagia	-0.623	-1.142, -0.104	0.019
Preferred treatment			
Prefer Oral treatment	-0.718	-1.153, -0.283	0.002
No preference	0.0546	-0.653, 0.762	0.878
Income			
£20,001-£30,000	-0.114	-0.674, 0.446	0.687
£30,001 -£40,000	0.254	-0.456, 0.964	0.479
£40,001-£50,000	1.041	0.320, 1.762	0.005
More than £50,000	0.0803	-0.505, 0.666	0.785
Married	-0.606	-1.064, -0.165	0.008
EQ-5Dbest	1.286	0.427, 2.145	0.004
R1: Found valuation difficult	1.999	-0.155, 4.113	0.063
Constant	5.015	2.898, 7.133	0.000

Adjusted R^2=0.394, n=93

In the case of the linear regression model for Oral treatment (Table 8.11) it can be seen that none of the variables for Mirena arose as significant predictors for WTP for Oral treatment. Overall, finding the question ‘difficult to answer’ (p=0.003) has a significantly negative impact on WTP for Oral treatment ($p<0.05$), but having ‘D5: difficulty with the hypothetical nature’ (p=0.015), ‘D1: not being used to valuing healthcare’ (p=0.002) and using ‘R8: cost of sanitary wear’ (p=0.044) as a proxy for WTP were positively associated with WTP ($p<0.05$). The following variables had a tendency towards significance; ‘D3: reasonable amount for the expected outcomes’ (p=0.06), ‘R9: misunderstanding the exercise’ (p=0.062) and ‘R3: affordability’ (p=0.078) but the confidence intervals include both negative and positive associations. Similar to the Mirena model, the link test revealed that the model is correctly specified (p=0.688). Thus it is not possible to reject the null hypothesis that an important variable has not been excluded and the Ramsey RESET test also shows that there are no omitted variables (p=0.8164). However since the R^2 for this regression is low (R^2=0.111), this could indicate that WTP is also predicted by something else that is not captured by any of the explanatory variables in this model. However, caution of interpretation of these models is necessary given the relatively small sample size.

Table 8.11 OLS linear regression for Oral treatment WTP values (ex-ante)

Variable	Coefficient	95% CI	P-value
R9: Misunderstood exercise	1.039	-0.0535, 2.132	0.062
Difficult to answer	-0.860	-1.430, -0.291	0.003
D5: Difficulty with hypothetical	1.111	0.217, -2.005	0.015
R3: affordability	0.364	-0.0412, 0.769	0.078
D1: Not used to valuing	1.021	0.378, 1.664	0.002
D3: Reasonable amount for treatment effects	0.399	-0.016, 0.814	0.06
R8: Cost of sanitary wear	0.682	0.018, 1.345	0.044
Constant	1.947	1.615, 2.279	0.000

Adjusted R^2=0.111, n=93

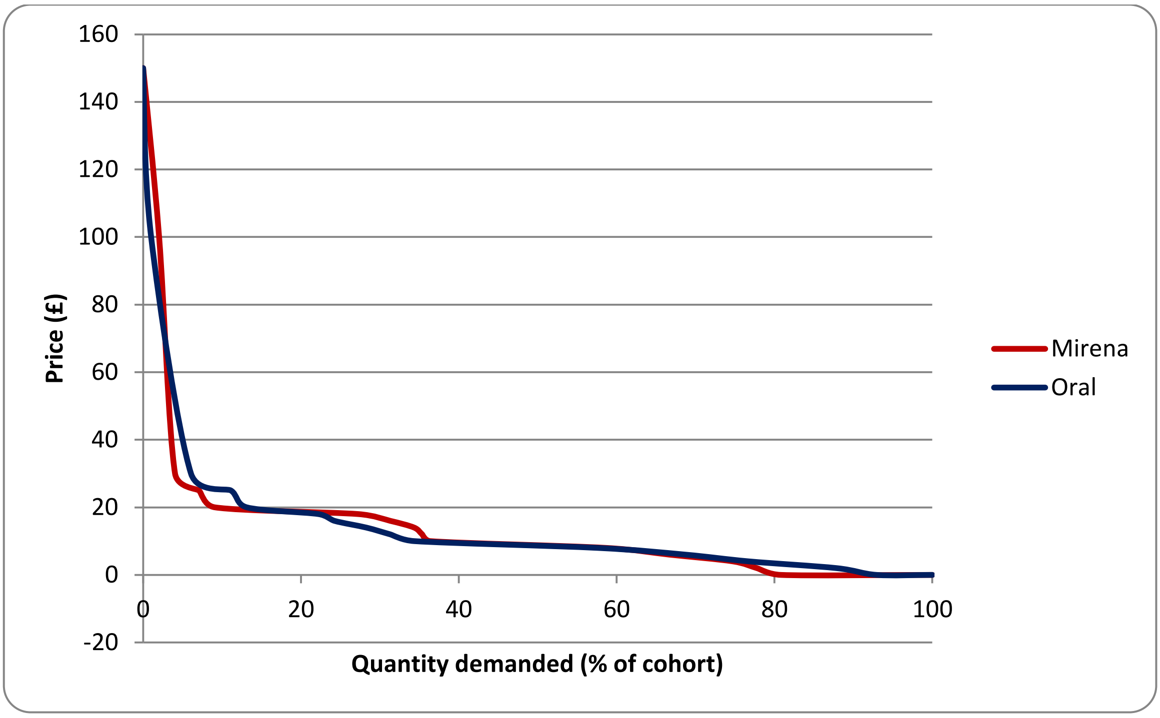
The results of the sensitivity analysis using interval regression for both Mirena and Oral treatment can be found in the Appendix 5.1. The model for Mirena is similar to the OLS model except 'R9: Misunderstanding the exercise' is included as a significant variable in the linear OLS model and is not in the interval regression. Also 'R5: cost of treatment' is not significant in the interval regression but is in the OLS model. The signs of the coefficients for the variables are the same in both the OLS and interval regression models and the size of the coefficients differ only slightly. The results of the interval regression for Oral treatment differ from the OLS model in that 'income', 'D3: reasonable amount for expected outcomes' and 'R3: affordability' are significant for interval regression but 'D3: reasonable amount for expected outcomes' and 'R3: affordability' only have a tendency towards significance in the OLS model. The signs of the coefficients of the OLS and interval regression models are the same, and the coefficients vary slightly. The interval regression model for both Mirena and

Oral treatment was shown to be statistically significant compared to the constant only model (likelihood ratio 48.64, p=0.000) (likelihood ratio 31.92, p=0.008).

8.2.5 Demand Curves

Figure 8.6 shows the demand curves for Mirena and Oral treatment. A demand curve represents the quantity demanded of a good according to changes in price. In a competitive market it is expected that as the price of the good decreases, the demand increases, thus producing a downward sloping curve (Sloman & Wride, 2009). When using contingent valuation, a hypothetical market is created which is contingent on the scenario provided. Thus it is expected that the demand curves in this WTP study behave as those in competitive markets. The demand curves for Mirena and Oral treatment resemble a 'typical' demand curve. It can be seen that from £60 upwards the demand curves for both treatments are inelastic as great changes in price (£150-£60) result in very small changes in quantity demanded. However, at lower prices, below £20, both treatments are elastic, as below the £20 price demand changes greatly from 5% to 80% of the respondents.

Figure 8.6 Demand curves for Mirena and Oral treatment



8.2.6 Preference Reversals and Inconsistent Responses

Table 8.12 presents the preferred treatment against the greater WTP value. It shows that in 39% of cases, the respondent has stated that they have a preferred treatment but gave an equal WTP to both treatments. Nine respondents who preferred Oral treatment gave a greater WTP value to Mirena. One respondent who preferred Mirena gave a WTP that was higher for Oral treatment. Seven respondents who had no preference did not give equal WTP values, but gave a greater value to Mirena or Oral treatment.

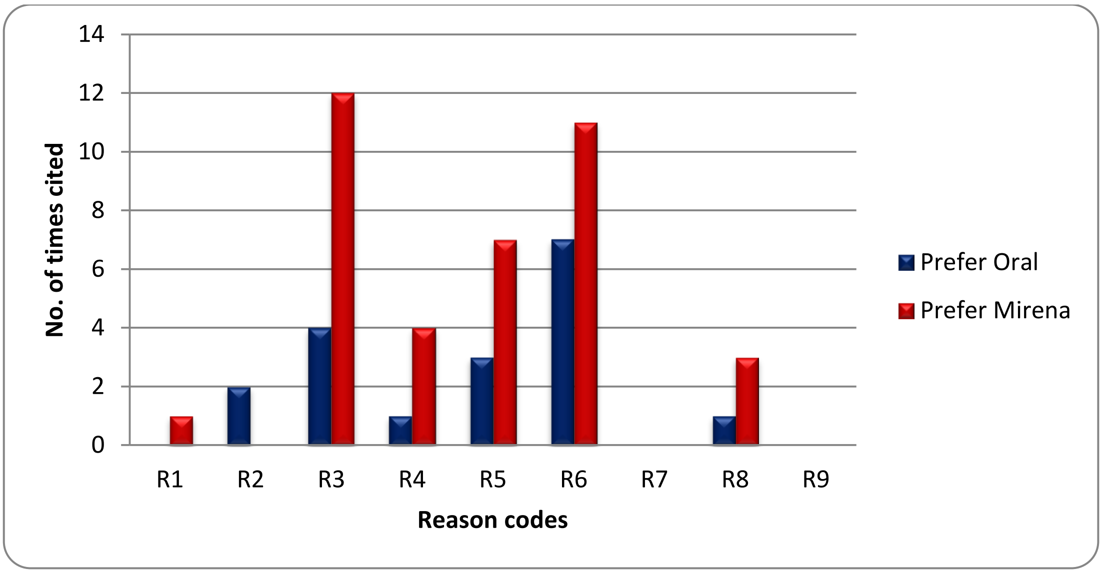
Table 8.12 Preferred treatment against greatest WTP value provided

Preferred treatment	**Is WTP Oral > WTP Mirena**			
	No (%)	Equal (%)	Yes (%)	Total
Oral	9 (9%)	13 (13%)	17 (18%)	39 (40%)
Mirena	21 (22%)	25 (26%)	1 (1%)	47 (48%)
No preference	5 (5%)	4 (4%)	2 (2%)	11 (11%)
Total	35 (36%)	42 (43%)	20 (21%)	97

WTP; willingness-to-pay

The qualitative reasons for the respondents who provided equal WTP values even though a preference for treatment is stated are displayed in Figure 8.7. The figure shows that the two most commonly cited reasons for equal WTP values are related to 'R6: the effect of treatment' and 'R3: affordability' for both Oral treatment and Mirena. However, it should also be noted that it is possible that the respondent has a preference for treatment, but this preference may not be particularly strong, which could result in equal WTP values (Ryan & San Miguel, 2000). Therefore these 39% of cases do not need to be treated as inconsistent responses and are therefore included in the sample.

Figure 8.7 Reason for equal WTP value according to preferred treatment



R1: difficulty ; R2: nominal amount; R3: ability to pay; R4: reasonable value; R5: cost of treatment; R6: effects of treatment; R7: protest; R8: cost of sanitary wear; R9: misunderstood

The qualitative reasons for preference reversals, where a respondent has stated that they prefer one treatment but provide a greater WTP value for the other treatment, were also checked to identify why these individuals provided inconsistent responses. It can be seen from Table 8.13 that the most commonly cited reason that respondents who preferred Oral treatment gave a greater WTP value for Mirena was related to 'R5: the cost of the treatment'. Thus, these respondents expected the Mirena to cost more than Oral treatment and therefore gave a greater WTP. Similarly, the only reasons given why respondents who preferred Mirena gave a greater WTP for Oral treatment includes 'R5: the cost of treatment' and 'R4 providing a reasonable WTP value for treatment'.

Table 8.13 Reasons for inconsistent WTP values

Category	Explanation	Prefer Oral, > Mirena WTP	Prefer Mirena > Oral WTP
R1	Subject expressed difficulty estimating WTP owing to: - Difficult to answer - Cannot put a price on healthcare	0 (0%)	0 (0%)
R2	WTP based on nominal amount - Arbitrary sum/ guess/ out of thin air	0 (0%)	0 (0%)
R3	WTP reflects ability to pay (affordability) - Maximum affordable amount given current situation	3 (30%)	0 (0%)
R4	WTP reflects reasonable value - NHS should pay but this is a reasonable limit	2 (20%)	1 (50%)
R5	WTP reflects cost of treatment - Attempted to estimate cost - Used a comparator such as prescription costs	4 (40%)	1 (50%)
R6	WTP reflects effects of treatment - In terms of effectiveness outcomes - In terms of process utility	0 (0%)	0 (0%)
R7	Protest expressed at idea of payment - Paid National Insurance and taxes/ NHS should pay - Women cannot help the condition/ treatment should be free	0 (0%)	0 (0%)
R8	Related to cost of sanitary wear - Washing clothes/wipes/painkillers	1 (10%)	0 (0%)
R9	Misunderstood exercise but provided WTP value	0 (0%)	0 (0%)
Total		10	2

NHS; National Health Service, WTP; willingness-to-pay.

8.2.7 Intensity of Preferences

The intensity of the preferences were assessed to identify why the WTP for Mirena (£15.11) was not greater than the WTP for Oral treatment (£15.38) given that Mirena was the most preferred treatment in the sample (48%).

The findings in Table 8.14 show that overall the WTP for Oral treatment is 2% greater than the WTP for Mirena. This finding contradicts the results of the most preferred treatment as Mirena was the most preferred in the sample (48%). Those that preferred Mirena were willing to pay 28% more for Mirena than Oral treatment, which is as theory would predict. Those that preferred Oral treatment were willing to pay 66% more for Oral treatment. This finding could indicate that those who preferred Oral treatment had a stronger preference for Oral treatment than Mirena, which may explain the greater overall WTP value for Oral treatment.

Table 8.14 Intensity of preferences

Preferred treatment	WTP ratio
All cases	WTP Oral/ WTP Mirena = 1.02
Mirena	WTP Mirena/ WTP Oral = 1.28[a]
Oral treatment	WTP Oral/ WTP Mirena= 1.66
No preference	WTP Mirena/ WTP Oral = 1.19[a]

[a] To identify the intensity of preferences for Mirena and No preference it is more appropriate to divide Mirena by Oral treatment, rather than the reverse ('Mirena' -WTP Oral/ WTP Mirena =0.78/ 'No preference' WTP Oral/WTP Mirena = 0.86)

However, those respondents that had no preference for treatment were willing to pay 19% more for Mirena than Oral treatment. The results of the qualitative reasons for WTP values were checked to determine the reason for providing a greater WTP for Mirena. All 5 respondents who stated that they had no preference but gave a greater WTP value for Mirena provided at least one reason for their WTP values. 'R4: WTP reflects a reasonable value but the NHS should pay', 'R5: WTP reflects the cost of treatment' and 'R6: WTP reflects effects of treatments' were each cited twice and 'R3: affordability' was cited once. The effects of treatments impacting on WTP may be related to Mirena having better long-term effects than Oral treatment. The perceived greater cost of Mirena has also led to greater WTP values for

Mirena and it seems from R4 that a 'Reasonable value for treatment, but the NHS should pay' may also be related to the outcomes for the two treatments presented in the scenario description.

8.2.8 Summary of WTP with Excluded Subgroups

Table 8.15 details the WTP results according to the individual subgroups that were excluded, to identify the impact of excluding these responses on WTP.

Table 8.15 WTP according to excluded subgroups (ex-ante)

Analysis	Mean [SD]	Median	No. of Obs
Mirena			
Sample analysed	£15.11 [24.54]	£10	99
Without misunderstood	£15.60 [25.03]	£10	94
Without equal WTP	£15.49 [25.32]	£10	92
Preference reversals	£15.27 [24.62]	£10	98
Full sample	£14.11 [24.00]	£10	106
Oral treatment			
Sample analysed	£15.38 [22.39]	£10	99
Without misunderstood	£14.69 [21.17]	£10	94
Without equal WTP	£15.99 [23.06]	£10	92
Preference reversals	£16.01 [23.36]	£10	90
Full sample	£14.37 [21.97]	£10	106

SD; standard deviation, WTP; willingness-to-pay.

It can be seen from Table 8.15 that the only subgroup that alters the findings, that Oral treatment generates a greater WTP than Mirena, is the exclusion of those who misunderstood the exercise. The exclusion of protests, preference reversal, those who provided equal WTP values despite having a preferred treatment changes the mean WTP very slightly. In the case of the 5 respondents who misunderstood the exercise the mean WTP for Mirena is 91 pence greater than Oral treatment.

8.2.9 Discussion

In this first part of the chapter the results of the WTP feasibility study from an ex-ante perspective were presented.

The results showed that Mirena was the preferred treatment but that the mean WTP for Oral treatment (£15.38) was slightly greater than the mean WTP for Mirena (£15.11). However this difference was not statistically significant. The reasons for the wide confidence intervals for both treatments and the subsequent lack of statistical significance are likely to be related to the nature of the payment scale. As a full numerical scale is not used and it was shown that respondents tended to value treatments within the £10-£20 range, there are very few numbers that respondents can select between those values, which could have then led to overlap in confidence intervals. Furthermore, women may see the two treatments as something they may want to try at different times in their life, or may consider the other treatment as the next option if the first did not improve QoL. Therefore they could be more indifferent about the treatments.

Those women who preferred Oral treatment were willing to pay 66% more for Oral treatment than Mirena, compared to those who preferred Mirena, who were willing to pay 28% more for Mirena than Oral treatment. Thus it is possible that Mirena did not generate a greater mean WTP because those who preferred Oral treatment had a stronger preference for Oral treatment than those who preferred Mirena. Alternatively, it could be argued that an embedding effect may have been observed, where the respondents may have failed to distinguish between the two treatments and were simply providing a WTP for any treatment (Kahnemann & Knetsch, 1992).

The two most commonly cited reasons for WTP values were related to the 'effects of the treatments' and 'affordability'. Therefore in this study, most respondents valued the treatments according to their associated process utility and outcomes, as is required to elicit valid WTP values. It is expected that 'affordability', i.e. ability to pay, would be a commonly cited consideration for providing a WTP value (Donaldson, 1999). The mean WTP results presented against income groups demonstrated that generally mean WTP does increase as income increases but that the relationship is not monotonic. Thus, this study shows that it may not necessarily be the case that those who have a higher household income will be willing to pay more for treatment. This would suggest that in this case, resource allocation would not be skewed towards the rich, as is commonly suggested in the literature as a downfall of the welfarist approach (Coast, 2004; Gold et al., 1996). However, the lack of a monotonic relationship between income and WTP is said to suggest that theoretical validity is not demonstrated.

Overall women did not consider the WTP question to be difficult to answer. The reasons provided by those respondents that did not find the question difficult to answer were related to providing a reasonable value for the expected outcomes. Whilst those who did find the question difficult cited not being used to valuing healthcare, which is not an unexpected finding given that generally UK citizens do not pay at the point of consumption of healthcare.

The qualitative reasons were found to be important predictors of WTP for Oral treatment, but not the socio-demographic variables, whilst predictors of WTP for Mirena included both qualitative reasons and socio-demographic factors. The results of the sensitivity analysis using interval regression varied only slightly. The removal of various subgroups of WTP values; including protests and inconsistent responses did not lead to a difference in which treatment generates the greatest WTP, except in the case of misunderstanding the exercise as Mirena had a slightly higher mean WTP than Oral treatment. However as the difference was so small, the results would not change the findings of a cost-benefit analysis (CBA) (presented in Chapter 9).

Strengths and Limitations

This is the first study to elicit WTP for Mirena and Oral treatment from an ex-ante perspective in menorrhagia. A potential limitation is that patients with experience of the condition and the treatments were included in the sample, which does not strictly meet the ex-ante perspective criteria (described in Chapter 3) but it does represent the female general population who could need the treatment. However, the number of women with experience of the treatment was not determined. It is possible that the inclusion of these respondents in the ex-ante perspective is likely to affect the validity of the WTP results as it is not clear whether these respondents based their WTP values on their own experience of the treatments or the descriptions provided. Further, despite the WTP questions for Mirena and Oral treatment not being randomly ordered, as WTP for Oral treatment always followed WTP for Mirena, the findings suggest that the results were not impacted by ordering. Mirena did not receive the highest mean WTP value, even when the respondent preferred Oral treatment, which is expected to occur due to the ordering effect (Stewart et al., 2002). This finding also suggests that the respondents did consider the two treatment scenarios as two separate elicitations tasks. According to economic theory, if respondents considered the two treatments as one elicitation task, the respondent's value would be expected to be greater for the first treatment than the second. As this was not the case in the WTP study presented in this chapter, the findings suggest that the respondents did not apply the same budget to both treatments. However, it is unclear whether an anchoring effect occurred where the respondent 'anchored' their second WTP value for Oral treatment, on the value provided for Mirena.

Comparison with Other Studies

In comparison to other WTP studies, the number of protest answers (n=6) and non-response to the WTP question (n=4 in each treatment arm) was relatively low. It is often observed that a large proportion of the sample report protest answers (Dalmou-Matarrodona, 2001). Similarly, there were few genuine zero WTP values (n= 3 for Oral treatment and n=15 for Mirena) which eliminated the need to use one of the commonly used methods, a two-part model, which accounts for zero values (Donaldson, 1998). Inconsistent responses and preference reversals were observed in this study where individuals had a preference for a treatment but either gave a greater WTP to another treatment or gave equal WTP values to both treatments. As mentioned in the QoL review chapter, Ryan and San Miguel (2000) carried out research to determine the reliability of WTP. In their ex-ante study on treatments for menorrhagia, they found that 30% of respondents provided inconsistent responses. In which case, a respondent reported a preferred treatment and gave a greater WTP value to another treatment. In contrast to Ryan and San Miguel's findings, the percentage of inconsistent responses was much lower in this study, as only 10% of responses were considered to be inconsistent. Furthermore, the exclusion of preference reversals (inconsistent responses), in this study, did not change which treatment had the greatest mean WTP value.

Ryan and San Miguel (2000) also found the cost of treatment to impact the mean WTP values, whilst in the study reported in this chapter, affordability and effect of treatment were the most commonly cited reasons for inconsistent responses and not cost of treatment. However, as noted previously, the reason for any inconsistent responses observed within the study may be related to random error as explained by Harless and Camerer (1994). Hence it is possible that these inconsistent responses occurred by chance due to a lapse in concentration by the individual (McIntosh et al., 2006).

It has been also shown in the literature that respondents tend to select prominent numbers, such as £5, £10, £20 and so on, and that the selection of prominent numbers can be related to the respondent's perception of the difficulty of the task (Hertwigg et al., 1999). That is respondents are likely to provide less precise WTP values when they do not believe that they have an adequate knowledge of the good, in this case Mirena and Oral treatment (Whynes et al., 2007). The WTP values elicited were taken from the ex-ante perspective where respondents may have experience of menorrhagia but not necessarily both treatments. It was shown that a greater proportion of respondents selected prominent numbers for the WTP for Mirena compared to Oral treatment. It is possible that this result is due to respondents finding it easier to imagine having an Oral treatment, as they are likely to have been prescribed one in the past, than it is to imagine the intrauterine device, Mirena.

Finally, as a significant different between WTP for the two treatments was not observed an alternative method has been suggested for eliciting WTP to improve its ability to differentiate between treatments. That is, the marginal WTP approach, where the respondent is asked to state which treatment they prefer and are then asked to provide a WTP value to have their preferred treatment rather than their least preferred (Donaldson et al 1997). It has been suggested that this approach may improve discrimination between treatments because, instead of eliciting WTP for both treatments, WTP for the respondent's first choice treatment is elicited. Hence, as respondents are more aware of the choice between treatments this approach is expected to lead to more discriminate answers (Shackley and Donaldson, 2000), which might result in significant differences in WTP values for the alternative treatments.

8.3 Part 2: Ex-post Perspective

The results in this section refer to the ex-post perspective and the methods explained in the previous chapter (subsection 7.2.2). The structure for the ex-post perspective largely follows the structure of the ex-ante perspective. That is, the questionnaire response, WTP for current treatment (either Mirena or Oral treatment), the qualitative findings for reasons for WTP values and finding the WTP question difficult, the econometric analysis to identify predictors for current treatment and demand curves are reported. These sub-sections are followed by a summary of WTP values for various subgroups of responses, an assessment of the association of WTP against the condition-specific measures and EQ-5D, and finally a discussion related to the findings from the ex-post perspective.

8.3.1 Questionnaire Response

As outlined in Chapter 7, in the ex-post study, the WTP question was included in a self-complete questionnaire which was posted to women who were already part of the ECLIPSE trial. In this section, the response to the questionnaire and specifically the WTP questions are detailed.

Questionnaires were posted out in August 2012 to all 506 ECLIPSE patients, excluding those women who had withdrawn from the trial (n=165). The initial response rate to the questionnaire was (117/506) 23%. In November 2012, a reminder letter in addition to a second copy of the questionnaire and a pre-paid envelope was posted to women who had not responded. The findings of the ECLIPSE trial subsequently revealed that many women had stopped treatment, had surgery or experienced menopause at the 2 year time point (Box 8.1 below). Thus in an effort to increase response rates, the reminder letter asked women who were experiencing menopause to complete the questionnaire.

Box 8.1 Findings from the ECLIPSE trial

The information available from the ECLIPSE trial stated that at **2 years:**

- Out of the 286 women randomised to oral treatment, **105** women were still taking oral treatment.
- Out of the 285 women randomised to Mirena, **167** women still had the Mirena inserted.

It is known that **before** the 2 year time point:

- 80 women crossed-over from the Oral treatment arm to the Mirena arm
- 29 women crossed-over from the Mirena arm to Oral treatment arm

After the reminder letter was posted out to the remaining women the overall response rate increased to 32%. Whilst this response could be considered to be low for a clinical trial there are several factors that must be taken into consideration. Firstly, the 5 year time point of the trial must be taken into account as the trial response rate has also decreased to 80%. Secondly, Box 8.1 shows that at the 2 year time point, there are fewer numbers of women who are still taking either of the randomised treatments. As some of these women were at the 5 year follow-up at the time the exploratory work was undertaken, these numbers in Box 8.1 are likely to be an underestimation. However, accurate data at the current time point were not available.

Out of the 163 respondents that replied, 50 responses were excluded as the women either stated that they would not pay because they were no longer taking any treatment or they were no longer contactable as they were not living at the address provided to the ECLIPSE trial. A further 19 provided a WTP value for no treatment or surgery as WTP for current treatment was elicited. As surgery is not one of the two treatments researched in this exploratory work and it would be invalid to use WTP values for no treatment, these 19 responses were excluded. Finally 9 women who were not taking either of the randomised treatments protested and were therefore excluded. These responses will be considered in the discussion (8.3.8).

The WTP values and corresponding qualitative answers were also checked to identify non-response to the WTP question and protest answers. A breakdown of the qualitative information is provided in section 8.3.3. Out of the remaining 85 women who returned the questionnaire and were currently taking one of the randomised treatments, 3 (4%) women did not provide a WTP value for their current treatment and 11 (13%) protest answers were identified from the qualitative explanations offered for the WTP value. These 14 non-responses and protest answers were also removed from the analysis for reasons explained previously. Therefore the total number on which the analysis was conducted was 71 respondents. Two of the returned questionnaires were completed retrospectively as these women were no longer taking any treatment due to menopause.

Participants

The characteristics of the sample, excluding those who gave a protest answer or non-response, are presented in Table 8.16. It can be seen that 51 respondents gave a WTP value for Mirena, and 20 respondents gave a WTP value for Oral treatment. The average expected age of menopause was 54 years old for both treatments which is close to the typically suggested average age of menopause (55 years old) for the population. Many of the women, 76% and 85% in Mirena and Oral treatment respectively, stated that they had experienced menorrhagia for more than a year. The employment status of 78% and 80% for Mirena and Oral treatment is slightly higher than observed in the ex-ante study. Similar to the ex-ante study the household income categories show that one half of the sample lie at the extremes of the income categories. The proportion of respondents that have a household income of less than £30,000 is approximately 50% for Mirena and 55% for Oral treatment, which is slightly lower than the national average where 65% are below approximately £27,000 (Cribb et al., 2012).

Table 8.16 Sample characteristics (ex-post)

Variable	Mirena (n=51)	Oral treatment (n=20)	p-value
Expected age of menopause (yrs) [SD]	53.8 [2.59]	53.8 [2.24]	0.857
Age [SD]	47.7 [3.73]	49.5 [4.44]	0.043*
Marital status (%)			
Married or living with partner	40 (78%)	14 (70%)	
Not	11 (22%)	6 (30%)	
Difference across treatment			0.454
Visits to GP (%)			
Yes	4 (8%)	9 (45%)	
No	47 (92%)	11 (55)	
Difference across treatment			0.001*
Employment status (%)			
Employed (FT)/(PT)	40 (78%)	16 (80%)	
Not	11 (22%)	4 (20%)	
Difference across treatment			1.000
Household income (%)			
Less than 20,000	16 (33%)	6 (30%)	0.910
20,001-30,000	9 (18%)	5 (25%)	0.517
30,001-40,000	9 (18%)	1 (5%)	0.263
40,001-50,000	5 (10%)	4 (20%)	0.258
More than 50,000	10 (20%)	4 (20%)	1
Main earner (%)			
Yes	23 (46%)	9 (45%)	
No	27 (54%)	11 (55%)	
Difference across treatment			0.940
Prevented from daily activities (%)			
Yes	4 (8%)	7 (35%)	
No	47 (92%)	13 (65%)	
Difference across treatment			0.009*
Drugs for pain relief (%)			
Yes	12 (24%)	12 (60%)	
No	39 (76%)	8 (40%)	
Difference across treatment			0.005*
Time off work (%)			
Yes	1 (80%)	2 (10%)	
No	50 (20%)	18 (90%)	
Difference across treatment			0.189
Number on pain scale [SD]	1.85 [2.554]	5.03 [2.92]	
Satisfied with life (%)			
Very unsatisfied	5 (10%)	1 (5%)	
Slightly unsatisfied	4 (8%)	4 (20%)	
Neither satisfied nor dissatisfied	6 (12%)	4 (20%)	
Slightly satisfied	10 (20%)	2 (10%)	
Very satisfied	24 (49%)	9 (45%)	
Difference across treatment			0.484
How is QoL affected (%)			
Not at all affected	34 (72%)	3 (16%)	
Slightly affected	8 (17%)	7 (37%)	
Affected	4 (9%)	7 (37%)	
Extremely affected	1 (2%)	2 (11%)	
Difference across treatment			0.000*
Time with HMB (%)			
Less than a year	12 (24%)	3 (15%)	
More than year	39 (76%)	17 (85%)	
Difference across treatment			0.531

*p<0.05. SD; standard deviation, QoL; quality of life, GP; general practitioner, HMB; heavy menstrual bleeding

The socio-demographic characteristics of the study respondents were assessed to identify if there is a statistically significant difference between those who provided a WTP value for Mirena to those who provided a WTP value for Oral treatment. The Chi-squared test was used for categorical variables where the number of respondents in each group is greater than 5, and the Fisher's exact test was used for categorical variables where the number of respondents in each group has a value less than 5. Continuous data were non-normal and were tested using the non-parametric Wilcoxon Rank Sum test. The respondents who are currently taking Mirena are significantly less likely to visit the GP (p=0.001), be prevented from daily activities (p=0.009), to take drugs for pain relief (p=0.005) and have their Qol affected (p=0.000). These findings suggest that these respondents find Mirena to be a more effective treatment than Oral treatment. Further, the average age of those valuing Mirena is significantly lower than those valuing Oral treatment (p=0.043).

8.3.2 WTP Mirena and Oral Treatment

The mean WTP for each treatment was calculated and is presented in Table 8.17. The maximum average WTP for Mirena along with bootstrapped confidence intervals was £31.08 (95% CI £11.26-£55.31) and for Oral treatment was £17.45 (95% CI £9.89-£25.02). It can be seen from the table that the minimum and maximum values differ between the two treatments, with Mirena ranging from £0-500 and Oral treatment ranging from £0-50. In the case of Mirena two respondents were not willing to pay for Mirena and one was willing to pay the maximum on the payment scale (£500). Whilst for Oral treatment two respondents were not willing to pay for Oral treatment and three were willing to pay £50. As the WTP data are skewed the WTP values for both Mirena and Oral treatment were log transformed. As detailed in the methods chapter (7), each WTP value has a constant of 1 added to it to ensure that zero values would still be taken into consideration when the data are log transformed. A two sample t-test was carried out and determined that the difference between the two WTP values is not statistically significant (p=0.257) as the confidence intervals for the WTP values for the two treatments overlap.

Table 8.17 Descriptive statistics for WTP (ex-post)

Max WTP	Mirena	Oral treatment	Overall
Valid numbers	51	20	71
Mean	£31.08	£17.45	£27.24
Min -Max	£0-£500	£0-£50	£0-£500
SD	£70.24	£16.92	£60.33
Median	£16	£10	£10
Two sample t-test	p=0.257		

Min/Max; Minimum/Maximum, SD; standard deviation; WTP; willingness-to-pay

The mean WTP for both treatments is presented according to income (Table 8.18) and it can be seen that generally mean WTP increases with household income but that in some cases, with the small numbers, i.e. £40,001-£50,000, the mean WTP reduces from the previous household income category. Thus WTP does not necessarily increase with household income in the ex-post study.

Table 8.18 Mean WTP against household income (ex-post)

Household income groups	WTP Mirena Mean (n)	WTP Oral treatment Mean (n)
Less than £20,000	£49.44 (16)	£9.67 (6)
£20,001-£30,000	£12.33 (9)	£8.40 (5)
£30,001 - £40,000	£25.44 (9)	£50.00 (1)
£40,001- £50,000	£19.20 (5)	£20.75 (4)
More than £50,000	£33.80 (10)	£30.75 (4)

WTP; willingness-to-pay

In Table 8.19 the WTP values for each treatment are presented against the number of respondents who chose that WTP value, to explore the influence of prominent numbers on the WTP value.

Table 8.19 Frequency of WTP value (ex-post)

WTP	**Mirena (n)**	**Oral treatment (n)**
£0	2 (4%)	2 (10%)
£2	3 (6%)	0
£4	1 (2%)	1 (5%)
£6	1 (2%)	1 (5%)
£8	2 (4%)	2 (10%)
£10	15 (29%)	7 (35%)
£12	1 (2%)	0
£16	1 (2%)	0
£18	0	1 (5%)
£20	9 (18%)	2 (10%)
£25	3 (6%)	0
£30	3 (6%)	0
£40	2 (4%)	0
£45	0	1 (5%)
£50	5 (10%)	3 (15%)
£100	2 (4%)	0
£500	1 (2%)	0
Total	51	20

WTP; willingness-to-pay. Underlined values are the prominent numbers in the payment scale

In the case of Mirena, the most common WTP values selected were prominent numbers; those are £10 and £20. For Oral treatment £10 was the most common number selected, which again is a prominent number. However, the small sample size means that even when a prominent number is selected only three times, as in the case of £50 for Oral treatment, it would be the second most common WTP value selected on the payment scale. Therefore the emphasis on selecting prominent numbers is difficult to judge from this ex-post perspective.

8.3.3 Qualitative Findings

Following the WTP questions for current treatment, the women were asked to answer two qualitative questions. The first, related to the reason behind their WTP value and the second, asked women to explain why they did, or did not, find the WTP question difficult to answer. Similar to the ex-ante study, the themes were generated by content analysis.

Reason for WTP

In total, 138 respondents offered an explanation for their response to the WTP value. Only one code was required to categorise the written explanation in 102 cases. In 36 cases two or more category explanations were required. Ten categories of reasons for a WTP value were generated from the qualitative information from the entire sample of women, including protests and non-response. The categories of reasons for the entire sample and for the sample used in the analysis are presented in Tables 8.20 and 8.21.

Table 8.20 Explanation given for WTP value (full sample, ex-post)

Category	Explanation	Total n (%)
R1	Question deemed inapplicable on grounds of: - No longer taking treatment - Patient has had surgery	31 (17%)
R2	Subject expressed difficulty estimating WTP owing to: - Difficult to answer - Cannot put a price on healthcare	8 (4%)
R3	WTP based on nominal amount - Arbitrary sum/ guess/ out of thin air	1 (0.5%)
R4	WTP reflects ability to pay (affordability) - Maximum affordable amount given current situation	48 (26%)
R5	WTP reflects reasonable value - NHS should pay but this is a reasonable limit	12 (7%)
R6	WTP reflects cost of treatment - Attempted to estimate cost - Used a comparator such as prescription costs	5 (3%)
R7	WTP reflects effects of treatment - In terms of effectiveness outcomes - In terms of process utility	36 (20%)
R8	Protest expressed at idea of payment - Paid National Insurance and taxes/ NHS should pay - Women cannot help the condition/ treatment should be free	20 (11%)
R9	Related to cost of sanitary wear - Washing clothes/wipes/painkillers	17 (9%)
R10	Misunderstood exercise but provided WTP value	6 (3%)
Total		184

NHS; National Health Service, WTP; willingness-to-pay.

Among the 71 women who provided a WTP value for current treatment, 69 provided a reason for the value. Out of the 20 women who provided a WTP value for Oral treatment, 19 provided a reason for their answer. Out of the 51 respondents who provided a WTP for Mirena, 50 provided a reason for the WTP value. It can be seen from Table 8.21, which describes the sample that are used in the analysis, that for both Mirena and Oral treatment 'R3: affordability' and 'R7: effects of treatment' are the most commonly cited reason for a WTP value. It can also be seen that there are five cases where a respondent has misunderstood the WTP question. As it could be argued that these WTP values are not valid, a sensitivity analysis is conducted to estimate the mean WTP excluding these five respondents who misunderstood the WTP question, and is reported in section 8.2.6.

Table 8.21 Explanation given for WTP value (sample analysed, ex-post)

Category	Explanation	Total n (%)	Mirena n (%)	Oral n (%)
R1	Question deemed inapplicable on grounds of: - No longer taking treatment - Patient has had surgery	0 (0%)	0 (0%)	0 (0%)
R2	Subject expressed difficulty estimating WTP owing to: - Difficult to answer - Cannot put a price on healthcare	8 (7%)	8 (10%)	0 (0%)
R3	WTP based on nominal amount - Arbitrary sum/ guess/ out of thin air	1 (0.9%)	1 (1%)	0 (0%)
R4	WTP reflects ability to pay (affordability) - Maximum affordable amount given current situation	37 (35%)	27 (33%)	10 (38%)
R5	WTP reflects reasonable value - NHS should pay but this is a reasonable limit	8 (7%)	4 (5%)	4 (15%)
R6	WTP reflects cost of treatment - Attempted to estimate cost - Used a comparator such as prescription costs	5 (5%)	3 (4%)	2 (7%)
R7	WTP reflects effect of treatment - In terms of effectiveness outcomes - In terms of process utility	30 (28%)	24 (30%)	6 (23%)
R8	Protest expressed at idea of payment - Paid National Insurance and taxes/ NHS should pay - Women cannot help the condition/ treatment should be free	0 (0%)	0 (0%)	0 (0%)
R9	Related to cost of sanitary wear - Washing clothes/wipes/painkillers	13 (12%)	10 (12%)	3 (12%)
R10	Misunderstood exercise but provided WTP value	5 (5%)	4 (5%)	1 (4%)
Total		107	81	26

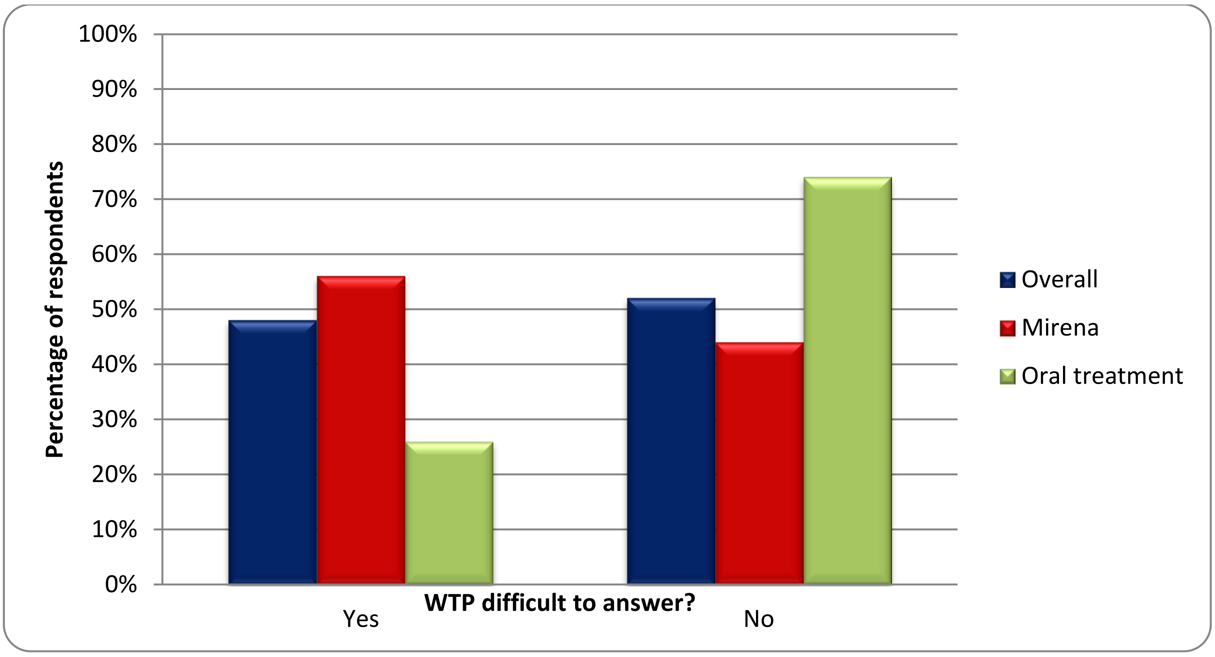
NHS; National Health Service, WTP; willingness-to-pay

Difficulty with WTP explanation

Overall, when responses from all the women who provided a WTP value for either Mirena or Oral treatment are considered, it can be seen from Figure 8.8, that approximately half of the women found the WTP question difficult to answer and the remaining half did not. 74% of the Oral treatment group did not find the question difficult, whilst more respondents (56%)

who had Mirena, did find the question difficult to answer and this difference is statistically significant (p=0.027).

Figure 8.8 Percentage of respondents that found WTP difficult to answer



Respondents were then asked to provide a written explanation for their answer. In total 79 respondents offered an explanation for the WTP value. Only one code was required to categorise the written explanation in 48 cases. In 31 cases two or more category explanations were required. Ten categories of explanation for their response to the difficulty question were generated using the qualitative information from the entire sample of women, including protests and non-response. The categories of explanations for the entire sample are presented according to whether the question was difficult to answer in Table 8.22.

Table 8.22 Explanation given for difficulty question (full sample, ex-post)

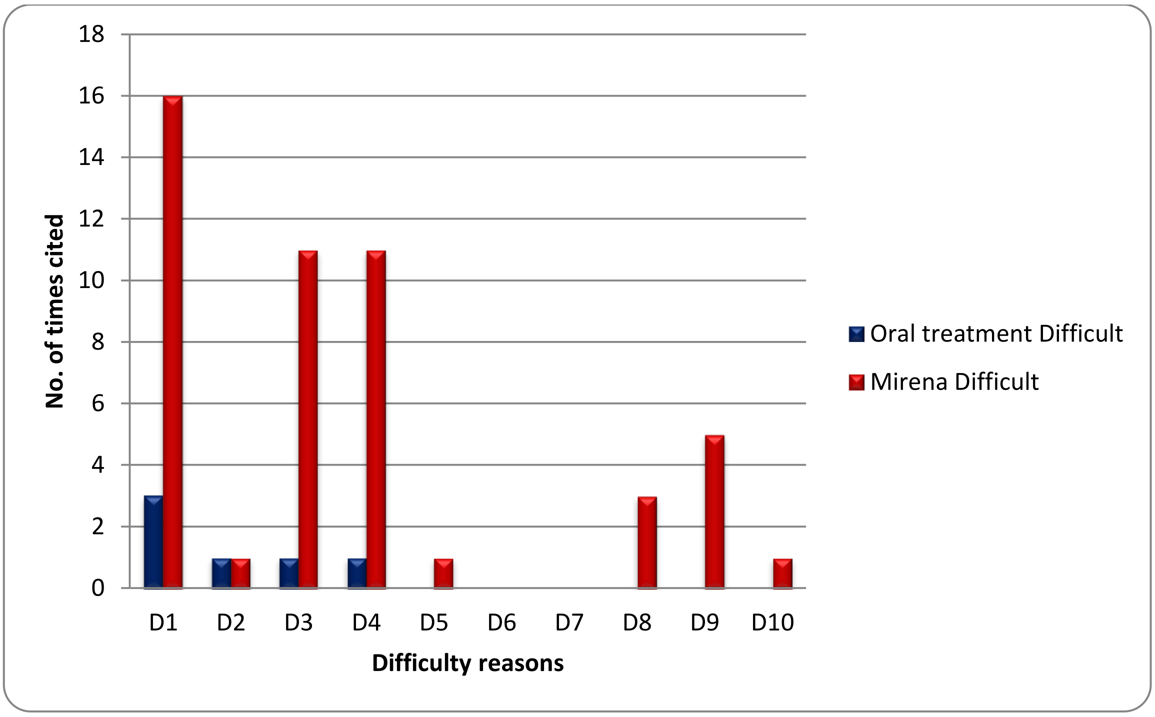
Category	Explanation	Difficult n (%)	Not difficult n (%)	Total n (%)
D1	Found valuation difficult - Not used to it/ difficult to quantify/Figure out of thin air - Not aware of cost implications	23 (34%)	1 (2%)	24 (21%)
D2	Consideration of actual costs Used prescription costs as a proxy - "I know the cost of prescription..."	2 (3%)	1 (2%)	3 (3%)
D3	Benefits of treatment – reasonable amount for expected outcomes - Due to way menorrhagia impacts lives - Due to nature of treatment effects	14 (21%)	12 (25%)	26 (23%)
D4	Ability to pay dictated amount	16 (24%)	9 (19%)	25 (22%)
D5	Based on cost of sanitary products	1 (1%)	3 (6%)	4 (3%)
D6	Explicitly explained WTP not difficult to answer - Balance of impact on QoL and affordability	0 (0%)	7 (15%)	7 (6%)
D7	Protest – I would not pay	2 (3%)	9 (19%)	11(10%)
D8	Recall issue to duration of satisfaction	3 (4%)	0 (0%)	3 (3%)
D9	Misunderstood exercise but provided WTP value	5 (7%)	3 (6%)	8 (7%)
D10	WTP reflects reasonable value - NHS should pay but this is a reasonable limit	1 (1%)	3 (6%)	4 (3%)
Total		67	48	115

NHS; National Health Service, WTP; willingness-to-pay

The most common explanation for those who did find the WTP question difficult was related to 'D1: Not used to valuing healthcare'. The most common explanation for those who did not find the WTP question difficult was 'D3: A reasonable amount to pay for expected outcomes' followed jointly by 'D4: ability to pay' and 'D7: protest'.

Among the 71 women who provided a WTP value for either Mirena or Oral treatment, 51 provided an explanation as to why they did or did not find the question difficult to answer. The categories for difficulty reasons were separated according to those who provided a WTP value for Mirena and those who provided a value for Oral treatment. Figure 8.9 shows that for those who found the question difficult to answer 'D1: not used to valuing healthcare' was the most common reason for both Oral treatment and Mirena followed by 'D3: a reasonable amount for the expected outcomes' and 'D4: ability to pay' for both Mirena and Oral. In addition to 'D2: consideration of actual treatment costs for Oral treatment'.

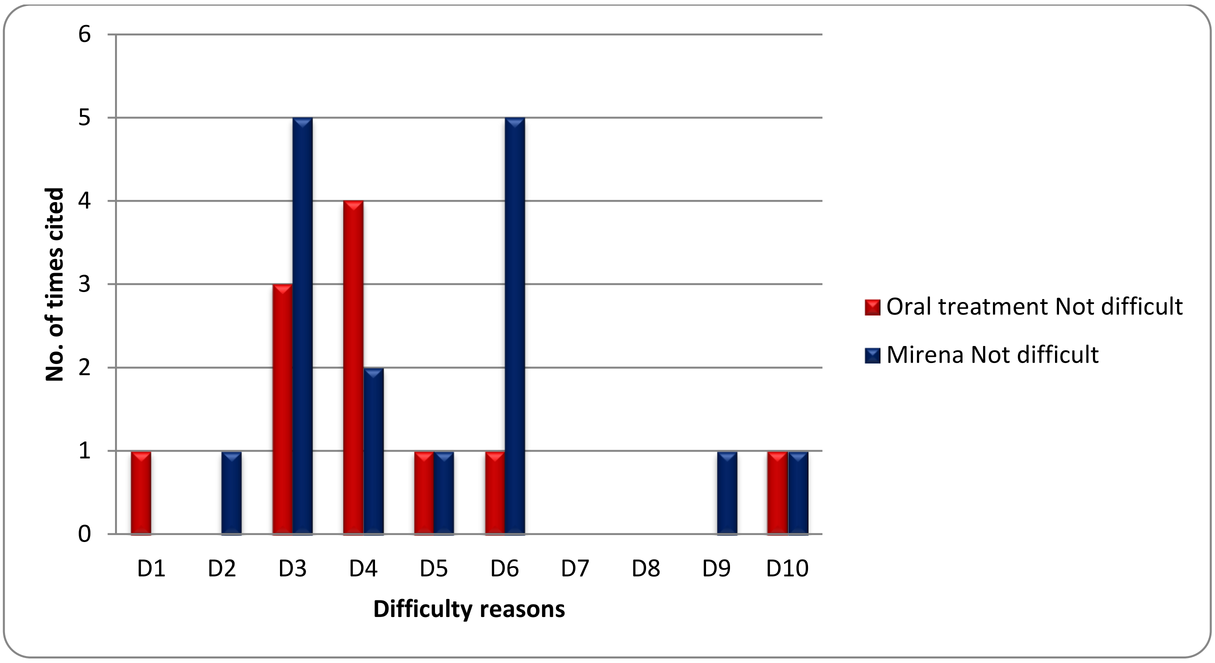
Figure 8.9 Reasons given from respondents who found the question difficult



D1: Not used to valuing; D2: actual costs; D3: reasonable amount for expected outcomes; D4: ability to pay; D5: cost of sanitary products; D6: not difficult explicitly stated; D7: protest; D8: Recall issue; D9: misunderstood; D10: reasonable value but NHS should pay

Figure 8.10 shows that for those who did not find the question difficult to answer for Mirena 'D3: reasonable amount for expected outcomes' and 'D6: explicitly explained the question is not difficult as it is a balance of impact on QoL and affordability' were the most commonly cited reasons, whilst the most common explanation for Oral treatment was 'D4: ability to pay' followed by 'D3: reasonable amount for expected outcomes'.

Figure 8.10 Reasons given for not finding the question difficult



D1: Not used to valuing; D2: actual costs; D3: reasonable amount for expected outcomes; D4: ability to pay; D5: cost of sanitary products; D6: not difficult explicitly stated; D7: protest; D8: Recall issue; D9: misunderstood; D10: reasonable value but NHS should pay

It can be seen from both Figures (8.9 and 8.10) that 6 respondents misunderstood the exercise (D9) but provided a WTP value. As it could be argued that these WTP values are not valid, a sensitivity analysis is conducted to estimate the mean WTP excluding these 6 respondents who misunderstood the WTP question (see section 8.2.6).

8.3.4 Econometric Analysis

Similar to the ex-ante perspective, it is not appropriate to conduct a two-part model due to the low proportion of zero WTP values in the sample. However, in this case, as there are only 4 respondents who provided a zero WTP across the sample, it is also inappropriate to carry out significance tests and present descriptive statistics in a univariable analysis. Nevertheless, the qualitative responses for those who provided zero values can be explored.

An OLS model, using the sample that includes the zero values, is used to identify predictors for WTP. As the number of respondents who provided WTP values for Oral treatment is small (n=20) an OLS linear regression cannot be run to identify predictors of WTP for Oral treatment. In order to explore the relationship between WTP for Oral treatment and other variables, a univariable analysis is conducted. An OLS model is run to identify predictors of WTP for Mirena and then a second OLS model is run to identify predictors of WTP for any treatment, i.e. Mirena and Oral combined.

Further, as there are very few respondents that did not respond to the question (n=3) and 11 protest answers (n=8 for Mirena and n=3 for Oral treatment), it would not be appropriate to conduct a regression model or univariable analysis to identify predictors for non-response or protest answers either.

In the following section the qualitative responses related to zero values is reported first and is followed by the univariable analysis for Oral treatment, the OLS linear regression for Mirena and the OLS linear regression on WTP for treatment, including Mirena and Oral treatment. As part of the sensitivity analysis, an interval regression model is run on WTP for Mirena and overall treatment.

Analysis of Zero WTP values

Overall, for both treatments, Mirena and Oral treatment, all 4 respondents who gave a zero value offered a reason for their WTP value, with 1 respondent providing a reason that was coded into two categories. ‘R2: difficulty completing the question’ was cited twice and ‘R4: affordability’, ‘R7: effect of the treatment’, and ‘R10: misunderstanding the exercise’ were each cited once.

Univariable analysis for Oral treatment

As an OLS model could not be run for Oral treatment due to the limited variation in the sample, a univariable analysis was conducted. The univariable analysis showed that those with a household income of more than £50,000 were willing to pay significantly more than those earning less than £20,000 ($p=0.038$; $p<0.05$). The amended EQ-5D score, with the amended wording of 'health during your cycle', was positively associated with WTP for Oral treatment ($p=0.000$); $p<0.05$). Hence those with a higher amended EQ-5D score were willing to pay significantly more for Oral treatment. However, the original EQ-5D, referring to 'health today', was not found to be a significant predictor for WTP.

OLS Linear Regression

The results of the OLS linear regression on the log transformed data for Mirena are presented in Table 8.23. As is typical for stepwise analysis all appropriate variables have been entered into the model and rejected or selected at a $p<0.1$. Similar to the ex-ante, to preserve a large sample size, the variables 'number on pain scale' and 'How is QoL affected' were dropped from the analysis as these variables contain 7% missing data. The same missing data are observed in both variables and do not occur as missing in any other variable. Two univariable analyses were conducted including 'number on pain scale' and 'How is QoL affected' to ensure that these omitted variables are not significant predictors of WTP for Mirena. In both cases, it was found that these variables are not significant at the 30% level, generating p-values of $p=0.625$ and $p=0.7150$ respectively. Similar to the ex-ante study, a flexible value of $p<0.3$ was used to determine significance at the univariable stage to ensure that the variables were not excluded unnecessarily.

In Table 8.23 it can be seen that WTP is significantly positively influenced by ‘R7: effects of treatment’ (p=0.01) and ‘D1: not used to valuing healthcare’ (p=0.008). Whilst ‘R10: misunderstanding the exercise’ (p=0.037), ‘expected age of menopause’ (p=0.015), and ‘R2: finding the valuation difficult’ (p=0.043) all have a significantly negative impact on WTP ($p<0.05$). However, in contrast to the ex-ante findings, income was not found to be a significant predictor of WTP for Mirena. Similarities between predictors of WTP for Mirena across the models include ‘expected age of menopause’ leading to a negative impact on WTP. Whilst the reason code related to ‘R1: having difficulty completing the question’ only had a tendency towards significance (p=0.063) in the ex-ante study. The link test revealed that the model type is correct and that the model is correctly specified (p=0.109). Thus it is not possible to reject the null hypothesis that an important variable has not been excluded. The Ramsey RESET test also shows that there are no omitted variables in the model (p=0.4453).

Table 8.23 OLS linear regression for Mirena WTP values (ex-post)

Variable	Coefficient	95% CI	p-value
Current MMAS	0.0144	-0.002, 0.031	0.087
R2: Found valuation difficult	-0.760	-1.494, -0.260	0.043
Expected age of menopause	-0.167	-2.30, -0.033	0.015
R7: effect of treatment	0.721	0.180, 1.262	0.01
R10: Misunderstanding the exercise	-1.094	-2.121, -0.067	0.037
D1: Not used to valuing	0.838	0.235, 1.441	0.008
Constant	10.064	3.219, 16.909	0.005

Adjusted R^2= 0.3142, n=49. MMAS; menorrhagia multi-attribute scale

In the case of the linear regression model for both treatments (Table 8.24) it can be seen that overall ‘D9: misunderstanding the exercise’ (p=0.031) has a significantly negative impact on WTP for both treatments ($p<0.05$), but having ‘difficulty with the question’ (p=0.009), ‘R7: effects of treatment’ (p=0.04) and the ‘amended EQ-5D’ (p=0.000) were positively associated with WTP ($p<0.05$). ‘R2: found valuation difficult’ (p=0.055), ‘R3: based on a nominal amount’ (p=0.08) and EQ-5D (p=0.064) had a tendency towards significance but the confidence intervals include both negative and positive coefficients. The link test again revealed that the form of the model is appropriate and that the model is correctly specified (p=0.436). Thus it is not possible to reject the null hypothesis that an important variable has not been excluded. The Ramsey RESET test also shows that there are no omitted variables (p=0.6761).

Compared to both the Mirena and Oral treatment models from the ex-ante perspective the only two variables that have appeared in this ex-post model, for overall treatment, that appear in either of the ex-ante models are ‘misunderstanding the exercise’, which has a significantly negative effect on WTP and in the regression overall treatment ‘difficult to answer’ had a positive effect on WTP whilst in the ex-ante model for Oral treatment ‘difficult to answer’ had a negative effect. Other possible variables that did not reach significance in this model, but did in the ex-ante models, include ‘cost of treatment’, ‘affordability’, and being ‘married’.

Table 8.24 OLS linear regression for overall treatment WTP values (ex-post)

Variable	Coefficient	95% CI	p-value
R7: effect of treatment	0.496	0.023, 0.968	0.04
D9: Misunderstood the exercise	-0.978	-1.861, -0.095	0.031
R3: Nominal amount	1.753	-0.216, 3.721	0.08
Difficult to answer	0.653	0.170, 1.136	0.009
R2: Found valuation difficult	-0.705	-1.427, 0.016	0.055
EQ5Damended	1.947	1.014, 2.880	0.000
EQ5D	-0.978	-2.016, 0.060	0.064
Constant	1.536	0.403, 2.670	0.009

Adjusted R^2=0.3247, n=67

The results of the sensitivity analysis using interval regression for both Mirena and overall treatment can be found in the Appendix 5.2. The OLS and interval regression models for

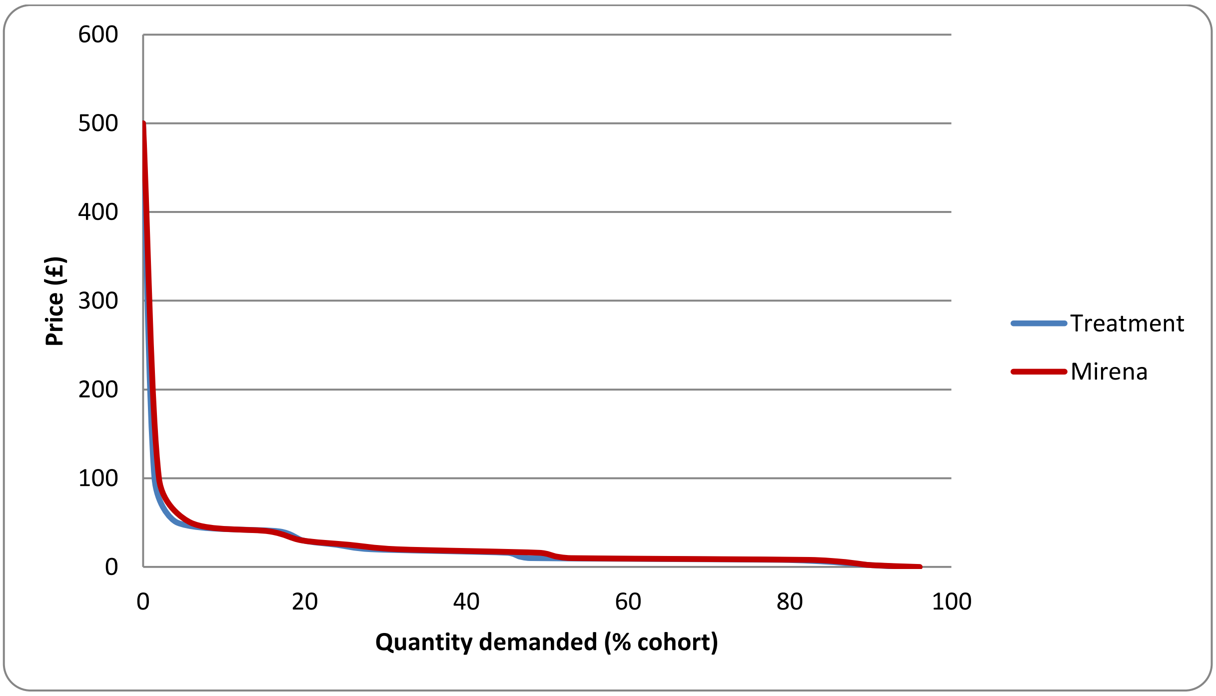
Mirena differ vastly in that only the variables ‘misunderstanding the exercise’ and ‘R7: effects of treatment’ are included in both models. Current MMAS has a tendency towards significance in the OLS model but is significant in the interval regression. The remaining variables (R3: WTP based on a nominal amount’, ‘time off work’, ‘D5: ability pay’, and ‘income’) that were found to significantly predict WTP for Mirena in the interval model are not identified as predictors in the OLS model.

Similarly OLS and interval regression models for the overall treatment also differ. The interval regression model does not include ‘D9: misunderstanding the exercise’ and ‘Difficult to answer’ as significant predictors of WTP for overall treatment but does include the remaining variables for the OLS model. Variables that have a tendency towards significance in the OLS model are significant in the interval regression (‘R2: Finding the valuation difficult’, ‘R3: WTP based on a nominal amount’). The interval regression model for overall treatment also includes ‘R10: misunderstanding the exercise’, ‘expected age of menopause’, ‘D1: not used to valuing healthcare’ and ‘prevented from daily activities’ as significant predictors of WTP for overall treatment (Mirena and Oral treatment). The interval regression model for Mirena and overall treatment was shown to be statistically significant compared to the constant only model (likelihood ratio 36.01, p=0.002) and (likelihood ratio 40.36, p=0.000).

8.3.5 Demand Curves

Figure 8.11 shows the demand curves for Mirena and overall treatment. Similar to the ex-ante demand curves it can be seen that the curves for Mirena and Oral treatment resemble a ‘typical’ demand curve. It can be seen that from £100 upwards the demand curves for both treatments are inelastic as great changes in price (£500-£100) result in very small changes in quantity demanded, but at lower prices, below £50, both treatments are elastic, because when prices are below £50 demand changes from 5% to 100% of the respondents. This is a similar result to the ex-ante study.

Figure 8.11 Demand curves for overall treatment and Mirena.



8.3.6 Summary of WTP Values for Excluded Groups

Table 8.25 presents the WTP results according to the individual subgroups that were excluded, to identify the impact of excluding these responses on WTP. The full sample group refers to the sample that was analysed in addition to the protest answers.

Table 8.25 WTP according to excluded subgroups (ex-post)

Analysis	Mean [SD]	Median	No. of Obs
Mirena			
Sample analysed	£31.08 [70.24]	£16	51
Without misunderstood	£34.53 [75.13]	£20	44
Full sample	£26.86 [66.01]	£10	59
Oral treatment			
Sample analysed	£17.45 [16.92]	£10	20
Without misunderstood	£18.38 [16.86]	£10	19
Full sample	£15.17 [16.83]	£10	23
Overall treatment			
Sample analysed	£27.24 [60.33]	£10	71
Without misunderstood	£29.65 [63.67]	£16	63
Full sample	£21.62 [56.86]	£10	89

SD; standard deviation

It can be seen from the table that, in every subgroup, mean WTP for Mirena is still greater than mean WTP for Oral treatment, but that mean and median values differ. Hence in this case, the inclusion or exclusion of protest answers and those who misunderstood the exercise made no difference to the results. However, as only 11 respondents gave zero protest values and 8 misunderstood the exercise if the numbers in each of these subgroups were greater the findings may differ.

8.3.7 Associations between Measures

This section reports the findings of the association between WTP, EQ-5D and MMAS as data were available at baseline on EQ-5D and MMAS from the ECLIPSE trial. In this case MMAS is considered to be the gold standard in menorrhagia as it has been shown to be valid (Pattison et al; 2011). The following section relates to WTP for both Mirena and Oral treatment combined (overall treatment).

As outlined in the methodology chapter (7), the association between the measures was assessed using Spearman's correlations analysis. The results are presented in Table 8.26 and show that WTP has a significantly positive relationship with change in MMAS. Thus, the greater the change in health state, as measured by changes in MMAS, the greater the WTP value. When compared to the association between change in EQ-5D and change in MMAS it can be seen that this association is not statistically significant. Thus WTP has a significant correlation with change in MMAS, whilst change in EQ-5D does not.

Table 8.26 Associations between measures

	Change in MMAS (rho)	WTP (rho)	Change in EQ-5D (rho)
Change in MMAS	1.0000		
WTP	0.2674*	1.000	
Change in EQ-5D	0.2265	-0.0158	1.000

*$p<0.05$. MMAS; menorrhagia multi-attribute scale, WTP; willingness-to-pay

The correlation values which compare the current MMAS with current EQ-5D and the amended EQ-5D are reported in Appendix 5.3 as this is not the focus of the thesis.

The percentage improvement in MMAS from baseline to current time point was also calculated. The percentage change was differentiated according to arbitrary categories to establish the extent of the improvement in disease-specific QoL from baseline. The categories, the number of respondents that fall into the categories and the mean WTP values for respondents in each category were then calculated and are detailed in Table 8.27.

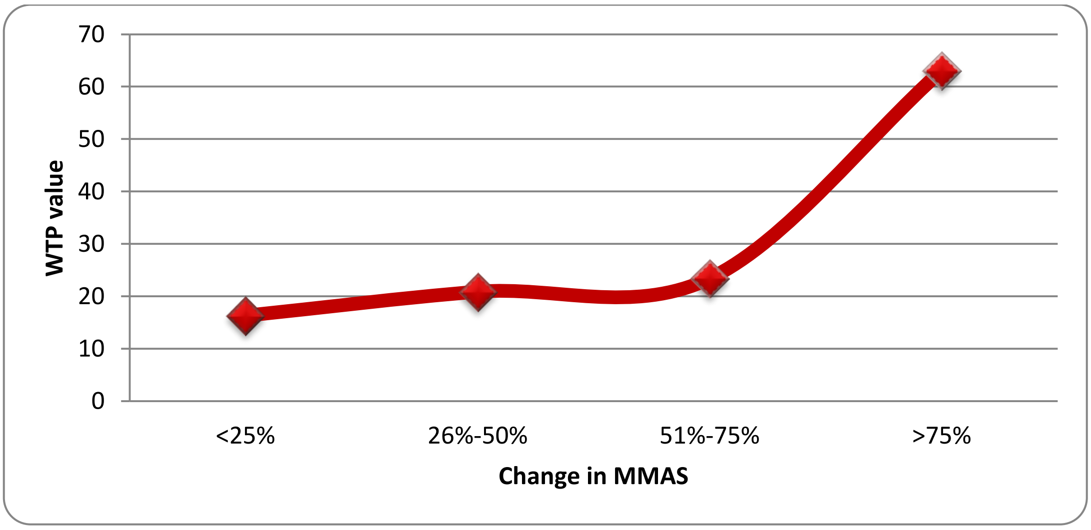
Table 8.27 Mean WTP against percentage improvement in MMAS

% change in MMAS	Number of observations	Mean WTP
<25%	21	£16.29
26% - 50%	14	£20.86
51% - 75%	24	£23.38
>75%	11	£63.09

MMAS; menorrhagia multi-attribute scale, WTP; willingness-to-pay

Figure 8.12 graphically presents the relationship between mean WTP and percentage change in MMAS. It can be seen from the figure that as the percentage improvement in MMAS increases, the mean WTP for treatment also increases. A Kruskal Wallis test was carried out to identify if the differences between the WTP values for each percentage change category are significant. Only the WTP values between '<25%' and '51%-75%' are found to be significantly increased as percentage change in MMAS increases ($p= 0.035$; $p<0.05$). Given that the significant difference was observed between the two groups with the greatest number of observations, it is likely that the remaining categories were not found to be significant due to the limited sample size for the groups '26-50%' and '>75%'. Thus there may not have been sufficient power to detect a significant difference between these groups.

Figure 8.12 Percentage change in MMAS against mean WTP



Overall, the results show that WTP behaves as would be expected. The greater the improvement in health state, the greater the WTP.

8.3.8 Discussion

In the second part of the chapter the results of the WTP feasibility study from the ex-post perspective were presented.

The results showed that WTP for Mirena (£31.08) was approximately twice as large as the WTP for Oral treatment (£17.45) but the difference was not statistically significant. The reasons for the lack of statistical significance may again be related to the nature of the payment scale and the use of prominent numbers, but also the small sample size which is discussed in 'Strengths and Limitations'. When the mean WTP values were presented against income it was shown that generally mean WTP increases with income but that the relationship is not monotonic. Out of those respondents who provided a WTP value for Mirena or Oral treatment the most commonly cited reason for WTP values were 'affordability' and 'effects of treatment'. Hence from the ex-post perspective, respondents valued the treatments and provided WTP as is expected by analysts. In the case of the ex-post perspective, approximately half of the sample found the WTP question difficult to answer (52%) and half did not (48%). The most commonly cited reason for finding the WTP question difficult was related to 'not being used to valuing healthcare' and for those who did find the question difficult to answer, the most commonly cited reason was 'a reasonable amount to pay for the expected outcomes'.

Predictors for WTP for Mirena and overall treatment using the OLS linear regression model primarily included qualitative reasons with fewer socio-demographic factors. However, in the interval regression sensitivity analysis, the results differed considerably to those using the OLS linear regression, particular for Mirena. The interval regression for Mirena included more qualitative reasons. As it is not possible to use the same test to assess the model fit, due to the nature of the models, it is difficult to establish which model is 'correct'. The reasons for this stark difference in models may be related to the limited sample size.

The results show that the inclusion and exclusion of subgroups such as those who misunderstood the question and those respondents that provided zero protest values did not change the mean WTP values. The mean WTP for Mirena was still greater than the mean WTP for Oral treatment in all cases. The findings from the tests of associations between change in measures from baseline showed that WTP is significantly associated with change in MMAS score. These findings suggest that the greater the improvement in the condition, the greater the WTP. Interestingly, change in EQ-5D was not shown to be significantly associated with change in MMAS, thus suggesting that WTP is more sensitive to changes in menorrhagia, which is likely to be due to WTP encompassing both health and non-health benefits.

Finally, a comment must be made on the number of respondents who provided WTP values for their current treatment when they stated that they were not currently taking any treatment. The responses where women provided WTP values for no treatment were excluded because it would be invalid to use WTP values for no treatment. 16 respondents provided a WTP value for no treatment, some of which were zero due to affordability. It is possible that these women may have been referring to a prior treatment or surgery, or perhaps placing a value on any treatment to alleviate menorrhagia. This misunderstanding may have possibly arisen due to the lack of clarity in the questionnaire due to the reference to 'current treatment' or perhaps the respondent may not have read the question as intended.

Strengths and Limitations

This is the first study to elicit WTP for Mirena and Oral treatment from an ex-post perspective in menorrhagia. Further, this is the first study to assess the correlation of WTP against change in MMAS, compared to the correlation of change in MMAS and change in EQ-5D.

A limitation of the study is the small sample size for WTP, in particular for Oral treatment. The small sample size, and subsequent insufficient power to detect a difference, is likely to have been the reason that a significant difference in mean WTP between Oral treatment and Mirena was not detected, despite the large apparent difference in values. However, it should also be noted that there were statistically significant differences in socio-demographic variables for those who provided a WTP for Oral treatment and those that provided a WTP value for Mirena, which could also be related to the lack of significance. The limited sample size for Oral treatment also meant that a regression model could not be run to identify predictors of WTP for Oral treatment. Finally, it is likely that the small sample size, will have also led to the lack of statistical significance between the different change in WTP and the MMAS percentage change groups. The two groups (<25% and 51-75%) that demonstrated the change were the two with the greatest sample size.

A further limitation of the ex-post perspective, that is unrelated to sample size, is the completion of the questionnaire by some respondents who were going through menopause. It is unclear to what WTP time frame these respondents were providing their WTP value for as the questionnaire states until menopause. Fortunately, the number of women who completed the questionnaire whilst experiencing menopause was very few (n=2) and this number is unlikely to affect the results.

A final limitation is that it may have been necessary to assess whether the respondents were experiencing any symptoms when providing a WTP value. Despite the advantage of eliciting WTP over extra-welfarist measures, being due to a specific recall period not being used, in order to truly determine whether the results are at all affected by the timing of assessment, the WTP values of women experiencing symptoms and those that were not could be compared.

Comparisons with Other Studies

The number of protest responses and non-response was low. Furthermore, few zero values (n=4) were observed which again is unexpected for a WTP study and eliminated the need to use one of the commonly used methods, a two-part model, which accounts for zero values (Donaldson, 1998). Whilst no other study has directly elicited WTP from an ex-post perspective in menorrhagia, San Miguel et al (2000) have indirectly elicited WTP in menorrhagia using conjoint analysis to compare conservative surgery against hysterectomy. Conjoint analysis is a method of assessing the importance of various attributes of a treatment (Ryan, 1999). However, the purpose of the study by San Miguel et al (2000) was to test the conjoint analysis method rather than to consider its use in menorrhagia. However, it was demonstrated that conjoint analysis could be used to assess WTP indirectly in the condition.

In the next part of this chapter the findings of the interviews with women from the ex-ante perspective are reported.

8.4 Part 3: Interviews

30 women, from the ex-ante perspective, who agreed to be interviewed whilst completing the questionnaire were approached.

8.4.1 Interview Response

7 of the 30 women (23%) responded to email and postal invitations. Two had recently had a hysterectomy and therefore could not attend within the specified time frame. Five agreed to attend an interview. One did not turn up to the interview, one agreed a date but could not subsequently attend due to changes in work commitments and therefore could not attend at all. Consequently three (10%) women attended an interview in February or March 2013. The interviews lasted between 20 and 60 minutes, with an average of 36 minutes. As outlined in Table 8.28, all of the interviewees had experience of menorrhagia, were white and were aged between 29-46 years old. As is typical in the qualitative literature, in the next section the interviews will be described according to the key themes identified from the thematic framework.

Table 8.28 Interview sample characteristics

Variable	Sample characteristics (n=3)
Experience of menorrhagia	3
Average age (yrs)	37
Preferred treatment	
• Mirena	1
• Oral	2
WTP Mirena	5 [0-10]
WTP Oral	39 [8-100]
Marital status	
• Married/ living with partner	2
• Single	1
Main income earner	
• Yes	2
• No	1
Income	
• <20,000	2
• 20,001-30,000	1
Employment	
• Employed FT	1
• Employed PT	2
Children in future	
• Yes	1
• No	2

FT; full-time, PT; part-time, WTP; willingness-to-pay

8.4.2 Descriptive Accounts of Interviews

The following themes were identified from the interviews and are described in this section, 'experience of menorrhagia', 'considerations related to treatments', 'reasons for WTP values', 'opportunity cost and the effect on the budget', and 'suitability of EQ-5D and WTP'.

1) Experience of menorrhagia

As all the women interviewed had experience of menorrhagia it was clear that they were relating their own experiences of the condition to the questions presented in the questionnaire. For example, when asked about their thoughts on the scenario description of menorrhagia, unsurprisingly the interviewees assessed whether the description reflected their own experience of menorrhagia.

I3: "I don't really have to imagine it because I know how it is […] I have to do a lot of that anyway"

I2: "yeah, I've been through all that"

I2: "I can relate to all that yeah I mean that I used to have to carry, not just extra sanitary protection but extra underwear as well…"

I1: "yeah it's true about the protection cause I only ever used tampax. I was using tampax and towels and socially like absolutely I wouldn't go out you know…"

But one respondent did attempt to separate her own experiences from the scenario presented on the treatments available, Mirena and Oral treatment.

I1 "…From my case I'd been told that I can't have the Mirena coil […] but putting that aside if I could have either [treatment] I would find it a lot easier to take tablets than I would the coil"

2) *Considerations related to treatments*

When deciding between the two treatments, the benefits and disadvantages of each treatment appeared to be primarily related to the process of care associated with each treatment. But the longer term benefits of Mirena were noted.

I1: "the thing that put me off with that Mirena was erm if there was any concerns… if …there was a side effect or it wasn't suitable the Oral treatment would discontinue because that's [the Mirena] fitted, that's what put me off…"

I2: "the idea of having something inserted into me for that long I don't like the sound of… having the thought of that in there as well makes me cringe…"

I2: "I'm no good at taking tablets I always forget so that one [Oral treatment] would probably be out the window…"

I2: "yeah, I'd rather have less bleeding and anxiety… I like the Mirena that there's no practical difficulties and bleed no more…. You could get on with what you normally do… whereas the Oral treatment you're still gonna get your periods"

I3: "… I would find it a lot easier to take tablets than I would the coil for practical issues…"

I3: "erm, I suppose with the coil probably it can take up to 6 months to have the effects of it when periods are bad you want the effects straight away…"

3) *Reasons for WTP values*

When providing reasons for WTP values, in addition to the points mentioned relating to the process and outcomes of treatment in the previous section, it became clear that some of the respondents based their WTP values on their perceived cost of production in addition to using prescription costs as a proxy, amongst other things.

I2: "ummh, I suppose because I dunno probably wouldn't make as many as the coils as you think Oral treatment would be bulk made you'd have to take quite a lot of them so it'd be cheaper"

I3: "yes, I think you would have indeed lots of repeat prescriptions of the tablets whereas the coil would be one fitted bit whereas the tablets there would be 28 tablets in the pack"

I3: "I suppose I wouldn't go [pay] more than a couple times more than normal prescription costs…"

Whilst other reasons for WTP values were unclear.

I2: "difficult err dunno anything over £10 just seems a lot of money"

I1: "I really don't know [laughter is heard] off the head and erm, I was trying, I did try to calculate how much my prescriptions was…"

And one respondent also focussed on the uncertainty of the treatment outcomes in her own experience when providing a WTP value for treatment.

I3: "I don't know it would all depend on the guarantee of it actually working, if it would work then there's pretty much no limit to what I'd pay…but there is no guarantee"

When the notion of WTP was explained and the respondent had the opportunity to reflect back on what was discussed in the interview, it was clear that they had not incorporated all relevant factors into their valuation for the WTP study. All of the women changed their WTP values upon reflection after taking into consideration all factors, such as the true value of treatment to them, costs aside. One woman also stated that upon reflection her opinion of the ease of completion of the WTP exercise had changed.

I3: "I'd find it a bit more difficult I think than I did yeah when I was initially filling it in because I just went with a low amount thinking what would I pay for prescriptions roughly that valuing the tablets a bit more to be willing to pay a bit more for them so I was kind of either side of prescription costs but thinking about it now bringing in other factors yeah, a bit harder to answer "

I3: I suppose it would be more difficult now justifying it a bit more I kind of put those amounts almost on instinct, what I felt right then at the time, now comparing the two different amounts yeah maybe I'm not quite as certain as I was, as I was back then."

4) Opportunity cost and the effect on the budget

When asked if they had considered the opportunity cost of not being able to spend the money on other things, it was observed that the women were not providing WTP values that would cause them to make any sacrifices in spending, which could be related to the amount of value they place on the treatments. In some cases, they misunderstood the exercise by then stating that in reality they would minimise the amount they paid through the use of pre-paid prescription cards.

I3: "If that was all the treatment I had to take, probably not much because that's not much difference to what I pay now either prescription costs or extra bits. I buy heat pads to stop the cramps and all other things like that, so it would probably balance itself out. If I'm still paying that on top of everything else, it would be a factor, it wouldn't make me immediately go into my overdraft but it would be a consideration to know that's another bill in effect every month that has to go out"

I1: "erm yeah I suppose it would be feasible you know because I do, I do I would probably get one of them prescription cards [laughter is heard] because they do help like [with the cost] they do give me 2 months' worth of you know when I go for my prescriptions so it doesn't cost so much erm but yeah, yeah I would"

I2: "probably the amount I could afford comfortably…so I'd not have to sacrifice anything"

5) Suitability of EQ-5D

When the respondents were asked what they thought about the EQ-5D questionnaire, many drew from their own experience of the condition to assess whether EQ-5D captured what is important to them.

I2: "I would be confined to bed, I wouldn't I'd still have no problems with self-care but I'd be unable to perform my usual activities and I'd be in extreme pain or discomfort and I'd be extremely anxious or depressed "

I1: "ok, erm I suppose in regards of like mobility erm when you're walking about and you've got heavy periods I don't know, ….erm I'd be less likely to do erm any usual activities. Only if the housework doesn't get done by me then nobody would do it erm pain/discomfort yeah absolutely yeah erm yeah you would be anxious and depressed you see …because of the way that it changes your moods then yeah, I would say I would be erm anxious and depressed yeah I'd go for moderately there…"

I3: "I suppose it touches a bit on the social activities that there might be, these are quite practical issues but sometimes you might not have problems with the activities but you don't want to do them because you're forcing yourself to do things that affect you rather than, it's not that you're not physically able to get out and go to work or whatever it's harder to do so and mentally the extra effort for it um I think they cover a good amount …"

One respondent discussed the issues with EQ-5D's recall period for assessing the condition stating that;

I3: "um, it can vary day by day even the week or 4 and a half weeks when I'm on my period. Some days you're fine, yes it's a bit more awkward as you have to go to the toilet more often, *you don't feel depressed, you still feel active enough, and the next day can be utterly exhausted, because you don't want to get out of bed at all. So it can work fine as a measure if it's your health today, but it isn't necessarily an average or good representative of the whole of the month or the whole week with your periods….. "*

8.4.3 Discussion

The interviews were carried out in an attempt to supplement the findings from the ex-ante WTP questionnaire to gain an in-depth knowledge of respondents understanding of the WTP question and their reasons for the WTP values. It can be seen from the small numbers of interview respondents and from the profile of the sample that the respondents who more closely reflected the criteria of an ex-ante perspective were not willing to attend an interview. Hence, only those respondents that have experience of menorrhagia were willing to attend the interview, which does not entirely meet the aim of the interviews of understanding ex-ante respondents' answers. But it does provide some insight into the respondents thought processes. Therefore the interpretation of these interview findings and their generalisability is limited.

However, whilst the findings from these particular women, with experience of menorrhagia, cannot be generalised to all respondents it was observed that the use of an interview based elicitation format may have improved the quality of the valuations provided by these particular women. The interviews showed that the women did not entirely understand the WTP question, perhaps because they themselves have the condition and were asked to imagine it. Two of the respondents were not able to explain why they chose the WTP value and also one stated that in reality she would use a prescription card to reduce the cost. Despite this finding, the respondents did discuss some of the limitations associated with EQ-5D, stating that some of the dimensions are not entirely suitable and that there are issues with the recall period used in the questionnaire.

Strengths and Limitations

The main strength of the interviews is that an understanding was gained of how respondents who have the condition provide WTP values when a scenario of the condition and the treatments are presented to them. However, a significant limitation of these findings is related to the small sample size and the sample characteristics of the respondents which have resulted in a lack of generalisability of the findings and a limited contribution to the literature.

Comparison with Other Research

To current knowledge, no other study has attempted to specifically identify respondents understanding of the WTP exercise in menorrhagia. However, previous studies have been carried out to assess the methodology of WTP through the use of such interviews. Smith (2007) carried out a thinkaloud study which suggests that higher WTP values are more stable than lower WTP values because respondents are likely to think about the values more than they would for lower WTP values. Baker et al (2008) carried out a systematic review of the qualitative research for WTP and found that there is a paucity of evidence in this area. They identified studies that showed, similar to the findings observed in this chapter, that respondents provide WTP values that would not sufficiently impact their budget to cause changes in spending habits.

Further Research

Additional interviews should be carried out to not only validate the WTP values provided in a questionnaire based format, but also to validate the responses provided in the qualitative component of WTP questionnaires. This would gain a more in-depth understanding of the suitability of WTP in menorrhagia in terms of respondents understanding of the exercise and their reasoning for WTP values. As ex-ante perspectives are currently recommended by the UK government, particular research should be focussed on the understanding of ex-ante WTP values in menorrhagia.

8.5 Overall Discussion of WTP Study

WTP was elicited from both the ex-ante and the ex-post perspective to explore the impact on the results of the theoretically preferred ex-ante perspective and the most commonly practiced ex-post perspective.

In the case of the ex-ante perspective, the difference between WTP for the two treatments was very small, with Oral treatment receiving a slightly greater WTP value than Mirena. Whilst the ex-post perspective showed that WTP for Mirena (£31.08) was approximately twice as large as the WTP for Oral treatment (£17.45), the difference was not significant in either case. It is possible that the larger difference in the ex-post group arose because the respondents were largely focussing on the benefits of their current treatment, whilst the embedding effect may have been observed in the ex-ante group as respondents may have been providing values for treatment in general. However, the stronger preference for Oral treatment may have also led to the ex-ante finding.

The wide confidence intervals and the lack of statistical significance in the ex-ante perspective are likely to be related to the nature of the payment scale, as most respondents appeared to provide values between £10-£20 and there are very few numbers in between these values for respondents to choose. Furthermore, the findings may suggest that the respondents are indifferent about the two treatments as at some point in their lives, they might want to try both treatments according to their needs, which may lead to more equal valuations. Whilst in the case of the ex-post perspective, despite the large difference observed in mean WTP, the small sample size is likely to have limited the ability to detect a statistically significant difference between WTP values.

Further, it can be seen that the confidence intervals for WTP values for both treatments in the ex-post perspective overlap with the confidence intervals of the ex-ante perspective, which is likely to be due to some women, in the ex-ante perspective, having experience of the condition and potentially the treatments. It could also be interpreted that as the confidence intervals for Oral treatment are fairly similar between the two perspectives (£10-£25 for ex-post and £11-£20 in the ex-ante), it is easier to imagine Oral treatment than it is Mirena. The confidence intervals between the ex-ante and the ex-post perspectives for Mirena are much more varied than those for Oral treatment, (£10-£20 for ex-ante and £11-£55 for ex-post) which could indicate that having experienced menorrhagia, as the ex-post women have, the benefits of Mirena weigh out the negative effects associated with initial treatment. Alternatively, as the ex-post women are providing WTP values after much longer than 6 months of treatment, typically 5 years, the beneficial effect of Mirena of causing bleeding to cease will have occurred for several years. Whilst, in the ex-ante perspective, the scenario for Mirena describes the side effects of causing bleeding to become worse during the first 6 months and then eventually the cessation of bleeding. Hence the respondents from the ex-ante perspective are accounting for these side effects when providing a WTP value, which the ex-post perspective women may not remember vividly or incorporate into their WTP value.

In both the ex-ante and the ex-post perspective studies it was shown that generally mean WTP does increase as income increases but that the relationship is not monotonic. The lack of a monotonic relationship between income and WTP is said to suggest that theoretical validity is not demonstrated. It is likely that, in both perspectives, the small number of respondents that fall into each of the five income categories is the reason for the lack of a relationship. A larger sample size may prove to be a better indicator of the relationship between WTP and income. A possible solution could be to collapse the income categories into smaller groups to reduce the spread of responses. However, the categories used are those that are typically used in other published studies and therefore enable comparability with other studies.

The two most commonly cited reasons for WTP values for both the ex-post and the ex-ante perspective were related to the ‘effects of the treatments’ and ‘affordability’. Therefore, in this study, most respondents valued the treatments according to their associated process utility and outcomes, as is required to elicit valid WTP values. It is expected that ‘affordability’, i.e. ability to pay, would be a commonly cited consideration for providing a WTP value.

Unexpectedly, the majority of women did not consider the WTP question difficult to answer from the ex-ante perspective, whilst there was an approximately equal number of those who did and did not find the question difficult in the ex-post perspective. The most commonly cited reasons for those who did and those who did not find the question difficult were the same for both perspectives. The reasons provided by those respondents that did not find the question difficult to answer were related to providing a reasonable value for the expected outcomes. Those who did find the question difficult cited not being used to valuing healthcare, which is not an unexpected finding given that generally UK citizens do not pay at the point of consumption of healthcare.

Predictors for WTP from both perspectives primarily included qualitative reasons with fewer socio-demographic factors. This finding suggests that the inclusion of qualitative reasons in the WTP elicitation process is important. A higher proportion of protest responses were observed in the ex-post perspective (13%) compared to the ex-ante perspective (6%). It is possible that those who have the condition and immediately require the treatments are more prone to strategic bias, where the respondents do value the treatment but fear they may have to pay for it in the future and therefore state that somebody else should pay (i.e. NHS).

The results of the associations between measures also showed that WTP is significantly associated with change in MMAS score, whilst change in EQ-5D was not. These findings suggest that the greater the improvement in health state, the greater the WTP. Thus WTP is more sensitive to changes in menorrhagia, which is likely to be due to WTP encompassing both health and non-health benefits.

Very little can be drawn from the interview findings, as the sample interviewed does not meet the typical definition of an ex-ante perspective respondent. All three women interviewed had experience of menorrhagia. However, it was observed that some of the women did not fully understand the WTP exercise, but it cannot be unreservedly stated that this is due to the WTP exercise itself as it could also be related to the women having experience of the condition and being asked to imagine the scenarios and the treatments.

8.5.1 Strengths and Limitations

This is the first study to elicit WTP for Mirena and Oral treatment from both the ex-ante and the ex-post perspective, and it is the first to compare the findings from the two perspectives in menorrhagia. Another strength is that once the questionnaires were developed, they were checked by clinical experts in menorrhagia, by psychologists and health economists to assess their face and content validity. Further, rather than basing the ex-ante questionnaire scenarios, for menorrhagia and treatment effectiveness, on expert opinion alone or expected outcomes, they were based on observed evidence from the ECLIPSE trial, which increases the reliability of the findings.

Although it should be noted that it was difficult to draw comprehensive comparisons between the ex-ante and ex-post perspectives because of the small sample size in the ex-post perspective. The OLS linear regression could not be carried out in the ex-post perspective for Oral treatment therefore a comparison could not be drawn between the predictors of WTP for Oral treatment across the two perspectives.

A limitation of the study is that some variables were not forced into the stepwise regression models. Variables such as income can be thought to relate to the theoretical validity of WTP and their inclusion in the regression model could increase the explanatory power of the model. However, variables are not always forced into regression models in WTP studies and in this case as WTP has not necessarily been shown to increase monotonically with income it was decided that forcing income into the regression model would be of limited value, particularly in cases where the stepwise model did not consider income to improve the explanatory power of the model.

A final limitation is the inability to assess respondents understanding of the WTP exercise, along with the inability to supplement the information on reasons for WTP values given in the questionnaires through interviews. Very few women attended an interview, all three women gave WTP values from an ex-ante perspective but had experience of menorrhagia, which meant that the findings could not be generalised to the wider ex-ante population. Therefore, very little can be drawn from the findings of the interviews.

8.5.2 Comparison with Other Studies

In comparison to other WTP studies, the number of protest answers and non-response to the WTP question in both perspectives were relatively low. It is often observed that a large proportion of the sample report protest answers (Dalmou-Matarrodona, 2001). Similarly, there were few genuine zero WTP values in both perspectives which eliminated the need to use one of the commonly used methods, a two-part model, which accounts for zero values (Donaldson, 1998).

The value for Oral treatment is 12% higher in the ex-post perspective compared to the ex-ante perspective value. The value for Mirena is 51% greater in the ex-post perspective than the ex-ante perspective. It is often reported that values provided from patients (ex-post perspective) are greater than those observed by the general population (Brazier & Rowen, 2011), and this was also observed in the WTP study reported here. These differences in values amongst perspectives are often observed and the reasons not to use patient values relate to issues around adaptation, which is explored further in section 9.5 of the next chapter (Brazier & Rowen, 2011).

Other studies have used welfarist measures in menorrhagia. The study by Ryan and San Miguel (2000) similarly elicited WTP in menorrhagia using a payment scale approach. However, WTP was elicited to compare hysterectomy against conservative treatment to assess the reliability of WTP in general, rather than to assess outcomes in menorrhagia. Finally another study has used a welfarist method in menorrhagia to compare conservative surgery against hysterectomy (San Miguel et al., 2000). However, a conjoint analysis was carried out and WTP was elicited indirectly in the choice experiment. Similar to the study by Ryan and San Miguel (2000), the aim of the study was to assess the use of conjoint analysis rather than assessing outcomes in menorrhagia.

8.6 Conclusion

In this chapter the results of Part 1: the ex-ante perspective, Part 2: the ex-post perspective and Part 3: the interviews are reported. It was found that the treatment with the greatest WTP differed according to the perspective used. From the ex-ante perspective Oral treatment had the greatest WTP value and from the ex-post perspective Mirena had the greatest WTP value. Respondents completed the WTP exercise by focussing on the benefits of treatment and affordability. A higher percentage of respondents in the ex-post perspective found the WTP questions difficult to answer than in the ex-ante perspective, and WTP was found to have a greater correlation with change in MMAS than change in EQ-5D did in the ex-post perspective. The number of women who agreed to attend an interview was too small to obtain any generalisable findings.

The next chapter reports the CBA comparing Mirena and Oral treatment using the WTP data reported here from the ex-ante perspective.

CHAPTER 9. COST-BENEFIT ANALYSIS

9.1 Introduction

In this chapter a cost-benefit analysis (CBA) comparing Mirena against Oral treatment is reported. It has been discussed in the previous chapters that there are concerns with the use of current extra-welfarist outcome measures based on EQ-5D and SF-6D in menorrhagia. Further in Chapter 6, where the economic evaluation alongside the ECLIPSE trial was reported, it was observed that when used in an economic evaluation the different measures provide different recommendations regarding which treatment is most cost-effective.

Now that the benefits, in terms of willingness-to-pay (WTP), for both Mirena and Oral treatment have been elicited using the welfarist framework (reported in Chapter 8), the WTP outcomes are combined with the costs associated with each treatment in a welfarist economic evaluation. The CBA will use the WTP data from the ex-ante perspective only and the justification for this will be explained in the discussion (section 9.5). The objective of this chapter is therefore to report an economic evaluation from the welfarist perspective to establish the cost-benefit of Mirena compared to Oral treatment. These findings can then be compared to those using the extra-welfarist (cost-utility) analysis reported previously (Chapter 6).

The reporting of this economic evaluation follows CHEERS guidelines (the checklist is presented in Appendix 3.2) (Husereau et al., 2013).

Firstly, the methods of CBA are presented. As the methods related to the WTP elicitation exercise and resource use have been reported in detail in Chapters 6 and 7, in this chapter these methods will be only briefly described. The analysis and results sections follow and finally a discussion on the findings is reported.

9.2 Methods

A CBA was carried out based on an outcome of WTP, from an ex-ante perspective. The analysis is related to the primary care setting and provides an assessment of the difference in costs and WTP between interventions over a 24-month time horizon. A 24-month time horizon and a National Health Service (NHS) perspective for costs were taken to enable comparisons to be drawn between the findings from the CBA and the findings of the cost-utility analysis (CUA) reported in Chapter 6. The societal perspective for costs was not used in this evaluation and the explanation for this will be considered in the discussion, section 9.5.

9.2.1 Participants and Study Design

The participants and study design is reported in detail in Chapter 7, section 7.2.1. Briefly, 110 women were recruited from any gynaecological outpatient clinic in the Birmingham Women's Hospital between December 2012 and January 2013. Women who did not necessarily have experience of menorrhagia or its treatments were sought, so all women attending an appointment were approached to complete a booklet questionnaire, either in the clinic or at home. Those respondents who took the questionnaire home to complete were given a stamped addressed envelope. Women were asked to value two treatments Mirena and Oral treatment which are described in detail in section 2.5 of Chapter 2.

The South West England research ethics committee and all relevant local committees approved the study.

9.2.2 Outcome Measures

As described in Chapter 7, section 7.2.1, data on the monetary outcome measure were collected in terms of WTP for both Mirena and Oral treatment. As women were not expected to have experience of menorrhagia, the questionnaire included a scenario description of menorrhagia, and its two alternative treatments, Mirena and Oral treatment (section 7.2.1). WTP was elicited for both treatments as a maximum monthly, out of pocket cost, in the payment scale format. The booklet questionnaire also collected information on socio-demographic details. The average WTP value for each treatment was calculated.

9.2.3 Cost and Resource Use

As the majority of women were not taking either Mirena or Oral treatment, primary data were not available for these women, so cost data from the economic evaluation of the ECLIPSE trial, reported in Chapter 6, were used as the best available source and were considered an appropriate proxy for the costs. Resource use was collected as part of the ECLIPSE trial and the unit costs were applied, as described in section 6.2.4 of Chapter 6. Briefly, costs were collected from a UK NHS perspective and the general healthcare costs for both treatments included healthcare staff costs and the cost of the treatments. The costs of Mirena and Oral treatment were estimated from the British National Formulary (BNF, 2011). Staff costs were calculated using the nationally recognised reference costs (Curtis, 2011). The overall costs for both Mirena and Oral treatment at the 2 year time point in the ECLIPSE trial included crossover between treatment arms, as the analysis within the trial was intention to treat. Table 6.4 of Chapter 6, section 6.4, reports the average cost of Mirena and Oral treatment per person to be £430 and £330 respectively. These costs are used to enable comparability between the CBA and the CUA results. The strengths and limitations of using these costs are explored in the discussion.

9.3 Analysis

The net benefit (NB: WTP minus the cost of treatment) of each treatment along with the bootstrapped 95% confidence intervals around the mean values are presented. Bootstrapping enables an assessment of uncertainty in the mean value, by randomly sampling values, with replacement from the observed values. Multiple samples are drawn, as 1000 bootstrap datasets are generated using STATA (v11.0), and each dataset is considered to be a reiteration of the trial (Glick et al., 2007). The distribution of the bootstrapped values is then presented graphically.

If the NB of the treatment is positive a welfare gain is said to occur and the intervention should be implemented. If the NB for the treatment is negative there is a welfare loss and the intervention should not be implemented. The treatment with the greatest NB should be recommended as the first line treatment. For completeness the incremental NB is also presented which shows the difference between the NB's. (i.e. Mirena NB – Oral NB). In this case if the incremental NB exceeds zero, Mirena would be considered the most cost-beneficial intervention. Similar to the economic evaluation reported in Chapter 6, the recommended discount rate of 3.5% was also applied to both the costs and outcomes as the evaluation time horizon was beyond 1 year (NICE, 2008). The WTP values derived for both Mirena and Oral treatment were based on a monthly amount, to obtain the present value the WTP value was discounted for every month up to and including 24 months. All costs are reported in 2011 prices in UK (£) sterling using the UK hospital and community health services index (Curtis, 2011).

The base case analysis is presented using the cost data described above, which relates to the outcome of the economic evaluation alongside ECLIPSE, reported in Chapter 6, and is based on an intention to treat analysis.

A deterministic sensitivity analysis using alternative cost data, which were not related to the model or based on intention to treat analysis, was carried out to identify the impact of the use of different types of cost data. In this sensitivity analysis the primary resource use and cost data related to the exclusive use of either Mirena or Oral treatment for 2 years were applied, thus treatment cross-over and movement to other treatment types was not considered. Table 9.1 outlines the cost data used in the sensitivity analysis. As Oral treatment comprises a range of pharmaceutical treatments the average cost of Oral treatment was weighted according to the frequency with which each treatment is prescribed, similar to the evaluation alongside the ECLIPSE trial. For this sensitivity analysis it should also be noted that the WTP scenario description refers to Oral treatment, but one treatment, Depo-provera, is an injection and not an oral tablet. This medication was only prescribed twice in the trial data so the weight given is very small and its effect on the overall cost data is as little as 20 pence. As the difference was very small and the inclusion of this medication provides a strict comparison between the CBA and CUA, Depo-provera was included in the calculation to form the average cost of Oral treatment. Similar to the economic evaluation reported in Chapter 6, the cost of repeat prescriptions was calculated as a weighted average, as described in section 6.2.4.

Table 9.1 Cost data used in sensitivity analysis

	Unit cost	Source
LNG-IUS		
Consultation (GP 10 mins)	£26.67	Curtis 2011/ expert opinion
Insertion		
GP (20 mins)	£53.33	Curtis 2011/ expert opinion
Practice nurse (20 mins)	£17.00	Curtis 2011/ expert opinion
Device cost	£88.00	BNF 62
Sterile pack (insertion)	£21.63	NICE (inflated to 2011)
Follow-up		
6 week review: (GP 10 mins)	£26.67	Curtis 2011/ expert opinion
3 month: (GP 10 mins)	£26.67	Curtis 2011/ expert opinion
Oral treatment		
Progestogen (Cerazette)	£8.68	BNF 62
Tranexamic acid (Cyclokapron)	£14.30	BNF 62
Mefenamic acid (Ponstan)	£15.72	BNF 62
Norethisterone	£2.18	BNF 62
Combined oral contraceptive (Microgynon)	£2.82	BNF 62
Methoxyprogesterone acetate injections (Depo-provera)	£6.01	BNF 62
Consultation: (GP 10 mins)	£26.67	Curtis 2011/ expert opinion
Review of medication (GP 10 mins)	£26.67	Curtis 2011/ expert opinion

Cost data are reported in UK £ sterling. The cost year is 2011. GP; general practitioner, LNG-IUS; levonorgestrel-releasing intrauterine system

9.4 Results

The response rate, socio-demographic details, and reasons for WTP answers are reported in detail in Chapter 8. Briefly, 99 women provided a WTP value for Mirena and Oral treatment, their average age was 37 years old and 80% of women said they had experience of heavy periods at one time in their lives, but this may not necessarily mean experience of heavy periods over consecutive cycles as defined by menorrhagia (see Chapter 8, section 8.2.1). Typically women did not consider the WTP question to be difficult to answer and the two most commonly cited reasons for a WTP value were related to the effect of the treatment and affordability.

The base case results using cost data from the ECLIPSE trial are reported below, followed by the results from the sensitivity analysis.

9.4.1 Base Case Results

When using cost data from the ECLIPSE decision model it can be seen in Table 9.2 that the maximum average WTP for Oral treatment was 13% higher than the cost of the intervention, whilst the maximum average WTP for Mirena was 15% lower than the cost of treatment over 2 years.

Table 9.2 Base case results: mean WTP and cost of treatment

Intervention	WTP [95% CI]	Cost	Source of costs
Mirena	£365.08 [£247.02-£483.14]	£430	ECLIPSE trial
Oral treatment	£371.67 [£267.00-£476.33]	£330	ECLIPSE trial
Mean difference	£-6.59	£100	

Cost data are reported in UK £ sterling and refers to 2011. Costs are rounded to the nearest 10.

The NB (WTP minus cost of treatment) for each treatment along with the bootstrapped 95% confidence intervals around the mean value is reported in Table 9.3. These base case results show that Oral treatment provides a positive NB of £45, resulting in a welfare gain, and Mirena produces a negative NB of £-68, leading to a welfare loss. The incremental NB exceeds zero suggesting that Oral treatment is cost-beneficial compared to Mirena. However, the confidence intervals between the NB's for both treatments do overlap, suggesting that the difference is not statistically significant and that there is an indifference between which treatment is most cost-beneficial. However based on the mean values Oral treatment would be considered the most cost-beneficial intervention.

Table 9.3 Base case net benefit results

Intervention	Net benefit	95% CI	Incremental Net Benefit
Mirena	£-67.94	£-186.16 – £50.28	
Oral treatment	£45.29	£-55.14 – £145.72	£113.23

CI; confidence interval

A clear illustration of the welfare gain and welfare loss produced by these treatments can be seen in Figures 9.1 and 9.2 where the bootstrapped NB's for Mirena and Oral treatments are presented. Each point on the x-axis represents one bootstrapped plot. It can be seen in Figure 9.1, that in the majority of cases Oral treatment produces a positive NB, as a greater proportion of the bootstrapped NB values lie above £0, whilst Figure 9.2 for Mirena is the inverse of that for Oral treatment, showing that in the majority of cases Mirena would produce a negative NB, as a greater proportion of the bootstrapped NB values lie below £0. Despite the confidence intervals overlapping, this distribution plot of bootstrapped values suggests that Oral treatment is more likely to be cost-beneficial relative to Mirena.

Figure 9.1 Base case results: bootstrapped net benefits - Oral treatment

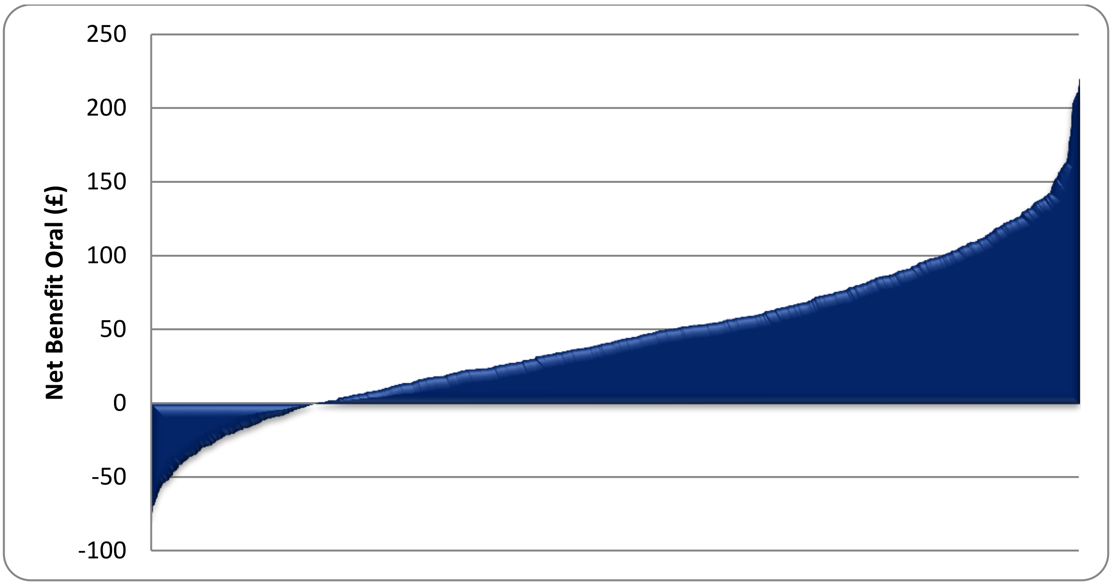
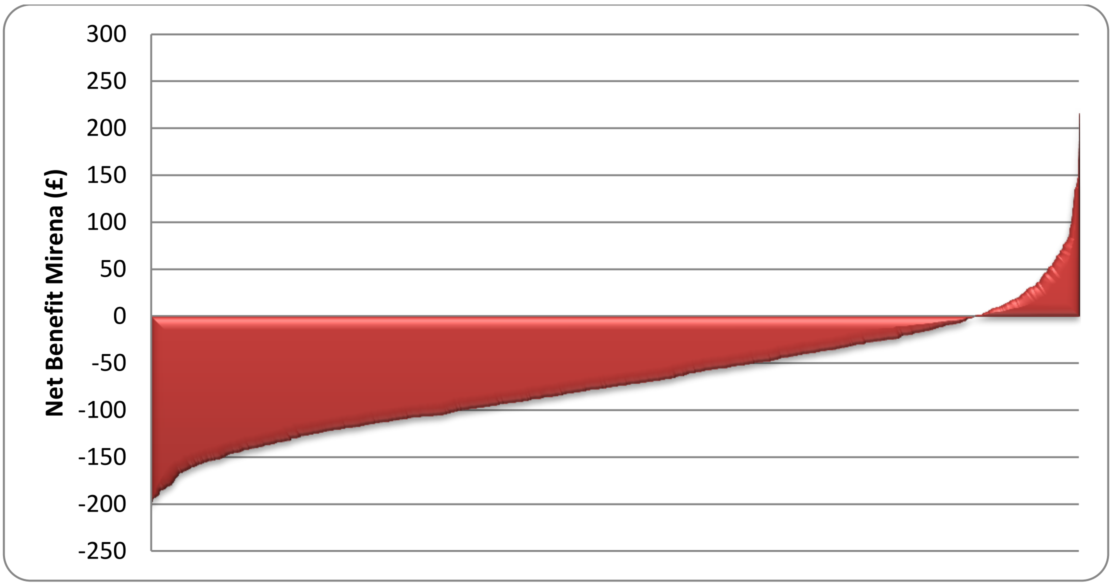


Figure 9.2 Base case results: bootstrapped net benefits - Mirena



9.4.2 Sensitivity Analysis

In the sensitivity analysis where the cost data are based on the primary costs incurred by using Mirena and Oral treatment alone, it can be seen from Table 9.4 that mean WTP for Mirena is 41% greater than the cost of Mirena. Whilst the mean WTP for Oral treatment is 280% greater than the cost of Oral treatment.

Table 9.4 Sensitivity analysis: mean WTP and cost of treatment

Intervention	WTP [95% CI]	Cost	Source of costs
Mirena	£365.08 [£247.02-£483.14]	£260	See methodology
Oral treatment	£371.67 [£267.00-£476.33]	£100	See methodology
Mean difference	£-6.59	£160	

Cost data are reported in UK (£) sterling and the year is 2011. Costs are rounded to the nearest 10.

The NB and the 95% bootstrapped confidence intervals are presented in Table 9.5, both treatments result in a welfare gain. The NB's for Mirena and Oral treatment are £105 and £274 respectively. Oral treatment generates a greater welfare gain than Mirena. The incremental NB exceeds zero suggesting that Oral treatment is the most cost-beneficial compared to Mirena. However, similar to the base case results the NB confidence intervals for both treatments overlap, suggesting that the difference between treatments is not statistically significant, but based on the mean values Oral treatment would be considered the most cost-beneficial intervention.

Table 9.5 Sensitivity analysis net benefit results

Intervention	Net benefit	95% CI	Incremental Net Benefit
Mirena	£105.48	£-10.11 – £221.07	
Oral treatment	£273.83	£168.01– £379.65	£168.35

CI; confidence interval

The bootstrapped NB's for each treatment are plotted in Figures 9.3 and 9.4 to illustrate the range of values. The plots show that in the case of Oral treatment (Figure 9.3) the entire range of values generate a welfare gain, as the NB's are all above £0. Whilst in the case of Mirena (Figure 9.4) some of the values do produce a welfare loss which is represented by the values below a NB of £0. Overall, the NB for Mirena is not as great as that for Oral treatment, as the positive NB values for Mirena lie between £0 and £250 and the NB values for Oral treatment lie between £0 and £450. Hence these bootstrapped plots would suggest that despite the overlap between the confidence intervals of the treatment, Oral treatment is likely to be considered the more cost-beneficial intervention.

Figure 9.3 Sensitivity analysis results: bootstrapped net benefits for Oral treatment

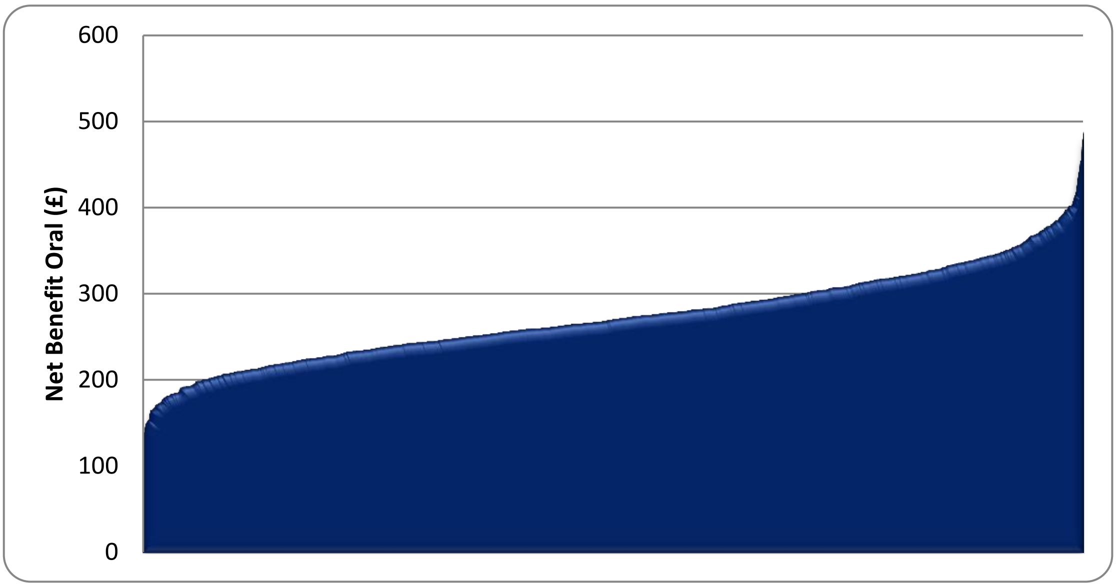
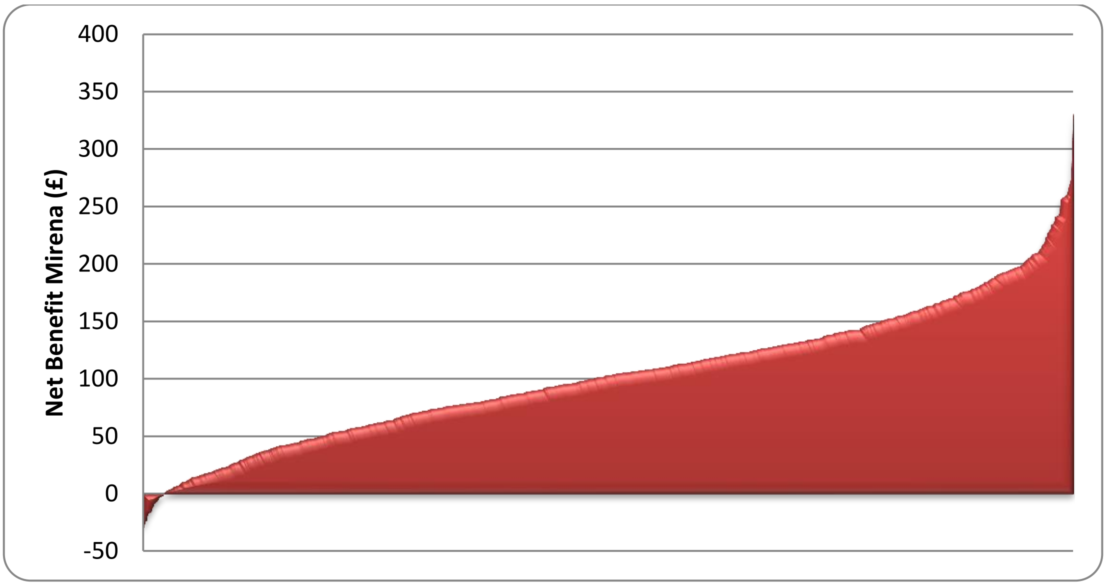


Figure 9.4 Sensitivity analysis results: bootstrapped net benefits for Mirena



9.5 Discussion

In this chapter an economic evaluation, using CBA to compare Mirena against Oral treatment is reported. The results show that Oral treatment is both less costly, as it costs £100 less than Mirena, and generates on average £6.60 more benefits than Mirena, but this difference in benefits is greater by a non-significant amount. The overall NB of Oral treatment was £42 and generated a welfare gain as it has a greater mean WTP relative to the mean treatment cost. Mirena generates a welfare loss of £-68 as the mean WTP for Mirena is less than the cost of the treatment. The incremental NB was £113 in favour of Oral treatment. Hence based on the mean NB and mean incremental NB values, Oral treatment would be recommended as the first line treatment for menorrhagia.

The findings from the sensitivity analysis show that whilst both Oral treatment and Mirena generated a welfare gain of £274 and £106 respectively, Oral treatment generated a greater welfare gain as the NB was 61% greater than the NB for Mirena. Similar to the base case findings, despite the lack of statistical significance between the NBs of the two treatments, based on the mean NB values and the mean incremental NB values (£168 in favour of Oral treatment), Oral treatment is considered the most cost-beneficial intervention.

Therefore the base case analysis and sensitivity analysis both support the finding that based on the mean NB values, Oral treatment would be considered the most cost-beneficial treatment and recommended as the first choice treatment in clinical practice, when a CBA is used.

9.5.1 Strengths and Limitations

This is the first study, to current knowledge, that applies a CBA from an ex-ante perspective to compare Mirena against Oral treatment in menorrhagia.

It could be argued that the study did not strictly use an ex-ante perspective as some women with experience of menorrhagia and its treatments were used. However, it could also be argued that the sample does to some extent reflect the at risk population group, which would be made up of both women who have and do not have the condition. This research is a feasibility study carried out to compare the cost-benefit of two treatments for menorrhagia and to demonstrate the merits of using a CBA when comparing Mirena against Oral treatment.

It was not deemed appropriate to carry out a CBA from the ex-post perspective for several reasons. First the limited data available (20 for Oral treatment and 51 for Mirena) meant that any conclusions drawn from the findings would not be particularly meaningful and could not be generalised to the wider population. Second, it would also be unsuitable to carry out a CBA from the ex-post perspective because the women used to elicit WTP were several years beyond immediate treatment. Hence the beneficial effects of the treatments will have occurred for several years and these women may not account for the initial side effects and difficulties with treatments when providing a WTP value. Third, decision-makers do not recommend the use of an ex-post perspective because in a publicly financed healthcare system society funds healthcare, therefore, all of society's views should be taken into consideration when valuing outcomes. Societal values are also recommended due to issues of adaptation, that is, individuals with a condition may not be able to desire properly because they have adapted to the condition, therefore society should value outcomes on their behalf. Finally, theoretically the ex-ante CBA is recommended for a publicly funded healthcare system (O'Brien & Gafni,

1996) so for all these reasons outlined above the CBA from the ex-ante perspective is considered the most suitable.

In the ex-post perspective a low response rate was observed as only 20 WTP values were elicited for Oral treatment and 51 were elicited for Mirena. However, a greater difference between the average WTP values for treatment was observed in the ex-post perspective than in the ex-ante perspective. These differences were likely to arise due to the ex-post women having current experience of the treatments, hence these women are fully informed about the condition and the treatment as they have first-hand experience of both. A discussion around the difference in results between the two perspectives is provided by Brazier et al (2005) and they state that there is potential for patient values to be taken into account to some degree. If this is the case then the ex-ante perspective WTP values which have been used in the CBA analysis reported here, are reasonably informed, since some women do have experience of menorrhagia and in some cases, its treatments. This is not unexpected given that women were recruited from gynaecological outpatient clinics. However, the proportion of patients with these experiences may not be representative of the female general population.

The costs used in the base case were taken from the average overall results of the model-based economic evaluation reported in Chapter 6. These cost data were used to enable comparability between the CUA and CBA as they reflect changes that women in each treatment arm are likely to experience. However a limitation of this approach is that the probabilities of moving to different states (i.e. stopping treatment, moving to the other treatment and having surgery) were not presented in the WTP scenario and hence women did not consider this when providing a WTP value. Incorporating these probabilities and other states into the WTP scenario would have made the elicitation exercise cognitively burdensome. However, when using cost data that are only related to the WTP scenario the same treatment was found to be superior. In this case although the overall cost-benefit decision did not differ, the extent of the welfare gain produced by Oral treatment compared to Mirena did vary, and was dependent on the cost data used in the CBA.

Whilst this study was the first to carry out a contingent valuation and CBA of two non-surgical treatments for menorrhagia, it was not possible to conduct a comprehensive CBA as societal costs were not available. As mentioned in the introduction of section 9.2 only the costs from an NHS perspective were considered in this evaluation. The potentially relevant costs to healthcare and society are outlined below.

NHS (healthcare) costs include; treatment cost and consumables, healthcare providers time (Clinician and Nurse) for administration of treatment, initial consultations and follow-up appointments.

Societal costs include; lost productivity which is related to time off work or reduced productivity at work, travel time and travel costs, out of pocket prescription costs and over the counter medication

Of these societal costs, it is hypothesised that the two most likely to have an impact are lost productivity and out of pocket prescription costs. Incorporating these costs however is not straightforward as it is possible that the WTP outcome already incorporates the relevant societal costs as women were presented with a scenario that took into consideration the effect of menorrhagia and its treatments on work/daily routine. This notion of potential 'double-counting' has been reported extensively in the literature (Drummond et al., 2005). Hence the increase in productivity as a result of treatment will be captured by the benefits and by incorporating changes in productivity into the cost side of the equation, it is possible that the benefits of treatment are double counted. The only other more relevant societal cost that may not be captured is the cost of prescriptions. The qualitative findings, reported in Chapter 8, showed that many women used the cost of prescriptions as an anchor for their WTP value and so these prescription costs may to some extent be taken into consideration with the benefits. Given that these out of pocket prescription costs are only incurred in the Oral treatment arm, for some of the treatments, it could be considered as bias in favour of Oral treatment not to consider this cost. However, on average, across all the Oral treatments, this cost is likely to be incurred eight times over 2 years, and as the standard UK cost of a prescription is £7 the exclusion of this societal cost is unlikely to change the overall recommendation of the most cost-beneficial treatment.

9.5.2 Comparison with Other Studies

To current knowledge, this is the first study to carry out a CBA comparing Mirena against Oral treatment. A recent CBA has been carried out but in the area of spinal surgery in Switzerland, where WTP was elicited from the ex-post perspective using patient values (Haefaeli et al., 2008). The authors elicited the ex-post WTP and conducted the CBA but suggested that further methodological work be carried out on the use of ex-ante WTP values, as this perspective is recommended for publicly funded healthcare systems.

As reported in Chapter 6, other studies have carried out CUAs but none have directly compared Mirena against Oral treatment. When compared to the findings of the CUA carried out in Chapter 6, it can be seen that the recommendations for first line treatment in the CBA are the same as those of the CUA using SF-6D. Both analyses show indifference between treatments, with Oral treatment being recommended as first line treatment for menorrhagia. However, when the CUA was estimated using EQ-5D it can be seen that these results and those of the CBA reported in this chapter do not concur. As decision-makers currently recommend EQ-5D for the valuation of outcomes, Mirena would be considered the most cost-effective treatment, despite other measures demonstrating that Mirena is not the most cost-effective intervention. These findings have significant implications for decision-makers which will be explored further in the next chapter.

9.5.3 Implications and Further Research

The results of the CBA provide evidence that would be seen to question the recommendations of the NICE guidelines which suggest that Mirena (LNG-IUS) should be used as the first line treatment for menorrhagia. The results reported in this chapter show that there is somewhat of an indifference about which treatment is most cost-beneficial, with the results tending to favour Oral treatment as the most cost-beneficial intervention. The characteristics of the study population did not completely reflect the ex-ante criteria as a high proportion (80%) of the sample had experience of menorrhagia or its treatments. In order to provide generalisable and robust evidence to decision-makers it would be beneficial to conduct the economic evaluation using a study population, with a larger sample size, but with fewer women who have experienced menorrhagia and its treatments to more closely reflect a general population and an ex-ante perspective.

9.6 Conclusion

In this chapter the CBA comparing Mirena against Oral treatment was reported. The results suggest that Oral treatment was the most cost-beneficial intervention in both the base case and sensitivity analysis, and would be most likely recommended as the first line treatment for menorrhagia. In the final chapter of the thesis the overall discussion is reported, summarising the entire thesis and providing reflections of welfarist and extra-welfarist approaches to valuing outcomes.

CHAPTER 10. DISCUSSION AND CONCLUSION

10.1 Introduction

The objective of this research was to:

1) Determine the value of levonorgestrel-releasing intrauterine system (LNG-IUS) compared to usual medical treatment in menorrhagia, and
2) Consider the suitability of current measures available for assessing the value of outcomes and interventions in menorrhagia.

In essence, to meet these objectives, a comparison was drawn between welfarist and extra-welfarist approaches to valuing outcomes in menorrhagia.

In the clinical condition of menorrhagia, an assessment of quality of life (QoL) is the primary indicator of treatment success. So it is particularly important to ensure that a suitable measure that accurately captures patients concerns and experiences is available for use. Menorrhagia is a particularly interesting condition to consider for two reasons; first the condition is chronic but symptoms occur in episodes and second, the condition is known to greatly impact on non-health aspects as well as health aspects of women’s lives. These are two properties that can make valuing outcomes in menorrhagia more problematic than other conditions.

In this thesis, alternative approaches to valuing the outcomes in menorrhagia have been assessed by exploring the use of outcome measures from the two theoretical frameworks in health economics, welfarism and extra-welfarism. The extra-welfarist outcome measure, the quality adjusted life year (QALY), is currently recommended by decision-makers such as the National Institute of Health and Care Excellence (NICE). In contrast, the willingness-to-pay

(WTP) approach to valuing outcomes which is based on a welfarist framework is not currently recommended but has potential advantages compared to extra-welfarist measures and these may be potentially beneficial to conditions such as menorrhagia.

To compare the alternative frameworks for valuing outcomes in menorrhagia, data were collected using the extra-welfarist instruments of EQ-5D and SF-6D, and the welfarist approach of WTP was also used. An economic evaluation comparing LNG-IUS against usual medical treatment was carried out using measures from each framework to compare the cost-effectiveness decision.

In this chapter a summary of the key findings from the entire thesis is presented. This is followed by reflections on welfarist and extra-welfarist measures in menorrhagia, strengths and limitations, main findings in relation to other research, implications for policy, further research recommendations and finally a conclusion of the thesis.

10.2 Summary of the Findings

The conclusions from the systematic review in Chapter 4 of economic evaluations comparing LNG-IUS and usual medical treatment were that there was no reliable evidence on the cost-effectiveness of these treatments. Furthermore, this conclusion was drawn from all the available published evidence to date which was typically based on secondary data, with the exception of one small study that used primary data but was not carried out in the UK (Brown et al., 2006). This suggested there was a clear need to carry out a UK based economic evaluation alongside a trial that directly compared LNG-IUS and usual treatment to inform clinical practice. Further, when exploring the use and assessment of outcome measures in menorrhagia, the findings from the previous literature, reported in Chapter 5, showed that there is no consensus on the most suitable measure. There is concern that extra-welfarist measures, such as EQ-5D and SF-6D, may be unsuitable for this particular medical condition because of their narrow health-related focus. Evidence suggested that women do not consider menorrhagia to be solely a health related condition, as impact on social life, daily activities and family life/relationships were shown to be important to women with menorrhagia (Shaw et al., 1998). Hence these measures were purported to have poor face validity and women were found to be unsure whether the instruments should be completed with reference to general health or menorrhagia. The condition's periodic nature and the standard recall periods of these extra-welfarist measures also meant that the results would be dependent on the timing of assessment.

The first part of the empirical work was carried out, (and is reported in Chapter 6) to determine whether the use of the different extra-welfarist measures, EQ-5D and SF-6D, leads to the same recommendation regarding the most cost-effective treatment for menorrhagia. A model-based economic evaluation alongside the ECLIPSE trial was carried out from a National Health Service (NHS) perspective with a 2 year time horizon. A Markov model was developed using patient level data to accurately capture the associated changes in QoL and repetitive resource use associated with the pathways of the non-curative trial treatments, LNG-IUS and usual medical treatment. The results of the economic evaluation, which was first estimated using EQ-5D, and then re-estimated using SF-6D, showed that the most cost-effective treatment differed depending on the outcome measure used. Hence these measures were shown to capture different aspects of QoL and these findings highlighted the importance of using the most appropriate measure. Using EQ-5D, LNG-IUS was shown to cost more but was also more effective than usual medical treatment, generating an ICER of £1600 per QALY. This is within the NICE guidelines threshold of £20,000 per QALY for recommending new interventions into practice. Hence using EQ-5D, LNG-IUS was recommended as the first line treatment for menorrhagia and this recommendation was supported in all sensitivity analyses.

In contrast, the utilities derived from SF-6D presented results that suggested usual treatment dominated LNG-IUS, thus implying that usual treatment should be recommended as the first line treatment for menorrhagia. This result was again supported by all relevant sensitivity analyses. Therefore the recommendation to decision-makers was shown to differ depending on the instrument used. These findings coupled with those of the systematic review in Chapter 5, illustrated that there are concerns about the use of these extra-welfarist measures in menorrhagia, which therefore presented a case for the exploration of a measure from the alternative welfarist framework, namely WTP, which enables a broader assessment of wellbeing and may overcome the issue of timing of assessment.

The second part of the empirical work, (reported in Chapters 7 and 8) involved eliciting WTP for LNG-IUS (termed Mirena in the questionnaire) and usual medical treatment (termed Oral treatment in the questionnaire) to consider the use of WTP in menorrhagia in a feasibility study. The methodology is reported in Chapter 7 and explains that WTP was elicited from the theoretically preferred ex-ante perspective, where expected utility for the change is elicited, and also from the most commonly practiced ex-post perspective, where WTP for actual utility after the change is elicited. This was with the intention of drawing comparisons between the two perspectives. A questionnaire was developed for each perspective to elicit information on WTP, socio-demographic details and other information on QoL. A payment scale format was used and maximum WTP was elicited as a monthly out of pocket cost up until menopause. When WTP was elicited from the ex-ante perspective, women were presented with a scenario of menorrhagia and its treatments and were asked to provide a WTP value for both Mirena and Oral treatment. The scenario for menorrhagia and the expected outcomes of treatment were based on the disease-specific menorrhagia multi-attribute scale (MMAS) outcomes of the ECLIPSE trial data. Process utility was also included in the scenario for the relevant treatments, (Mirena and Oral treatment), to ensure that the results were not constrained to the time point of assessment. Hence the respondents were presented with information on the initial process of care in addition to the longer-term outcomes, therefore when future alternative treatments become available and if a similar scenario was presented, these WTP values could be compared against those presented in Chapter 7. The WTP values from the ex-post perspective were elicited from women who were enrolled in the ECLIPSE trial and had experience of menorrhagia and its treatment. Therefore, the WTP was for current treatment.

The analyses of WTP were reported in Chapter 8. Part 1 referred to the analyses of the ex-ante perspective, and Part 2 referred to the analysis of the ex-post perspective. In Part 1, Mirena was the preferred treatment but despite this, the mean WTP for Oral treatment (£15.38) was slightly greater than the mean WTP for Mirena (£15.11). It appeared that women who preferred Oral treatment were willing to pay 66% more for Oral treatment than Mirena, whereas women who preferred Mirena were willing to pay 28% more for Mirena than Oral treatment – a smaller difference in WTP. In Part 2 when WTP was elicited from the ex-post perspective, the women from the ECLIPSE trial who had experience of menorrhagia and either Mirena or Oral treatment were asked to complete a postal questionnaire and provide a WTP value for their current treatment. In contrast to the ex-ante perspective Mirena had the highest average WTP value with the WTP for Mirena (£31.08) being approximately twice as large as the WTP for Oral treatment (£17.45).

In both perspectives it was shown that respondents who provided WTP values were completing the questionnaire as is expected and required by analysts. The most commonly cited reason for WTP was related to the effects of treatment, hence respondents were focusing on the effects and value of the treatments to themselves, providing support for the use of WTP as a suitable measure of outcome for menorrhagia. In the regression analyses, reasons for WTP values, rather than socio-demographic information, were shown to be more important predictors of WTP in both the ex-ante and ex-post perspective. When WTP was presented against income it was shown that generally mean WTP increases with income but the relationship was not shown to be monotonic, which according to previous research would suggest that theoretical validity has not been demonstrated. However, the sample size and number of respondents in each income category could be the reason for this finding.

In Part 2 with the ex-post perspective, the association between measures was investigated. As baseline data from the ECLIPSE trial were available it was possible to assess the association between the ex-post perspective WTP and the change in EQ-5D against the change in disease-specific MMAS from baseline. The associations were assessed by Spearman's correlation and showed that there was a significant correlation between WTP and change in MMAS. However, a significant correlation was not observed between MMAS and EQ-5D, which would suggest that WTP is detecting certain effects, most likely the non-health effects, which EQ-5D does not capture.

Finally, in Chapter 9 a cost-benefit analysis (CBA) using ex-ante WTP comparing Mirena and Oral treatment was reported. This analysis helped to determine whether the most cost-beneficial intervention differs between the welfarist approach when WTP is used and the extra-welfarist approach when EQ-5D and SF-6D is used. To enable comparisons between the welfarist and extra-welfarist approaches, both economic evaluations were undertaken using a 2 year time horizon. As the opportunity to collect ex-post WTP data was available it was initially decided that an attempt should be made to compare the CBA results across perspectives. However upon reflection it was realised that a CBA from the ex-post perspective would not be theoretically recommended for a publicly funded healthcare system, in addition to the lack of sufficient data available. As decision-makers recommend societal values in a publicly funded healthcare system, because society contributes to funding the provision of healthcare, and theoretically the ex-ante perspective would be most suitable for a publicly funded healthcare system, the findings from the ex-ante CBA are likely to be more acceptable to decision-makers than those of the ex-post perspective. In both the base case and the sensitivity analysis, it was found that Oral treatment was the most cost-beneficial intervention as it was both less costly and resulted in a greater WTP value, hence producing the greatest welfare gain. The results of the CBA recommend Oral treatment as the first line treatment for women with menorrhagia. This finding of Oral treatment being the most cost-beneficial intervention concurred with the findings of the extra-welfarist SF-6D, but not the decision-maker recommended EQ-5D. The meaning of the findings is explored in more detail in section 10.5, but the findings across the measures demonstrate that both the SF-6D and WTP do not support the findings of the currently recommended EQ-5D measure.

10.3 Strengths and Limitations of Research

10.3.1 Strengths

The main strength of this research is that empirical primary data on the alternative instruments were collected and used throughout to value the outcomes in menorrhagia.

First, a comprehensive systematic review was carried out to determine which measures have been used and assessed in menorrhagia. The review illustrated the lack of consensus on the most suitable measure to use in menorrhagia and highlighted the issues associated with the currently used EQ-5D and SF-6D.

Second, to current knowledge, this study is the first to comprehensively assess all the existing evidence on outcomes for menorrhagia and carry out two contrasting economic analyses to encompass both a welfarist and extra-welfarist perspective. The conflicting findings observed from the comparison of the economic evaluation using alternative extra-welfarist measures EQ-5D and SF-6D in menorrhagia were based on primary data taken directly from the largest multi-centre randomised controlled trial carried out to date in the UK for menorrhagia. Further, the economic evaluation was carried out using a Markov model based on trial data and expert opinion, and developed using sophisticated methods. Hence the findings from a robust data set and analysis have illustrated that the selection of instruments to measure QoL in menorrhagia is very important as different decisions on cost-effectiveness were observed. This could potentially lead to policy-makers making decisions on interventions for clinical practice based on ambiguous data.

Third, the questionnaires eliciting WTP were developed and subsequently assessed by clinical experts in menorrhagia, by psychologists and health economists, increasing the content and face validity of the questionnaires. The scenarios presented to respondents in the questionnaire on menorrhagia and the effectiveness of the interventions were based on tangible evidence from the ECLIPSE trial, rather than expert opinion alone or expected outcomes as these may not actually be observed in practice. From this respect, providing WTP for actual observed outcomes increases the reliability of the findings. Further, the initial protocol for the empirical work for the WTP research was assessed by an international expert in the field of applying WTP in practice.

Finally, the research reported here is the first to: elicit WTP from an ex-ante and ex-post perspective; carry out a CBA from the ex-ante perspective comparing LNG-IUS to usual treatment in menorrhagia; compare SF-6D and EQ-5D in menorrhagia; and draw a comparison between the cost-effectiveness decision of the cost-utility analyses (CUA) and the CBA.

10.3.2 Limitations

The main limitations of this thesis are outlined in this section.

First, in the economic evaluation alongside the ECLIPSE trial, reported in Chapter 6, missing data in EQ-5D and SF-6D were not imputed using conventional methods such as multiple imputation due to the nature of the analysis. As the trial data are used in a decision model, it would be necessary to impute missing data at the individual level rather than overall at the QALY level for each time point. As this has not been attempted before and is not yet a recommended method, it did not seem appropriate. To some extent the probabilistic sensitivity analysis would compensate for not handling missing data.

Second, it was not possible to calculate a suitable sample size required for the WTP work, in Chapters 7-9, because a prior study in the area has not been carried out, therefore a value for what constitutes a meaningful change or difference in WTP has not been identified to be able to base the calculation upon. This meant that the study was not sufficiently powered to test for the statistical significance of the difference in WTP values. However, as the a-priori distribution of WTP values has been presented in this thesis, it can now be used to determine the required sample size in future studies.

Third, insufficient WTP data were available from the ex-post perspective to enable a comparison to be drawn between the WTP values of the ex-ante and ex-post perspectives. This also meant that a CBA from the ex-post perspective could not be carried out but it is likely that this may not have been entirely suitable as explained in the discussion of Chapter 9. Further, it could be argued that the ex-ante perspective WTP values and subsequent CBA do not entirely meet the criteria for an ex-ante perspective because men's values were not considered and women who do have experience of menorrhagia were also included. Therefore caution to these limitations must be considered, when drawing conclusions from this research.

Finally, there was a difficulty in recruiting women to attend interviews to more comprehensively assess reasons for WTP values and their understanding of the question. As outlined in Chapter 8, there are likely to be a multitude of reasons for respondents to not have attended the interviews. Some respondents stated that they had recently had a hysterectomy making attending an interview within the stated timeframe difficult. For others, the timing of the interviews (Easter holiday period) or unexpected inclement weather (multiple heavy snow days) could have represented insurmountable difficulties. Furthermore, attending an interview would have meant that women were required to take time off work which could expectedly lead to the poor response rate. In hindsight, a telephone interview may have been more appropriate.

10.4 Main Findings in Relation to Other Research

10.4.1 Comparison of EQ-5D and SF-6D

A difference in utilities observed between SF-6D and EQ-5D was observed in Chapter 6 and this led to the different cost-effectiveness decisions. Previous studies have similarly observed that SF-6D and EQ-5D generate different utilities for the same patient in seven other conditions (Brazier et al., 2004; Whitehurst & Bryan, 2011; Whitehurst et al., 2011). Brazier et al (2004) also observed these differences and found that one of the reasons for these differences is likely to be due to the ceiling effects associated with EQ-5D, as SF-6D is able to discriminate between health at the top end of the utility scale, which is anchored by 0 and 1. It was suggested that when full health is observed on EQ-5D but not SF-6D it is likely that the dimensions affected are those that are not comprehensively captured by EQ-5D such as vitality, mental health and physical functioning (Brazier et al., 2004). These additional SF-6D attributes are more likely to be those impacted by menorrhagia. Whilst the differences observed in this thesis are likely to be similar, there are likely to be additional specific reasons associated with the nature of menorrhagia, as discussed previously, but these reasons were not explored in this thesis.

10.4.2 WTP Research

Whilst this is the first study to elicit WTP for LNG-IUS and usual medical treatment, prior studies have attempted to elicit WTP in menorrhagia. A conjoint analysis study has been carried out in menorrhagia comparing the effectiveness of conservative surgery against hysterectomy (San Miguel et al., 2000). Cost was used as an attribute in the choice experiment to indirectly elicit WTP from patients with the condition. However, the purpose of the study by San Miguel et al (2000) is more related to testing the conjoint analysis method rather than attempting to assess the suitability of the measure in menorrhagia. Nonetheless it was demonstrated that the method could be used to assess WTP indirectly in the condition but that further research is required. The study by Ryan and San Miguel (2000), discussed in the systematic review in Chapter 5, also used WTP in menorrhagia, however again, not with the purpose of assessing outcomes in menorrhagia but instead to assess the reliability of WTP. In this study WTP was elicited directly using a payment card in a questionnaire based format to compare hysterectomy against conservative treatment from an ex-ante perspective. Hence no study has aimed to investigate the use of WTP in menorrhagia, instead the previous studies have used the data from women with menorrhagia to assess the methodology of WTP itself.

10.4.3 Economic Evaluations in Menorrhagia

The first CUA in menorrhagia was carried out by Sculpher (1998) and he compared surgical interventions for the condition. Given that, at the time, recommendations on the instrument to use were not presented by decision-makers, Sculpher considered EQ-5D but then used TTO to generate QALYs. Utilities were elicited for several condition-specific health state scenarios directly through TTO using patient values. It is not unexpected that one of the first economic evaluations in the area differs markedly to those carried out today. As demonstrated in the CUA in Chapter 6, economic evaluations are now required to use generic health state descriptions that are valued by society using EQ-5D.

In Chapter 4 it was demonstrated that all prior economic evaluations in the area that had focused on either LNG-IUS or usual medical treatment had been carried out from the extra-welfarist theoretical framework. As expected, given current recommendations, CUAs were carried out using either EQ-5D or SF-6D to generate the QALY outcome. Whilst LNG-IUS had been used as a comparator in these evaluations, it was often compared to surgical interventions and was not always found to be the most cost-effective intervention. A CBA was not identified. The prior evaluations were generally found to be of poor quality, based on secondary data and/or were conducted outside the UK, which has implications for the transferability of the findings. The CUAs carried out in Chapter 6 were the first to directly compare LNG-IUS against usual medical treatment using data from a robust trial. It was also the first to carry out the economic evaluation using the recommended EQ-5D and to re-estimate the evaluation using SF-6D to compare the cost-effectiveness results in menorrhagia. Other studies have assessed whether a difference in cost-effectiveness is observed in other conditions and similar to the findings presented in Chapter 6, they have also shown that the degree of cost-effectiveness, and the cost-effectiveness decision, can differ across the two measures (Davis et al., 2012; Sach et al., 2009).

Finally, to current knowledge the CBA reported in Chapter 9 is the first to compare LNG-IUS to usual medical treatment in menorrhagia. Another recent CBA has been carried out in the area of spinal surgery in Switzerland (Haefaeli et al., 2008). In this spinal study, the WTP was elicited from the ex-post perspective using patient values. Initially the payment scale method was considered but was not used because issues around starting point bias were observed in the pilot study. Thus WTP was elicited as an open ended and closed ended question, in addition to willingness-to-accept. The authors were able to successfully elicit the ex-post WTP and conduct the CBA, but suggested that methodological work should be carried out on the comparison of findings across the ex-ante and ex-post perspective, because the ex-ante method is recommended for publicly funded healthcare systems.

10.5 Reflections on Welfarist and Extra-welfarist Measures in Menorrhagia

In summary, in the context of menorrhagia, it has been shown that there are concerns around the use of the extra-welfarist measures EQ-5D and SF-6D. The findings from the welfarist approach demonstrate that it is feasible to elicit WTP and conduct a CBA within this condition.

The focus of this thesis was to explore the differences between the welfarist and extra-welfarist measures. The reasons behind the difference in utilities derived and cost-effectiveness results between the extra-welfarist measures were not explored as it was considered to be beyond the scope of the objective. However, potential reasons for these differences are likely to be due to the measures capturing different aspects of QoL, and the different recall periods presented in the questionnaires, amongst other things. SF-6D refers to health 'during the past month' which is likely to lead to an underestimation of the QALY. Whilst EQ-5D could lead to an under or overestimation of the QALY due to the recall period used of 'health today'. For example, if the woman is experiencing the symptoms on the day she completes EQ-5D, the QALY gain from treatment is likely to be overestimated, and underestimated if she completes the EQ-5D questionnaire when she is not experiencing symptoms. This is due to the calculation of the QALY, i.e. the linear interpolation of the value and the use of area under the curve to generate the QALYs.

What can be drawn from the evidence presented in this thesis is that the combined focus on health-related QoL and the specified time-frame used in extra-welfarist measures poses difficulties for completion and could result in misleading decisions about which treatment is most cost-effective. It has been shown in Chapter 5 that these measures do not capture what is important to women with menorrhagia and that the questions are presented in a manner that women find difficult to comprehend given the chronic, but episodic nature of the condition. However, the primary advantage of the use of extra-welfarist measures in menorrhagia is that they are currently recommended to generate QALYs. A QALY is argued to be the most suitable outcome due to its focus on health related QoL, which is considered most appropriate for decision-makers to allocate healthcare resources (NICE, 2008). Hence, decision-makers such as NICE have signed up to the extra-welfarist school of thought with its focus entirely on health and its aim to maximise health. It is increasingly recognised that this school of thought that focuses entirely on health is not suitable for certain conditions where benefits can additionally lie beyond health (McIntosh et al., 2010). However, it is important to state here that the findings presented in this thesis are a criticism of the current extra-welfarist measures EQ-5D and SF-6D and not all extra-welfarist multi-attribute utility instruments. But EQ-5D and SF-6D are the only recommended extra-welfarist measures in the UK to have a value set based on the UK adult population.

The advantage of the use of WTP is that it is based on economic theory and captures a broader assessment of wellbeing than extra-welfarist measures. In menorrhagia it is known that benefits beyond health are important to sufferers and that these benefits are unlikely to be captured using the current extra-welfarist measures (Shaw et al., 1998). This would therefore have a negative effect on policy implications for healthcare resource use as treatments may not be shown to be cost-effective and would not be recommended for clinical practice because an inappropriate measure was used to assess the benefits. Evidence for the use of WTP in menorrhagia has been presented in Chapter 8 and has shown that WTP is capturing aspects of life that are important to women, most likely the non-health benefits, that extra-welfarist measures do not capture. Furthermore, WTP values from both the ex-ante and the ex-post perspective were elicited and women were shown to focus on the value of the treatments, as is required. Very few protest responses were observed and a low non-completion rate provides an indication of the acceptability of the method from both perspectives. Hence the potential for the use of WTP in menorrhagia has been demonstrated but the reasons behind the WTP values and the respondents understanding of the exercise would need to be verified with the use of in-depth interviews.

The way in which WTP should be used is open to debate. In keeping with theory and decision-makers recommendations for the use of societal values, the ex-ante perspective WTP should be used. A further potential benefit of using the ex-ante perspective WTP is its ability to overcome the issue of timing of assessment. As these women are not experiencing the condition, the results are unlikely to be affected by the timing of assessment, which is the case when patients complete the extra-welfarist measures.

Despite these positive findings, there are obstacles against the use of WTP. As mentioned in Chapter 3, it is clear that the use of WTP does not currently ‘fit’ within the guidelines provided by decision-makers. A wide range of alternative extra-welfarist techniques are recommended prior to the consideration of any other information. Other methods are suggested in the guidelines but can only be presented as additional to information using the extra-welfarist approach. Decision-makers’ aversion to the welfarist method is due to the perception that WTP is based on ability to pay alongside on-going methodological issues (NICE, 2004). Practice is unlikely to change instantaneously but the gradual accumulation of evidence against the use of extra-welfarist measures in menorrhagia may encourage decision-makers to consider alternative techniques.

In relation to the findings of this thesis, it is not surprising that WTP and the change in the disease-specific measure MMAS from baseline are correlated. As WTP is elicited specifically for the treatment of the condition it will always inherently be more sensitive than EQ-5D and SF-6D, as these extra-welfarist measures are designed to capture general aspects of health. However, it is interesting that change in EQ-5D and change in MMAS did not have a significant correlation. This lack of correlation suggests that the changes that occur on the disease-specific measure are not comprehensively captured by EQ-5D. This finding could be excused because these extra-welfarist measures are not designed to capture disease-specific health. However, as mentioned previously, the issue with the extra-welfarist measures arises because menorrhagia is not primarily a health related condition, but healthcare resources are used to treat the condition. Therefore, these extra-welfarist measures are not comprehensively valuing outcomes in menorrhagia because they do not sufficiently capture the impact of the condition. It is the extent to which these measures do not capture these aspects that is important and with the remit of decision-makers broadening to include health and care excellence, this issue of how to incorporate benefits beyond health must be resolved.

A second issue is the episodic nature of the condition as it means that the timing of assessment will always arise as an issue when attempting to value outcomes in such conditions. Sculpher et al (1996) has previously argued that it is difficult to determine when the correct time would be to elicit values. Therefore the resolution to this problem is related to identifying which measure is considered to most comprehensively assess values in these episodic conditions. It is clear that the results from EQ-5D will be the most influenced by the timing of assessment due to its recall period. Given that decision-makers currently recommend extra-welfarist measures it would appear that SF-6D is less likely to be affected by the timing of assessment, as the measure will at least capture some information related to the condition. Yet it must be recognised that the extent to which it is able to detect attributes, i.e. non-health, relevant to the condition is limited. Whilst WTP captures the non-health benefits and is less likely to be affected by timing of assessment, it is argued that currently information from several measures should be provided to decision-makers to ensure the relevant information is fully captured.

Further, whilst in this thesis the decision-maker recommended EQ-5D shows a greater difference in QALYs between the two interventions, in favour of LNG-IUS, the SF-6D and WTP do not support this evidence. SF-6D and WTP suggest indifference between the treatments that is slightly more in favour of usual treatment. It is possible that the result of WTP and SF-6D agreeing is due to chance, or perhaps because SF-6D has a more comprehensive recall period than EQ-5D therefore it has captured some of what is important to women, as explained previously. Hence SF-6D is more sensitive and more closely follows the results of WTP. Although EQ-5D shows that LNG-IUS is more effective, which corresponds with the trial disease-specific measure, the correlation analysis showed that EQ-5D and the disease-specific measure are <u>not</u> correlated significantly and therefore the reasons for this agreement is not clear. So as the majority of the evidence suggests that there is a limited difference between the effectiveness of the two treatments, perhaps there is a case to take into account women's preferences rather than following a prescriptive list of recommendations. In menorrhagia, there may be scope for preference based treatment allocation as originally discussed by Sculpher (1998) and most recently mentioned by Roberts et al (2011). Sculpher discussed that there may be potential for a preference based treatment system in menorrhagia rather than assuming the 'all or nothing' approach to decision-making. Both usual treatment and LNG-IUS are already, and will be, used in clinical practice because LNG-IUS provides contraception, which is not required by all women. Furthermore, the randomised allocation of the ECLIPSE trial, and clinical practice, is based on whether women have a preference for contraception. This coupled with the relative indifference between treatments may mean that there may be scope for the use of preference-based treatment allocation in menorrhagia.

In summary, the three pieces of information provided by the three measures EQ-5D, SF-6D and WTP are providing decision-makers with different information and therefore the findings presented in this thesis would suggest that they should not be considered individually but rather altogether. The differences observed in the results are likely to be caused due to a number of factors, as all three measures have different elicitation methods, each measure captures different aspects of QoL and each measure refers to different recall periods. It is difficult to directly compare the measures for these reasons, although comparisons are often drawn between EQ-5D and SF-6D because they are used to generate the same QALY outcome. Therefore, the most suitable method of assessing the measures is by comparing the cost-effectiveness decisions and attempting to identify the potential reasons for these differences. Based on the current evidence it is difficult to categorically say which measure is more suitable than the other, therefore the information from the measures should be considered altogether, as explained further in section 10.6. WTP should currently be considered as a complement to these extra-welfarist measures, it cannot be recommended as the sole measure in menorrhagia because there is not currently sufficient evidence to support this. But a clear case has been presented in this thesis for WTP in menorrhagia because it more comprehensively overcomes the issues associated with extra-welfarist measures but further research exploring WTP within the area is required, particularly through the use of interviews.

10.6 Implications for Policy

Current recommendations from the NICE guidelines state that LNG-IUS should be implemented as the first line treatment for menorrhagia (NICE, 2007). The cost-effectiveness results of the economic evaluation carried out alongside the ECLIPSE trial using the decision-maker recommended EQ-5D correspond with the NICE guideline recommendations, as LNG-IUS was shown to be cost-effective. However, the recommendations from the NICE guidelines were not supported when SF-6D was used to generate QALYs in the economic evaluation. Hence the implications for clinical practice are dependent on the measure used, clearly indicating a need to ensure that the most suitable measure is used. It has been shown in the literature that neither measure is considered to be entirely suitable for menorrhagia. It should be noted that there was strong support from the principle investigators of the ECLIPSE trial to explore the use of WTP in menorrhagia because the shortcomings of the extra-welfarist measures in menorrhagia are recognised. This evidence suggests that there is a need for policy-makers to consider the findings from measures other than EQ-5D and SF-6D when evaluating interventions for menorrhagia. The potential for the use of WTP in menorrhagia has been demonstrated in this thesis, though further research is required. The findings from both SF-6D and WTP provide evidence against the guidelines issued in the UK that recommend the use of the extra-welfarist EQ-5D for use in all conditions.

NICE recommend that measures other than EQ-5D should not be used without justification (NICE, 2008). Their list of recommendations for alternative methods to valuing outcomes is presented in Box 3.1 of Chapter 3. In this section a case against the use of EQ-5D and the subsequent suggested methods in the guidelines is presented, similar to the manner required by NICE to justify the use of an alternative method;

1. Evidence of the limited psychometric properties and use of EQ-5D and SF-6D in menorrhagia have been presented in Chapter 5, suggesting that neither measure is able to fully capture the concerns and experiences of women with menorrhagia
2. EQ-5D and SF-6D have provided conflicting evidence in an economic evaluation alongside a trial, reported in Chapter 6.

3. The use of an alternative generic multi-attribute utility instrument would not be suitable given that a value set using a UK population has not been generated. Furthermore, a condition-specific multi-attribute utility instrument cannot be used as one has not been developed for menorrhagia.
4. Mapping from the condition-specific measure is unsuitable because the recommended measure that the condition-specific measure should be mapped on to is EQ-5D, which has been shown to have limited use in menorrhagia.

There is a need for policy-makers to develop a standard framework for evaluating conditions where the benefits can lie beyond health, as EQ-5D and SF-6D are known to be problematic in these cases. The use of WTP in menorrhagia has been demonstrated. However, whilst the research in this thesis shows the feasibility of using WTP in menorrhagia, further research does need to be carried out within the area. Until additional evidence in support of WTP is available, it is recommended that a cost-consequence analysis be used as each measure is providing additional information that should be taken into consideration. Therefore, due to limited evidence, the recommended approach is neither a welfarist nor extra-welfarist approach, instead it is a decision-maker approach where decision-makers are presented with all of the relevant information that is currently available in a cost-consequence analysis, as shown in Table 10.1. Thereby enabling decision-makers to make decisions about resource allocation in menorrhagia with a more informed opinion.

Table 10.1 Presentation of a cost-consequence analysis

Treatment	EQ-5D	SF-6D	WTP (ex-ante)	Cost
LNG-IUS	1.580	1.198	365.08	430
Usual treatment	1.513	1.200	371.67	330
Difference	0.070	0.002	6.59	100
Overall	ICER: 1600	Dominates	INB: -110	-
Cost-effective treatment	**LNG-IUS**	**Usual treatment**	**Usual treatment**	**-**

LNG-IUS; levonorgestrel-releasing intrauterine system, ICER; incremental cost-effectiveness ratio; WTP; willingness-to-pay

10.7 Further Research Recommendations

Further research is required in a number of areas to further advance the field. Whilst the feasibility of using WTP in menorrhagia has been shown, it is recognised that subsequent studies with larger sample sizes will need to be carried out to validate the findings. Additionally, it may be beneficial to carry out some in-depth interviews to determine the respondents understanding of the elicitation exercise. There are also still some methodological issues that need to be resolved before WTP will be implemented on a wider scale. The research around the methodology of WTP has shown that the WTP values are sensitive to the elicitation format and also to the use of an interview or questionnaire. Further, the sensitivity of WTP to the size of the good has also been questioned. Whilst sensitivity was shown in this thesis, the issue needs to be assessed on a wider scale. Even the method by which WTP should be assessed is debated i.e. whether it be through psychometric testing or whether theoretical validation is more important (O'Brien & Gafni, 1996). It is agreed amongst health economists that these differences exist but a framework has yet to be produced that outlines how to move forward within the area. For the use of WTP to progress in menorrhagia, and in other conditions, there first has to be in place a set of guidelines or recommendations for how contingent valuation techniques should be used in health care, which analysts should follow. Without such guidelines, it will be difficult to convince decision-makers of the merits of WTP due to the methodological differences observed in studies.

An alternative welfarist method, which is still in its infancy in health economics, could be considered in menorrhagia. The stated preference discrete choice experiment is gaining increased credibility for valuing outcomes. It appears to have so far received a more welcoming reception from health economists to the conventional contingent valuation method as WTP values can be elicited indirectly for changes in separate and in groups of attributes (Ryan & Gerrard, 2003). It has been argued to be able to overcome some of the shortfalls of direct elicitation of WTP (Hanley et al., 2001). However, methodological work does need to be carried out to establish this.

Next steps for extra-welfarist measures include research into identifying the drivers behind the differences between the findings of SF-6D and EQ-5D. More in-depth analyses could be carried out to assess the sensitivity to change of EQ-5D and SF-6D compared to the disease-specific measures MMAS. It would also be interesting to identify which domains on SF-6D are affected by the condition compared to the domains affected by EQ-5D by using the changes in response to changes in MMAS. Similarly, it will be important to assess whether a meaningful change in MMAS is reflected in both the extra-welfarist measures and whether this change is matched by the respondents' report of general health.

Further, as the EQ-5D-5L has now been developed (Herdman et al., 2011) to help overcome the issues with ceiling effects of the EQ-5D-3L, there may be scope for testing this measure against WTP, though the results are likely to be similar to those found here as the recall period remains the same and the non-health benefits are not captured. The amended EQ-5D measure developed and briefly touched on in this thesis could also be more thoroughly assessed. Rather than using a recall period of ‘health today’, a period of ‘during your cycle is used’ to determine to what extent this change improves the sensitivity of the measure, as an amended EQ-5D is likely to be more readily accepted by decision-makers.

10.8 Conclusion

The aim of this thesis was to draw a comparison between welfarist and extra-welfarist approaches to valuing outcomes in menorrhagia. To meet this aim, first a CUA was carried out using the extra-welfarist measure EQ-5D and was re-estimated using SF-6D; second, a questionnaire was developed to elicit the welfarist WTP from an ex-ante and an ex-post perspective; and third a CBA was carried out using the WTP values from the ex-ante perspective. The findings between the alternative extra-welfarist measures were compared, and these, in turn, were compared with the CBA results. The success of the completion of WTP was also assessed. The key contributions of this thesis are that it has been demonstrated that in the case of a chronic condition, with episodic symptoms, current recommendations from decision-makers on which outcome measures should be used are not supported by evidence and this has significant implications for clinical practice. Two other measures, SF-6D and WTP, did not support the findings of the recommended EQ-5D measure suggesting that it should not be used alone in this condition. A second contribution is that the potential for the use of the somewhat controversial welfarist WTP measure was also demonstrated, as it was shown to have an association with changes in the disease-specific measure whilst EQ-5D did not. Several concerns around the use of the EQ-5D and SF-6D extra-welfarist measures in menorrhagia have been shown. The feasibility of the use of the welfarist WTP also demonstrated. However, there is still much scope for further research to be carried out to understand what aspects of QoL, if any, the current extra-welfarist measures are capturing, and to solidify the need to consider the use of a welfarist measure, such as WTP, in a condition like menorrhagia.

APPENDIX 1. SYSTEMATIC REVIEW OF ECONOMIC EVALUATIONS: SEARCH TERMS AND QUALITY ASSESSMENT OF STUDIES

A1.1 Search Terms

Ovid Medline

1. menorrhagia.mp.
2. exp Menorrhagia/
3. exp "cost"/
4. Cost$.mp.
5. heavy menstrual bleeding.mp.
6. HMB.mp.
7. 1 or 2 or 5 or 6
8. economic$.mp.
9. exp economics/
10. exp "Costs and Cost Analysis"/
11. 3 or 4 or 8 or 9 or 10
12. levonorgestrel releasing intrauterine system.mp.
13. LNG-IUS.mp.
14. exp tranexamic acid/
15. tranexamic acid.mp.
16. exp mefenamic acid/
17. mefenamic acid.mp.
18. exp progestogen/
19. norethisterone.mp.
20. progestogen.mp.
21. levonorgestrel intrauterine system.mp.
22. oral contraceptive.mp.
23. 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
24. 7 and 11 and 23
25. limit 24 to yr="2006 -Current"

EMBASE

1. menorrhagia.mp.
2. exp Menorrhagia/
3. exp "cost"/
4. Cost$.mp.
5. heavy menstrual bleeding.mp.
6. HMB.mp.

7. 1 or 2 or 5 or 6
8. economic$.mp
9. exp economics/
10. exp "Costs and Cost Analysis"/
11. 3 or 4 or 8 or 9 or 10
12. levonorgestrel releasing intrauterine system.mp.
13. LNG-IUS.mp.
14. exp tranexamic acid/
15. tranexamic acid.mp.
16. exp mefenamic acid/
17. mefenamic acid.mp.
18. exp progestogen/
19. norethisterone.mp.
20. progestogen.mp.
21. levonorgestrel intrauterine system.mp.
22. oral contraceptive.mp.
23. 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
24. 7 and 11 and 23
25. limit 24 to yr="2006 -Current"

PsychInfo

1. menorrhagia.mp.
2. Cost$.mp.
3. heavy menstrual bleeding.mp.
4. HMB.mp.
5. 1 or 3 or 4
6. economic$.mp
7. exp economics/
8. exp "Costs and Cost Analysis"/
9. 2 or 6 or 7 or 8
10. levonorgestrel releasing intrauterine system.mp.
11. LNG-IUS.mp.
12. tranexamic acid.mp.
13. mefenamic acid.mp.
14. norethisterone.mp.
15. progestogen.mp.
16. levonorgestrel intrauterine system.mp.
17. oral contraceptive.mp.
18. 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17
19. 5 and 9 and 18
20. limit 19 to yr="2006 -Current"

Science Citation Index And Social Science Citation Index (Web of Science)

1. TS= (Menorrhagia OR “heavy menstrual bleeding” OR HMB)
2. TS= (Cost$ OR "Cost analysis" OR economic$)
3. TS=("Levonorgestrel releasing intrauterine system" OR "Levonorgestrel intrauterine system" OR "LNG-IUS" OR "tranexamic acid" OR "mefenamic acid" OR "Oral contraceptive" OR progestogen OR norethisterone)
4. #3 AND #2 AND #1

DARE and NHS EED (Cochrane library)

1. MeSH descriptor: [Costs and Cost Analysis] explode all trees
2. MeSH descriptor: [Menorrhagia] explode all trees
3. MeSH descriptor: [Economics] explode all trees
4. MeSH descriptor: [Tranexamic Acid] explode all trees
5. MeSH descriptor: [Mefenemaic Acid] explode all trees
6. MeSH descriptor: [Contraceptives, Oral] explode all trees
7. MeSH descriptor: [Progestins] explode all trees
8. Levonorgestrel releasing intrauterine system:ti,ab.kw (Word variations have been searched)
9. Levonorgestrel intrauterine system: ti,ab.kw (Word variations have been searched)
10. LNG-IUS: ti,ab.kw (Word variations have been searched)
11. Mefenamic acid or tranexamic acid or oral contraceptive$ or progestogen or norethisterone: ti,ab.kw (Word variations have been searched)
12. HMB or heavy menstrual bleeding: ti,ab.kw (Word variations have been searched)
13. Cost$ or economic$: ti,ab.kw (Word variations have been searched)
14. (#2 or #12) and (#1 or #3 or #13) and (#4 or #5 or #6 or #7 or #8 or #9 or #10 or #11) From 2006-2013, in other reviews and economics evaluations

A1.2 Assessment of quality of studies included in systematic review of economic evaluations in menorrhagia

Study (Country)	Assessment of quality
You et al 2006 (Hong Kong)	• The cycle length of one year is not suitable – does not capture all relevant changes in QoL and resource use • Used median utility values - mean is deemed to be more suitable as it accounts for all individuals equally (Roberts et al., 2011) • Data source for utility values – Hong Kong study using UK utility values – TTO using patient values • Reason for using triangular distribution in the PSA instead of more typically observed distribution was not specified
Brown et al 2006 (New Zealand)	• A decision tree was used – does not allow for recurrent events which are required given the chronic nature of the condition • No QALYs reported • Uncertainty not comprehensively assessed – a PSA was not carried out, particularly important given small sample size of study (n=70) • Data on QoL was collected at baseline and 24 months only – due to the non-curative nature of the non-surgical treatments this frequency is insufficient to capture changes in QoL. Despite resource use being captured at 3, 12 and 24 months there was no explanation for not collecting QoL
Clegg et al 2007 (UK)	• Combined TTO using patient values and EQ-5D valuation methods in the same analysis – different valuation methods provide different results. Additionally, TTO was valued using patient values and EQ-5D was valued using general population values, the utilities are not comparable • Combined utility values from UK (TTO) and Finland (EQ-5D) in the same analysis – utility value sets are not transferable across countries and a combination is likely to result in a poor evaluation • Used a combination of mean and median utility values in the same analysis– mean is deemed to be more suitable as it accounts for all individuals equally. The use of both without justification indicates a poor assessment of the data sources • Where unavailable utility values were estimated from other states – some estimation methods are unclear • Used rates and outcomes for hysterectomy from Finnish study, despite acknowledging that the use of hysterectomy in UK clinical practice is no longer recommended or as commonly used as the time the study was carried out.
NICE 2007 (UK)	• A comprehensive assessment of uncertainty of results was not carried out, which is particularly important given the poor data used. Only a one-way sensitivity analysis was carried out, not a PSA • Based on Finnish study and used these utility values for a UK population – the value sets are not transferable • No direct comparisons of treatments were carried out, despite the use of a decision model – incremental differences need to be assessed to determine which treatment is most cost-effective relative to another • Two treatments, oral and injected progestogen, were recommended for clinical practice without any evidence of cost-effectiveness
Roberts et al 2011 (UK)	• Combined TTO using patient values and EQ-5D valuation methods in the same analysis – different valuation methods provide different results. Additionally, TTO was valued using patient values and EQ-5D was valued using general population values, the utilities are not comparable • Combined utility values from UK (TTO) and Finland (EQ-5D) in the same analysis – utility value sets are not transferable across countries and a combination does not produce a robust evaluation *However, a comprehensive assessment of uncertainty associated with the model inputs was carried out in the PSA*

Study (Country)	Assessment of quality
Lete et al 2011 (Spain)	• Claimed to use EQ-5D because it is validated, but actually used TTO from Sculpher (1998) study which was based on patient values – limits credibility and confidence in the analysis • TTO values taken from the UK population for a Spanish study – utility value sets are not transferable across countries • Used median utility values - mean is deemed to be more suitable as it accounts for all individuals equally • Six month cycle for model is not suitable- does not capture all associated changes in QoL and resource use • PSA distribution was not specified for utilities – given the utility data used was also from a different country it is particularly important to assess uncertainty, using a PSA, in the utility data
Ganz et al 2013 (US)	• Utilities were taken from a Finnish study that used EQ-5D for a US population - utility value sets are not transferable across countries • Data comes from women who have at baseline bleeding greater than 80ml per cycle – Population group does not reflect current women who are treated with menorrhagia as an objective assessment of blood loss is no longer considered suitable. • Probabilities of treatment success of oral and LNG-IUS were taken from studies that recruited women who were bleeding in excess of 80ml and utilities were taken from the Finnish study which did not use the objective assessment – utilities may differ between those who subjectively complain of menorrhagia and those who are objectively bleeding in excess of 80ml per cycle. Women who bleed in excess of 80ml are also likely to suffer with anaemia which may impacts QoL and the probabilities of treatment success may not be the same for those who bleed in excess compared to those recruited due to a subjective assessment of impact on QoL • Applied utility for successful LNG-IUS to all treatments – LNG-IUS is likely to be associated with a higher utility value than other treatments because it stops bleeding whereas others only reduce bleeding hence assuming utility for successful LNG-IUS is the same for all treatments is not a suitable assumption

LNG-IUS; levonorgestrel-releasing intrauterine system, PSA; probabilistic sensitivity analysis, QoL; quality of life, TTO; Time trade off.

APPENDIX 2. SYSTEMATIC REVIEW OF OUTCOME MEASURES: SEARCH TERMS AND DATA EXTRACTION TABLES

A2.1 Search Terms

Ovid Medline

1. exp Menorrhagia/
2. Menorrhagia.mp.
3. Heavy menstrual bleeding.mp.
4. HMB.mp.
5. exp questionnaires/
6. questionnaire$.mp.
7. exp quality of life/
8. quality of life.mp.
9. QoL.mp.
10. Exp Outcome assessment (Healthcare)/
11. Outcome assessment.mp.
12. Exp psychological test/
13. Psychological test$
14. Exp Interview, Psychological/
15. Interview$
16. Exp psychometrics/
17. Psychometric$
18. 1 or 2 or 3 or 4
19. 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17
20. 18 and 19
21. Limit 20 to (humans and female)

Source – EMBASE Classic + EMBASE (Ovid)

1. exp menorrhagia/
2. menorrhagia.mp.
3. heavy menstrual bleeding.mp.
4. HMB.mp.
5. exp questionnaire/
6. questionnaire$.mp.
7. exp "quality of life"/

8. quality of life.mp.
9. QoL.mp.
10. exp outcome assessment/
11. outcome assessment.mp.
12. exp psychologic test/
13. psychological test$.mp.
14. exp interview/
15. interview$.mp.
16. exp psychometry/
17. psychometric$.mp.
18. 1 or 2 or 3 or 4
19. 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17
20. 18 and 19
21. limit 20 to (human and female)

CINAHL Plus Cumulative Index to Nursing, Allied Health Literature (EBSCO)

1. (“MH “Menorrhagia”) OR Menorrhagia OR heavy menstrual bleeding OR HMB
2. ("MH "Psychological Tests+) OR (MH "Questionnaires+") OR "questionnaire* OR (MH "Quality of Life+") OR (psychological test*) OR MH "Outcome Assessment") OR (outcome assessment) OR psychometric* OR (MH "Interviews+") OR interview* OR (MH "Psychometrics") OR QoL"
3. #1 AND #2
4. Limit 3 to Human and Female

Source – PsycInfo (Ovid)

1. Menorrhagia.mp.
2. heavy menstrual bleeding.mp.
3. HMB.mp.
4. exp Questionnaires/
5. questionnaire$.mp.
6. exp "Quality of Life"/
7. quality of life.mp.
8. QoL.mp.
9. outcome assessment.mp.
10. exp Psychometrics/
11. psychometric$.mp.
12. psychological test$.mp.
13. interview$.mp.
14. exp interviews/
15. 1 or 2 or 3

16. 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
17. 15 and 16
18. limit 17 to (human and female)

Source – Science Citation Index (Web of Science)

1. TS= (Menorrhagia OR "heavy menstrual bleeding" OR HMB)
2. TS=(questionnaire$ OR QoL OR "quality of life" OR "outcome assessment" OR "psychological test$" OR interview$ OR psychometric$)
3. 1 AND 2

Source – Social Science Citation Index (Web of Science)

1. TS= (Menorrhagia OR "heavy menstrual bleeding" OR HMB)
2. TS=(questionnaire$ OR QoL OR "quality of life" OR "outcome assessment" OR "psychological test$" OR interview$ OR psychometric$)
3. 1 AND 2

NHS EED National Health Service Economic Evaluation Database (Cochrane Library – Wiley)

1. interview OR psychometric in Technology Assessments and Economic Evaluations
2. questionnaire OR QoL OR quality of life OR outcome assessment OR psychological test in Technology Assessments and Economic Evaluations
3. heavy menstrual bleeding OR HMB OR menorrhagia in Technology Assessments and Economic Evaluations
4. MeSH descriptor **Menorrhagia** explode all trees
5. MeSH descriptor **Questionnaires** explode all trees
6. MeSH descriptor **Quality of Life** explode tree 1
7. MeSH descriptor **Outcome Assessment (Health Care)** explode all trees
8. MeSH descriptor **Psychological Tests** explode all trees
9. MeSH descriptor **Interviews as Topic** explode all trees
10. MeSH descriptor **Psychometrics** explode all trees
11. ((#3 OR #4) AND (#1 OR #2 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10))

Source – HTA Health Technology Assessment (Cochrane Library –Wiley)

1. interview OR psychometric in Technology Assessments and Economic Evaluations
2. questionnaire OR QoL OR quality of life OR outcome assessment OR psychological test in Technology Assessments and Economic Evaluations
3. heavy menstrual bleeding OR HMB OR menorrhagia in Technology Assessments and Economic Evaluations
4. MeSH descriptor **Menorrhagia** explode all trees
5. MeSH descriptor **Questionnaires** explode all trees
6. MeSH descriptor **Quality of Life** explode tree **1**
7. MeSH descriptor **Outcome Assessment (Health Care)** explode all trees
8. MeSH descriptor **Psychological Tests** explode all trees
9. MeSH descriptor **Interviews as Topic** explode all trees
10. MeSH descriptor **Psychometrics** explode all trees
11. ((#3 OR #4) AND (#1 OR #2 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10))

A2.2 Results (Data Extraction) Tables

Table 1. Study characteristics for economic measures that have been assessed for psychometric properties

Study	Comparator	Measure assessed					Psychometric property assessed			Comment
		SF-36	EQ-5D	MMAS	WTP	other	Validity	Sensitivity	Reliability	
Clark & Gupta 2004	Thermal balloon ablation		✓	✓				✓		MMAS was statistically significantly associated with satisfaction (p=0.001), whilst EQ-5D was not (p=0.08).
Coulter et al 1994	Surgery vs. non-surgery	✓				✓		✓		• Other measures: social impact (SI) score & satisfaction. • Improvement in QoL is generally reflected across all questionnaires. • Discrepancies were seen in menorrhagia specifically. SI shows moderate improvement for no surgery group, SF-36 shows a small improvement in 2 domains only • Those satisfied with treatment and with moderate menorrhagia, showed improved SI scores on social functioning and energy in SF-36. With Little to no influence on general health status. The attributes general health, role physical and physical functioning either showed a small change or no difference.
Garratt et al 1993	4 conditions vs. general population	✓					✓		✓	• In a group of patients with 4 different conditions assessed reliability through internal consistency (cronbach's alpha); validity through confirmatory factor analysis; & construct validity • For all conditions together, internal consistency of SF-36 was observed on all 8 scales. Factor analysis confirmed relevance of 8 scales: Physical functioning was most relevant in the group with 4 conditions. For menorrhagia specifically physical functioning was not statistically significantly different to the general population.
Garratt et al 1994	General population vs. 4 conditions	✓						✓		• In line with responses to SF-36 transition question mean improvements in all health states were observed. • Discrepancies were seen in menorrhagia specifically. • For the answer my health is much better: large improvements were observed on SF-36 for pain, energy, role physical, social functioning domains, moderate responses were observed for

Study	Comparator	Measure assessed					Psychometric property assessed			Comment
		SF-36	EQ-5D	MMAS	WTP	other	Validity	Sensitivity	Reliability	
										the remaining attributes except physical functioning which was small. • For the answer my health is somewhat better: the pain, energy, role physical, social functioning domains showed a moderate improvement (social and energy were borderline), the remaining domains showed a small improvement. • For the answer my health is the same: a small change was observed for pain, energy, role physical, social functioning domains • For the answer my health is somewhat worse: a small decrease was observed in for physical functioning, energy, role physical, social functioning domains. The general health attribute improved to moderate. • For my health is worse: a moderate reduction was observed for 2 attributes, and the reductions were small for the remaining attributes. General health reduced to negative-small and pain improved
Gokyildiz et al 2013	Menorrhagia population vs. population without disease	✓						✓		Adapted to Turkish scale. All domains of the SF-36 are significantly different in the menorrhagia population compared to the population without a condition
Habiba et al 2010	SF-36 (MCS & PCS) vs. MMAS	✓		✓						• Many women may find SF-36 inappropriate or difficult to answer – face validity is questioned • MCS/PCS were not statistically significantly improved even though some dimensions of SF-36 were significant. Therefore authors argue PCS/MCS scoring procedure may not accurately summarise subscales • A statistically significant improvement in SF-36 was observed in MCS only and MMAS. At baseline statistically significant association in MMAS and need for surgery. This association was observed for the SF-36 attributes role physical, pain, social functioning and mental health - but not for PCS or MCS.
Hehenkamp et al 2008	UAE vs. hysterectomy	✓	✓			✓				Other measure: HUI3. The SF-36 MCS was more sensitive than the PCS as it was significantly associated with patient satisfaction (p=0.01),

Study	Comparator	Measure assessed					Psychometric property assessed			Comment
		SF-36	EQ-5D	MMAS	WTP	other	Validity	Sensitivity	Reliability	
										whilst the PCS was not (p=0.191). Significant differences in QoL between treatment groups were observed in the PCS only.
Jenkinson et al 1996	Results vs. general population	✓					✓		✓	• Interviews revealed that women are unsure whether to complete for general health or periods. Therefore issues with face validity. Cyclical symptoms leads to ambiguity with questions with time frames. • Mental health, general health and social functioning showed lower internal reliability than Oxford general population, especially mental health and general health
Pattison et al 2011	LNG-IUS vs. oral treatment	✓	✓	✓			✓			• Overall score for EQ-5D had a poor correlation with MMAS. MMAS was shown to have face validity, convergent & discriminant validity, and test retest reliability. • Authors stated general health and Physical functioning attributes of SF-36 are heavily weighted to mobility and self-care which are not affected by menorrhagia, which are inappropriate items on SF-36. General health and EQ-5D were argued not to be precise enough to measure differences in patient groups as they refer to general condition rather than cycle
Ruta et al 1994	general population vs. 4 conditions	✓							✓	Assessed reliability through internal consistency using cronbach's alpha & test re-test. Found to be reliable for groups of patients only not individual patient management
Ryan et al 2000	conservative vs. hysterectomy : preference & WTP				✓				✓	Inconsistent responses (30%) between treatment preference and maximum WTP
Shaw et al 1998	Development & assessment			✓			✓			Main areas of concern: physical health, psychological health, family, social, work life, practical difficulties. The attributes ‘ Family life/relationships’ and ‘Physical health and wellbeing’ were given the greatest weightings

4 conditions - group of back pain, varicose veins, peptic ulcers, menorrhagia

HUI3; Health Utility Index mark 3, LNG-IUS; levonorgestrel-releasing intrauterine system, MCS; mental composite scale, MMAS; menorrhagia multi-attribute scale, PCS; physical composite scale, QoL; quality of life, WTP; willingness-to-pay.

Table 2. Study characteristics for economic measures in an economic evaluation or cost study

Study	Comparator	Economic measure				Non-economic measure					Comment
		SF-36	EQ-5D	VAS	other	PBAC	Ruta scale	Satis-faction	Anx & dep	other	
Brown et al 2006	LNG-IUS vs TBEA	✓				✓					Authors comments:SF-36 is insensitive. WTP should be used instead. (cost study)
Frick et al 2009	Hysterectomy vs EA	✓	✓	✓			✓				Cost study
Hurskainen et al 2001	LNG-IUS vs Hysterectomy		✓	✓	✓			✓	✓	✓	EQ-5D was the primary outcome – validated and universally accepted. Affect on sexual life was the other non-economic measure.
<u>5 yr follow-up:</u> Hurskainen et al 2004											
Kennedy et al 2003	Decision aid: interview vs leaflet vs control	✓	✓				✓	✓	✓		Authors comments: All dimensions on SF-36 were unlikely to show a difference between groups, but generic measures allow comparisons across studies. validated and universally accepted
Kilonzo et al 2010	MEA vs TBEA		✓								Authors comments: Sensitivity of EQ-5D in menorrhagia is questionable
Sculpher et al 1993	TCRE vs hysterectomy			✓				✓			A formal economic evaluation was not carried out. Costs and outcomes were presented separately
<u>2 yr follow-up:</u> Sculpher et al 1996	ER vs hysterectomy	✓		✓				✓			SF-36 validated and universally accepted
Sculpher, 1998	Hysterectomy vs TCRE				✓						TTO used. Authors comments: Menorrhagia is a chronic, episodic, condition and QoL results will be dependent on the timing of assessment. TTO widely used
tai pale et al 2009	Hysterectomy vs no treatment				✓						Utilities were derived using 15D
Van der wilt et al 2005	Blood loss chart vs no chart	✓				✓				✓	Menorrhagia severity index. SF-36 domains: physical functioning, vitality, pain, health perception - validated
Vuorma et al 2004	Decision aid booklet			✓	✓			✓	✓	✓	RAND-36: universally accepted. Other sex and inconvenience

Anx & dep; anxiety and depression, EA; Endometrial ablation, ER; endometrial resection, LNG-IUS; Levonorgestrel intrauterine system, MEA; Microwave endometrial ablation, PBAC; Pictorial blood assessment chart, QoL; quality of life, TBEA; Thermal balloon endometrial ablation, TCRE; Transcervical resection of the endometrium, TTO; time trade off, WTP; willingness-to-pay.

Table 3. Study characteristics for economic measures as a primary outcome in an effectiveness study

Study	Comparator	Economic measure				Non-economic measure				Comments
		SF-36	EQ-5D	VAS	other	PBAC	Satis-faction	Dep & Anx	other	
Bongers et al 2005	Balloon EA vs Bipolar EA	✓						✓	✓	Authors comments: SF-36: insensitive to treatment effect & difficult to answer. Other non-economic measure: Structured clinical history questionnaire &Rotterdam Symptom Checklist
Busfield et al: main study for Brown	LNG-IUS vs TBEA	✓				✓	✓		✓	Other non-economic measures: questions on menstrual symptoms
De Souza et al 2010	Effect on QoL	✓				✓			✓	SF-36 Mental composite scale and physical composite scale considered separately. Non-economic other: haemoglobin levels
Gorgen et al 2008	LNG-IUS			✓		✓			✓	Authors comments: VAS is too simplistic. Other non-economic measure: Daily life, libido, pain
kupperman et al 2004	Hysterectomy vs medical treatment	✓		✓					✓	SF-36 Mental composite scale and physical composite scale considered separately. Other non-economic measure: sexual functionings, sleep problems
Matsumoto et al 2007	Combined oral contraceptive				✓					Other economic measure: WHO-QoL
Olah et al 2005	EA			✓°						
Shawki et al 2009	LNG-IUS		✓			✓				Authors comments: EQ-5D is appropriate – changes in blood loss were reflected in EQ-5D

Dep & Anx; depression and anxiety, EA; endometrial ablation, LNG-IUS; Levonorgestrel intrauterine system, PBAC; pictorial blood assessment chart, QoL; quality of life, TBEA; thermal balloon endometrial ablation, VAS; visual analogue scale.

Table 4. Study characteristics for economic measures as secondary outcomes in effectiveness studies

Study	Comparator	Economic measures				Non-economic measures					Comments
		SF-36*	EQ-5D	VAS	other	PBAC	Satis-faction	Anx & dep	MMAS	other	
Abbott et al 2003	TBEA vs. Bipolar EA	✓	✓	✓		✓	✓			✓	Other non economic measure: sexual life
Chadha et al 2000	effectiveness of national guidelines	✓								✓	Authors comments: The only domains of SF-36 that are most likely to show a change are general health, role physical, mental health, social functioning. Other non-economic: condition specific questions
Clark et al 2011	Bipolar EA vs. TBEA		✓	✓			✓		✓	✓	Other non-economic measures: Amenorrhoea, acceptability, menorrhagia outcomes questions, sexual activity
Cooper et al 1997	Med treatment vs. TR	✓						✓		✓	Other non economic measures: Clinical questions, Bleeding and Pain Score
2 yr follow-up: Cooper et al 1999											
5 yr follow-up: Cooper et al 2001											
Cooper et al 1999	MEA vs TCRE	✓					✓			✓	Authors comments: Health related QoL should be used as definitive indicator of treatment success. Other non-economic measures: Acceptability (bleeding and pain)
2 yr follow-up: Bain et al 2002		✓					✓			✓	
5 yr follow-up: Cooper et al 2005		✓					✓			✓	
10 yr follow-up: Sambrook et al 2009		✓					✓			✓	
Cooper et al 1997	preference based vs non-preference based treatment	✓						✓		✓	Other non-economic clinical questions
Crosignani et al1997	ER vs hysterectomy	✓					✓	✓		✓	Other non-economic measure: impact on sexual life
Crosignani et al1997	ER vs LNG-IUS	✓				✓	✓				
Davis et al 2000	Oral contraceptive vs placebo	✓				✓				✓	Authors comments: SF-36: Physical health domain is most likely to be affected (self-care walking climbing stairs). Other non-economic condition-specific clinical

Dickersin et al 2007	Hysterectomy vs EA		✓							✓	Other non-economic measures: bleeding, pain, fatigue, sexual function, employment, housework, leisure activities
Edwards et al 2004	SDM vs risk communication skills	✓			✓		✓	✓			Other economic measure: Comrade questionnaire
Gupta et al 2013	LNG-IUS vs usual medical treatment	✓	✓	✓					✓	✓	Other non-economic measure: sexual activity questionnaire
Hawe et al 2003	TBEA vs YAG Laser	✓	✓	✓			✓			✓	EQ-5D was lower at baseline in one arm compared to another, this was not reflected in SF-12. SF-12 or SF-36: MCS and PCS separately. Other non-economic measures: Amenorrhoea, acceptability and sexual activity
Van der Kooji et al 2010 (**Hehenkamp follow-up**)	UAE vs hysterectomy	✓				✓	✓				Other non-economic measures: elimination of menorrhagia., sexual activity, body image,
Helal et al 2011	TBEA			✓			✓			✓	Other non-economic measures: volume of blood loss
Henshaw et al 2002	MEA vs LNG-IUS	✓					✓			✓	Other non-economic measures: acceptability
<u>5 year follow up to Bongers</u>: Kleijn et al 2007	LNG-IUS vs ER	✓					✓				
Malak & Shawki 2006	decision aid			✓		✓					EQ -VAS: Well-being, work performance, physical and sexual activity
Protheroe et al 2007	TBEA vs MEA							✓	✓	✓	Other non-economic measures: decision conflict scale
Ramazanzadeh et al 2012	LNG-IUS vs copper IUD	✓					✓			✓	Other non-economic measures: menstrual history
Sambrook et al 2009	Bipolar EA vs Balloon EA	✓	✓	✓		✓	✓	✓		✓	Other non economic measure: pain
Samuel et al 2009	outpatient EA vs inpatient EA						✓		✓	✓	Other non economic measure: general health

*****SF-36 or SF-12. Dep & Anx; depression and anxiety, EA; endometrial ablation, ER; endometrial resection, IUD; intrauterine device, LNG-IUS; Levonorgestrel intrauterine system, MEA; microwave endometrial ablation, MMAS; menorrhagia multi-attribute scale, PBAC; pictorial blood assessment chart, QoL; quality of life, SDM; shared decision making, TBEA; thermal balloon endometrial ablation, TCRE; Transcervical resection of the endometrium, TR; transcervical resection, UAE; uterine artery emboliszation, VAS; visual analogue scale.

APPENDIX 3. ECONOMIC EVALUATION ALONGSIDE ECLIPSE TRIAL: CHEERS GUIDELINES AND EXAMPLE OF DATA MANAGEMENT

A3.1 CHEERS Guidelines for Reporting Economic Evaluations

Section/item	Item No	Recommendation
Title and abstract		
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.
Introduction		
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.
Methods		
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.

Section/item	Item No	Recommendation
Measurement of effectiveness	11a	*Single study-based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.
	11b	*Synthesis-based estimates:* Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.
Estimating resources and costs	13a	*Single study-based economic evaluation:* Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.
	13b	*Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.
Results		
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input

Section/item	Item No	Recommendation
		values is strongly recommended.
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.
Characterising uncertainty	20a	*Single study-based economic evaluation:*Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).
	20b	*Model-based economic evaluation:* Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.
Discussion		
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.
Other		
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.

For consistency, the CHEERS statement checklist format is based on the format of the CONSORT statement checklist

A3.2 – Example of Data Management

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	patientid	treatment allocated	cycle change to mirena	cycle tx stopped	Tx stop after oral	cycle change to oral	well with oral	cycle new oral	cycle surgery	surgery after tx stop	surgery after change 2 oral	cycle mirena oral	cycle secnd new oral	cycle added oral	cycle mirena no oral	cycle secnd change to mirena
2	1001	LNG-IUS							2							
3	1002	LNG-IUS			16	14										
4	1004	LNG-IUS														
5	1006	LNG-IUS														
6	1009	LNG-IUS				5	6									
7	1010	LNG-IUS				2	3									
8	1014	LNG-IUS														
9	1015	LNG-IUS														
10	1018	LNG-IUS														

For example in column F it can be seen that patient 1002 (row 3) changed to usual medical treatment during cycle 14. Then column E shows that the same patient stopped treatment altogether at cycle 16.

APPENDIX 4. WILLINGNESS TO PAY METHODS: SCENARIOS, QUESTIONNAIRES AND INTERVIEW TOPIC GUIDE

A4.1 MMAS questionnaire weightings

1. Practical difficulties

1. I have no practical difficulties, bleed no more than I expect and take no extra precautions.	14.0
3. I have to carry extra sanitary protection and clothes because of the risk of flooding.	9.4
2. I have to carry extra sanitary protection with me but take no other precautions.	3.1
4. I have severe problems with flooding, soil the bedding and need to be close to a toilet.	0

2. Social life

1. My social life is unaffected during my cycle. I can enjoy life as much as usual.	10.0
2. My social life is slightly affected during my cycle. I may have to cancel or modify my plans.	6.7
3. My social life is limited during my cycle. I rarely make any plans.	2.7
4. My social life is devastated during my cycle. I am unable to make any plans.	0

3. Psychological health

1. During my cycle I have no worries I can cope normally.	14.0
2. During my cycle I experience some anxiety and worry.	8.4
3. During my cycle I often feel down and worry about how I'll cope	2.9
4. During my cycle I feel depressed and cannot cope.	0

4. Physical health and wellbeing

1. During my cycle I feel well and relaxed. I am not concerned about my health.	21.0
2. During my cycle I feel well most of the time. I am a little concerned about my health.	14.9
3. During my cycle I often feel tired and do not feel especially well. **1** am concerned about my health.	4.2
4. During my cycle I feel very tired and do not feel well at all. I am seriously concerned about my health.	0

***5.* Work/ daily routine**

1. There are no interruptions to my work/daily routine during my cycle.	18.0
2. There are occasional disruptions to my workldaily routine during my cycle.	11.3
3. There are frequent disruptions to my work/daily routine during my cycle	4.1
4. There are severe disruptions to my WorMdaily routine during my cycle.	0

***6.* Family life/relationships**

1. My family life/relationships are unaffected during my cycle.	23.0
2. My family life/relationships suffer some strain during my cycle.	14.0
3. My family life/relationships suffers quite a lot during my cycle.	5.3
4. My family life/relationships are severely disrupted as a result of my cycle.	0

A4.2 Ex-ante questionnaire

WOMENS' VALUES OF TREATMENT FOR HEAVY MENSTRUAL BLEEDING

Currently there are two possible ways of treating heavy menstrual bleeding (heavy periods) and we need to decide which is better. Although you may not personally experience heavy menstrual bleeding it is important to identify the views of women on the value of different treatment options.

We would be most grateful if you could complete the enclosed questionnaire. It should take no longer than 10 minutes to complete. Once completed please return the questionnaire using the stamped addressed envelope provided.

The questionnaire has two sections. The first section presents you with a scenario of heavy menstrual bleeding and will ask you about how much you would value different treatments for heavy menstrual bleeding. The second section will ask you some general questions about yourself.

There are no right or wrong answers. We are just interested in your views.

Please answer all the questions even if some may seem repetitive or less relevant, as it is important to get complete information.

The questionnaire is anonymous.

Your answers will remain confidential.

If you have any queries about completing this questionnaire do not hesitate to contact:

Miss Sabina Sanghera
Health Economics Unit
Public Health Building
University of Birmingham
Birmingham
B15 2TT
[redacted]

Thank you for your participation in this study.

Study ID | | | | |

Please check that you have answered <u>each question</u>

Section 1: Valuing treatments for heavy menstrual bleeding

TAKE YOUR TIME TO READ THE DESCRIPTIONS BELOW

The descriptions explain the experience of heavy menstrual bleeding (heavy periods) and the two treatment options.

Please imagine that you start having heavy periods from tomorrow, and what is described to you below is what you will experience up until the change (menopause).

- You have to carry extra sanitary protection and clothes because of the risk of flooding
- Your social life is slightly affected during your cycle. You may have to cancel or modify your plans
- During your cycle you experience some anxiety and worry
- During your cycle you often feel tired and do not feel especially well. You feel concerned about your health
- There are frequent disruptions to your work/daily routine during your cycle
- Your family life/relationships suffer some strain during your cycle

Now, rather than continue to live with heavy periods you could either be treated with Mirena coil (a coil that is designed to treat heavy periods) or Oral Treatment using tablets. Each treatment helps you to manage the bleeding in a different way:

Please turn over for description of treatments

Mirena

Mirena is a coil that is inserted into your womb by your GP or other qualified practitioner. The procedure usually takes a few minutes. Mirena can last for up to 5 years but it can be removed before if you wish. If you were given the Mirena, during the first 6 months you may experience irregular periods (bleeding in between periods) and your periods may not improve . Women are advised to persevere as the benefit of treatment can be seen after **6 months** and by 12 months most women will have stopped their periods. At 6 months on average you will:

- Have no practical difficulties, bleed no more than you expect and take no extra precautions
- Your social life is unaffected during your cycle and you can enjoy life as much as usual
- During your cycle you experience some anxiety and worry
- During your cycle you feel well most of the time and are a little concerned about your health
- There are no interruptions to your work/daily routine during your cycle
- Your family life/relationships suffer some strain during your cycle

Oral Treatment

Depending on the Oral Treatment you and your GP choose, you may either have to take tablets every day or just during your period. You may temporarily experience headaches or nausea or changes to your mood. If these persist you could change to a different Oral Treatment. When taking Oral Treatment you will see an **immediate** effect where on average you will:

- Have to carry extra sanitary protection but take no other precautions
- Your social life is slightly affected during your cycle and you may have to cancel or modify plans
- During your cycle you experience some anxiety and worry
- During your cycle you feel well most of the time and are a little concerned about your health
- There are occasional disruptions to your work/daily routine during your cycle
- Your family life/relationships suffer some strain during your cycle

1. **Out of these two treatment options Mirena or any Oral Treatment what is your preferred treatment?**

☐ Mirena ☐ Oral Treatment ☐ No preference

One way of measuring the value of different types of treatment for heavy menstrual bleeding is to ask you how much money you would be willing to pay for it. Of course, the treatments are provided free on the NHS and would stay free. This is simply a method of measuring the value you place on each treatment. So, imagine you do have to pay.

We believe that you should not have to pay for healthcare, other than prescription costs where necessary. The information you provide us will in no way be used to set or change prices for healthcare, it is simply a method of measuring how strongly you feel about the different treatment options and how much you value them.

There are no right or wrong answers. The amount you say could be large or small. Please keep in mind that you will need to provide values that are within your means i.e. please do not state values that you would not actually be able to pay. When thinking of a value assume that you would pay this amount every month until you reach menopause. Generally menopause occurs around 55 years of age.

If you expect menopause to occur earlier or later than 55 years, based on your family history please write the age below and use this age as your basis for duration of monthly payment:

My expected age of menopause is .... years old

PLEASE ANSWER BOTH QUESTIONS

2. **What would be the maximum monthly amount you would be willing to pay, out of your own pocket, for <u>Mirena</u>?**
Put a circle around the **maximum amount** you would pay

£0
£2
£4
£6
£8
£10
£12
£14
£16
£18
£20
£25
£30
£35
£40
£45
£50
£60
£70
£80
£90
£100
£150
£200
£250
£300
£350
£400
£450
£500

If more than £500 please state the exact amount: £_______________________

3. **What would be the maximum monthly amount you would be willing to pay, out of your own pocket, for <u>Oral Treatment</u>?**
Put a circle around the **maximum amount** you would pay

£0
£2
£4
£6
£8
£10
£12
£14
£16
£18
£20
£25
£30
£35
£40
£45
£50
£60
£70
£80
£90
£100
£150
£200
£250
£300
£350
£400
£450
£500

If more than £500 please state the exact amount: £_______________________

<u>The information you provide us will in no way be used to set or change prices for healthcare,</u>

4. **In the space provided below could you please tell us the reasons behind your answers to question 2 and 3?** *(What did you think about when choosing a maximum monthly amount?)*

Please check that you have answered <u>each question</u>

5. **Did you find questions 2 and 3 difficult to answer?**

☐ Yes ☐ No

6. **Could you please explain your answer to question 5**

Please turn over for remaining questions

Section 2: General information about yourself

Could you please tell us your age? ☐

7. What is your current status?

- ☐ Single
- ☐ Divorced
- ☐ Married/living with partner
- ☐ Separated
- ☐ Widowed

8. Would you like to have children in the future?

☐ Yes ☐ No

9. What is your current employment status:

- ☐ Employed (Full -time)
- ☐ Retired
- ☐ Long-term sick
- ☐ Employed (Part-time)
- ☐ Student
- ☐ Self-employed
- ☐ Unemployed
- ☐ Looking after family or home
- ☐ Other

*If other please state:*______________________________________

If employed, what is your current occupation? ______________________________________

10.Are you the main income earner in the household?

☐ Yes ☐ No

<u>If No</u>, what is the occupation of the main income earner? ______________________________

11.Could you please estimate the annual income of your <u>household</u> <u>before</u> deducting tax and national insurance (if you receive any benefits include them as income)?

- ☐ Less than £10,000
- ☐ £40,001 - £50,000
- ☐ £10,000 - £20,000
- ☐ £50,001 - £60,000
- ☐ £20,001 - £30,000
- ☐ £60,001 - £70,000
- ☐ £30,001 - £40,000
- ☐ More than £70,000

12.Have you experienced heavy menstrual bleeding?

☐ Yes No ☐

13. Could you please explain why you are visiting the Birmingham Women's hospital today?

....

14. Could you please indicate how satisfied or dissatisfied you are with your <u>life overall</u>: Please tick the box which you feel best describes how satisfied or dissatisfied you are.

- ☐ Very unsatisfied
- ☐ Slightly unsatisfied
- ☐ Neither satisfied nor unsatisfied
- ☐ Slightly satisfied
- ☐ Very satisfied

Please answer the questions by ticking one box in each group. Please indicate which statement best describes your own health <u>today</u>

Mobility	
I have no problems walking about	☐
I have some problems walking about	☐
I am confined to bed	☐
Self care	
I have no problems with self-care	☐
I have some problems washing or dressing myself	☐
I am unable to wash or dress myself	☐
Usual activities (e.g work, study, housework, family or leisure activities)	
I have no problems with performing my usual activities	☐
I have some problems with performing my usual activities	☐
I am unable to perform my usual activities	☐
Pain/Discomfort	
I have no pain or discomfort	☐
I have moderate pain or discomfort	☐
I have extreme pain or discomfort	☐
Anxiety/Depression	
I am not anxious or depressed	☐
I am moderately anxious or depressed	☐
I am extremely anxious or depressed	☐

15. Would you be willing to take part in an interview, that would last for about 30 minutes, to discuss your answers? *The interview will be conducted at your own convenience at the Birmingham Women's Hospital.*

☐ No ☐ Yes

Thank you for completing this questionnaire

A4.3 Ex-post Questionnaire

PATIENTS' VALUES OF TREATMENT FOR HEAVY MENSTRUAL BLEEDING

We would be most grateful if you could complete the enclosed questionnaire to help us better understand the affect that heavy menstrual bleeding has on your life and the value you would place on the different treatment options. It should take no longer than 15 minutes to complete. Once completed please return the questionnaire using the stamped addressed envelope provided within 2 weeks.

The questionnaire has three sections. The first section will ask you about your periods. The second section will ask you about how much you value treatments for heavy menstrual bleeding. The third section will ask you some general questions about yourself.

Please read through the instructions at the beginning of each section carefully. The questions are simple to complete. There are no right or wrong answers. We are just interested in your views.

Please answer all the questions even if some may seem repetitive or less relevant, as it is important to get complete information.

The questionnaire is anonymous.

Your answers will remain confidential.

If you have any queries about completing this questionnaire do not hesitate to contact:

Miss Sabina Sanghera
Health Economics Unit
Public Health Building
University of Birmingham
Birmingham
B15 2TT
[redacted]

Thank you for your participation in this study.

Section 1: Questions about your periods

In each of the following areas of health, select the statement that best applies to you and place a tick in the right hand side box provided. Please tick only one statement in each area.

---

1) Practical difficulties

a. I have no practical difficulties, bleed no more than I expect and take no extra precautions. ☐

b. I have to carry extra sanitary protection with me but take no other precautions. ☐

c. I have to carry extra sanitary protection and clothes because of the risk of flooding. ☐

d. I have severe problems with flooding, soil the bedding and need to be close to a toilet. ☐

---

2) Social life

a. My social life is unaffected during my cycle. I can enjoy life as much as usual. ☐

b. My social life is slightly affected during my cycle. I may have to cancel or modify my plans. ☐

c. My social life is limited during my cycle. I rarely make any plans. ☐

d. My social life is devastated during my cycle. I am unable to make any plans. ☐

---

3) Psychological health

a. During my cycle I have no worries I can cope normally. ☐

b. During my cycle I experience some anxiety and worry. ☐

c. During my cycle I often feel down and worry about how I'll cope. ☐

d. During my cycle I feel depressed and cannot cope. ☐

---

4) Physical health and wellbeing

a. During my cycle I feel well and relaxed. I am not concerned about my health. ☐

b. During my cycle I feel well most of the time. I am a little concerned about my health ☐

c. During my cycle I often feel tired and do not feel especially well. I am concerned about my health. ☐

d. During my cycle I feel very tired and do not feel well at all. I am seriously concerned about my health. ☐

5) Work/daily routine

a. There are no interruptions to my work/daily routine during my cycle.

b. There are occasional disruptions to my work/daily routine during my cycle.

c. There are frequent disruptions to my work/daily routine during my cycle.

d. There are severe disruptions to my work/daily routine during my cycle.

---

6) Family life/relationships

a. My family life/relationships are unaffected during my cycle.

b. My family life/relationships suffer some strain during my cycle.

c. My family life/relationships suffer quite a lot during my cycle.

d. My family life/relationships are severely disrupted as a result of my cycle.

Please turn over for remaining questions

Section 2: Valuing treatments for heavy menstrual bleeding

1. **What treatment(s) are you currently taking for your heavy periods?** *tick as many as applicable.*

- ☐ Mirena coil
- ☐ Ponstan (mefenamic acid)
- ☐ Contraceptive pill (any brand)
- ☐ Norethisterone
- ☐ No treatment
- ☐ Other (Please write the name) ……………………………………
- ☐ Copper coil
- ☐ Cyklokapron (tranexamic acid)
- ☐ Depo-provera
- ☐ Cerazette

2. **Have the treatment(s) that you take for your heavy periods changed since you last completed an ECLIPSE questionnaire (upon entering the trial 6mth/ 1 year/ 2 years/5 years)?**

☐ No ☐ Yes

3. **If YES, what treatment(s) were you taking before?** *Indicate as many as applicable*

- ☐ Mirena coil
- ☐ Ponstan (mefenamic acid)
- ☐ Contraceptive pill (any brand)
- ☐ Norethisterone
- ☐ No treatment
- ☐ Other (Please write the name) ……………………………………
- ☐ Copper coil
- ☐ Cyklokapron (tranexamic acid)
- ☐ Depo-provera
- ☐ Cerazette

Please turn over for remaining questions

One way of measuring the value of different types of treatment for heavy menstrual bleeding is to ask you what you would be prepared to give up to receive each treatment i.e. how much money would you be willing to pay for it.

We believe that you should not have to pay for healthcare, other than prescription costs where necessary. The information you provide us will in no way be used to set or change prices for healthcare, it is simply a method of measuring how strongly you feel about the different treatment options and how much you value them.

There are no right or wrong answers. The amount you say could be large or small. Please keep in mind that you will need to provide values that are within your means i.e. please do not state values that you would not actually be able to pay. When thinking of a value assume that you would pay this amount every month until you reach menopause. Generally menopause occurs around 55 years old.

If you expect menopause to occur earlier or later than 55 years old based on your family history please write the age below and use this age as your basis for duration of monthly payment:

My expected age of menopause is .... years

Please turn over for the question

4. What would be the maximum monthly amount you would be willing to pay, out of your own pocket, for your <u>current</u> treatment?

Put a circle around the **maximum amount** you would pay

£0
£2
£4
£6
£8
£10
£12
£14
£16
£18
£20
£25
£30
£35
£40
£45
£50
£60
£70
£80
£90
£100
£150
£200
£250
£300
£350
£400
£450
£500

If more than £500 please state the exact amount: £______________________

5. In the space below could you please tell us the reasons behind your answers to question 4?
(What did you think about when choosing a maximum monthly value?)

Please check that you have answered each question

6. Did you find question 4 difficult to answer?

☐ Yes ☐ No

7. Could you please explain your answer to question 6

Please turn over for remaining questions

Section 3: General information about yourself

8. Could you please tell us your age? ☐

9. What is your current status?

☐ Single ☐ Divorced

☐ Married/living with partner ☐ Separated

☐ Widowed

10. What is your current employment status:

☐ Employed (Full -time) ☐ Retired ☐ Long-term sick

☐ Employed (Part-time) ☐ Student ☐ Self-employed

☐ Unemployed ☐ Looking after family or home

☐ Other

*If other please state:*_____________________________________

If employed, what is your current occupation? _____________________________________

11. Are you the main income earner in the household?

☐ Yes ☐ No

If No, what is the occupation of the main income earner? _____________________________

12. Could you please estimate the annual income of your household before deducting tax and national insurance (if you receive any benefits include them as income)?

☐ Less than £10,000 ☐ £40,001 - £50,000

☐ £10,000 - £20,000 ☐ £50,001 - £60,000

☐ £20,001 - £30,000 ☐ £60,001 - £70,000

☐ £30,001 - £40,000 ☐ More than £70,000

13. How long have you had heavy menstrual bleeding?

☐ Less than 1 year ☐ 3 - 4 years

☐ 1 - 2 years ☐ 4 -5 years

☐ 2 - 3 years ☐ More than 5 years

We would be grateful if you could help our research into the best treatment for heavy periods by answering all of the following questions by ticking the appropriate boxes and providing any necessary additional information.

Thinking about the last 3 months:

14.Have you had to visit your GP due to heavy periods? ☐ No ☐ Yes

If YES, how many times did you visit your GP? ....

What for? ....

15.Has your GP had to visit you at home due to heavy periods? ☐ No ☐ Yes

If YES, how many times did your GP visit you? ....

What for? ....

16.Have you had to take time off work because of your heavy periods?

☐ No ☐ Yes ☐ Not currently working

If YES, how many days have you taken off work in the last three months? ....

17.Have your heavy periods prevented you from doing your other daily activities?

☐ No ☐ Yes

If YES, how many days in the last three months? ....

18.On average how many tampons/sanitary towels do you use during your period? ....

19.How regular is your cycle?

☐ Regular, I know when to expect my period
☐ Fairly regular, my period starts within a few days of when I expect
☐ Irregular, I cannot predict when my period will start
☐ I have bleeding on and off all the time

20.Do you take any drugs for pain relief during your periods?

☐ No ☐ Yes

If YES, what drugs? ☐ Ibuprofen (Neurofen) ☐ Paracetamol ☐ Aspirin

Other (please write the name) ....

On average how many do you take during your period? ....

21. **Please tick a number shown below to indicate how much pain you experience due to your periods**

0 1 2 3 4 5 6 7 8 9 10

No pain at all Worst imaginable pain

22. **Could you please indicate how much your quality of life is affected by your heavy menstrual bleeding. Please tick the box which you feel best describes how your quality of life is affected.**

- ☐ Not at all affected
- ☐ Slightly affected
- ☐ Affected
- ☐ Extremely affected

Could you please indicate how satisfied or dissatisfied you are with your life overall: Please tick the box which you feel best describes how satisfied or dissatisfied you are.

- ☐ Very unsatisfied
- ☐ Slightly unsatisfied
- ☐ Neither satisfied nor unsatisfied
- ☐ Slightly satisfied
- ☐ Very satisfied

Please answer the questions by ticking one box in each group. Please indicate which statement best describes your own health today

Mobility	
I have no problems walking about	☐
I have some problems walking about	☐
I am confined to bed	☐
Self care	
I have no problems with self-care	☐
I have some problems washing or dressing myself	☐
I am unable to wash or dress myself	☐
Usual activities (e.g work, study, housework, family or leisure activities)	
I have no problems with performing my usual activities	☐
I have some problems with performing my usual activities	☐
I am unable to perform my usual activities	☐
Pain/Discomfort	
I have no pain or discomfort	☐
I have moderate pain or discomfort	☐
I have extreme pain or discomfort	☐
Anxiety/Depression	
I am not anxious or depressed	☐
I am moderately anxious or depressed	☐
I am extremely anxious or depressed	☐

Please answer the questions by ticking one box in each group. Please indicate which statement best describes your own health <u>during your cycle</u>

Mobility	
I have no problems walking about	☐
I have some problems walking about	☐
I am confined to bed	☐
Self care	
I have no problems with self-care	☐
I have some problems washing or dressing myself	☐
I am unable to wash or dress myself	☐
Usual activities (e.g work, study, housework, family or leisure activities)	
I have no problems with performing my usual activities	☐
I have some problems with performing my usual activities	☐
I am unable to perform my usual activities	☐
Pain/Discomfort	
I have no pain or discomfort	☐
I have moderate pain or discomfort	☐
I have extreme pain or discomfort	☐
Anxiety/Depression	
I am not anxious or depressed	☐
I am moderately anxious or depressed	☐
I am extremely anxious or depressed	☐

Please turn over for remaining questions

23.Since you last completed an ECLIPSE questionnaire <u>(upon entering the trial 6mth/1 year/2 years/5 years)</u> have you been to hospital due to heavy periods?

☐ No ☐ Yes

If YES, what was this for?

Tests or investigations:

☐ Laparoscopy (camera via belly) ☐ Hysteroscopy (camera via vagina)

☐ Ultrasound scan

Other (please describe) ....

Treatment or surgery other than hysterectomy or endometrial ablation:

☐ Removal of polyps ☐ Removal of fibroids

☐ Treatment of endometriosis

Other (please describe) ....

Did you have to stay in hospital for the treatment? ☐ No ☐ Yes

If YES, how many nights? ....

Have you attended any follow-up clinics? ☐ No ☐ Yes

If YES, how many times? ....

Surgery:

☐ Hysterectomy

☐ Endometrial ablation (removal of lining of womb)

How many nights did you stay in hospital? ....

If you made two or more visits to hospital, please tell us about each visit – you can use the space at the end if necessary

....

24.Have you experienced any hot flushes/night sweats? ☐ No ☐ Yes

Thank you for completing this questionnaire

A4.4– Interview Topic Guide

Research Questions

1. Which outcome is most suitable
2. Was process and outcome utility taken into consideration
3. Were both health and non-health outcomes taken into consideration
4. Did they place a greater emphasis on non-health outcomes
5. Was WTP difficult to answer and why?
6. What did they take into consideration when completing the WTP questionnaire

Introduction

Firstly I'd just like to thank you for taking part in this study. My name is Sabina and I am a researcher at the University of Birmingham. As you know, we are carrying out some research to identify which measures should be used to value treatments for heavy periods. So we want to look at some of your answers to the questionnaire and talk a bit more about them.
The interview is very informal and completely confidential. Only my colleagues and I will see it and your name will not appear in anything we write. As you are already aware I will be recording the interview, this is so I can concentrate on what you are telling me rather than spending the whole time taking notes. As soon as we have written up the study all of the recordings will be destroyed.

To start off, I would like to find out a little bit more about you....

1. How is your health in general?
   - **1.** How do you feel during your monthly periods?
     - i. What makes you say that?
     - ii. What affect do they have on your everyday life/ normal activities? How so?
   - **2.** Have you ever been to your GP to seek advice about your periods?
     - i. Could you tell me a bit more about that?

If you remember, in the questionnaire we asked you to imagine you were experiencing heavy periods

Description of HMB

2. After reading the text in the questionnaire about heavy periods. How did you find placing yourself in the position of somebody with heavy periods?
    1. Could you imagine what it would feel like?
    2. What did you think of the description provided?

Description of treatments

3. So after the description of heavy periods we then described how the two treatments Mirena and Oral treatment would improve your periods. When reading the scenarios for the two treatments, was any part of the description more important to you than the other or was it all important? (process or outcome utility)
    1. Why was that?
    2. What did you like and dislike about each treatment?

WTP answers

So if we look at your answers to the questionnaire, particularly question 3 and 4 about the two treatment options and how much you would pay for them. You've said that you would be WTP £… for mirena and £…. For oral. And that your reason was…

4. Could you tell me a bit more about how you came up with value for each treatment?
    1. What type of things did you take into consideration?
        i. Did you think about how the periods would affect your life? [PROMPT: the cost of sanitary products, time off work or impact on family life]

    2. Was the affect of the heavy periods on your health (wellbeing and anxiety) equally important to you as affect on everyday activity (social life and family relationships)?

    3. You said previously that certain things related to the descriptions of each treatment were more important to you? How did that factor in to your decision-making here?
        i. How important was it to you when coming up with a value for the treatment?

4. Did you think about the impact on your monthly budget of having to pay the amounts -that you would have to sacrifice spending this money on one thing in order to pay for the treatment?
    i. What effect would paying this have on your monthly budget?

You're imagining you have the condition, what would be the point at which you would not be WTP any more?

Difficulty w/ WTP

5. You said in the questionnaire that you thought the WTP question was/was not difficult to answer.
    1. Can you tell me a bit more about why you thought that?
        i. What did you mean by…?

Preference reversal

5. You said that you prefer one treatment to the other but would be willing to pay the same amount for both treatments. Could you just explain your answer? (If preference reversal – why?)
    a. So if we set the cost of the treatments aside and think about purely how much you would value each treatment and how much of your monthly budget you would be willing to give up for this treatment, would you change the amount that you would be willing to pay for treatment, or keep them as they are?

Reconsider WTP value

6. Having thought about everything we have talked about so far would you like to change the amount that you would be willing to pay for the treatments or keep them as they are?

Appropriateness of questionnaires

7. How did you feel about the question on page 11 – EQ-5D (show questionnaire response)?
    a. Imagining you have heavy periods again, how would you feel answering these types of questions?
        i. Do you think it picks up on things that would be important to you? Why?

b. How do you think this measure compares to the WTP question? Difficult, easy? Why?

8. As you know we are trying to find out which measure is best for heavy periods. We use these measures to decide which treatment is better for women. Would you say that these measures (WTP or EQ-5D) are equally relevant or is one more than the other? And why?

9. Finally, in relation to heavy periods do you think there is anything missing from the questionnaire that we should have asked you?

Thank you very much for your time and for taking part. Your answers will be kept confidential and the recordings will be destroyed as soon as we have written up the study.

APPENDIX 5. WILLINGNESS TO PAY RESULTS: INTERVAL REGRESSION

A5.1: Interval Regression Results (Ex-ante)

Mirena

Variable	Coefficient	95% CI	P-value
Age of menopause	-0.0598	-0.096, -0.023	0.001
R5: Cost of treatment	0.459	-0.028, 0.945	0.064
R1: Found valuation difficult	1.923	-0.146, 3.995	0.068
R3: Affordability	0.601	0.144, 1.058	0.01
Experience of menorrhagia	-0.539	-1.053, -0.026	0.04
Preferred treatment			
Prefer Oral treatment	-0.749	-1.187, -0.311	0.001
No preference	-0.104	-0.814, 0.607	0.775
Income			
£20,001-£30,000	-0.174	-0.746, 0.397	0.55
£30,001-£40,000	0.416	-0.297, 1.129	0.253
£40,001-£50,000	1.156	0.440, 1.872	0.002
More than £50,000	0.187	-0.402, 0.777	0.533
Married	-0.670	-1.117, -0.223	0.003
EQ5Dbest	1.236	0.362, 2.110	0.006
Constant	4.675	2.577, 6.774	0.000

Log likelihood = -232.210, n= 93

Oral treatment

Variable	Coefficient	95% CI	P-value
R8: Cost of sanitary wear	0.713	0.079, 1.348	0.028
Difficult to answer	-0.846	-1.407, -0.285	0.003
R9: Misunderstanding exercise	1.040	-0.037, 2.117	0.059
D5: Difficulty with hypothetical	1.010	0.155, 1.865	0.021
D1: Not used to valuing	1.083	0.448, 1.717	0.001
Preferred treatment			
Prefer Oral treatment	0.260	-0.131, 0.652	0.193
No preference	-0.481	-1.130, 0.168	0.146
Income			
£20,001-£30,000	0.535	0.021, 1.048	0.041
£30,001-£40,000	0.850	0.215, 1.484	0.009
£40,001-£50,000	0.540	-0.088, 1.168	0.092
More than £50,000	0.194	-0.322, 0.711	0.461
D3: Benefit of treatment	0.417	0.016, 0.817	0.041
R3: Affordability	0.511	0.104, 0.918	0.014
Constant	1.238	0.742, 1.734	0.000

Log likelihood = -234.95, n=93

A5.2 Interval Regression (Ex-post)

Mirena

	Coefficient	95% CI	p-value
Current MMAS	0.044	0.023, 0.064	0
R7: effect of treatment	0.596	0.109, 1.082	0.016
R3: Nominal amount	1.781	0.015, 3.548	0.048
Time off work	4.309	1.949, 6.668	0
D5: Cost of sanitary products	-1.394	-2.673, -0.115	0.033
D10: Reasonable value	-1.189	-2.422, 0.044	0.059
Income			
£20,001-£30,000	-0.454	-1.163, 0.256	0.21
£30,001-£40,000	0.427	-0.319, 1.174	0.262
£40,001-£50,000	-0.403	-1.340, 0.534	0.399
More than £50,000	0.739	0.060, 1.418	0.033
R10: Misunderstanding exercise	-1.588	-2.506, -0.670	0.001
Constant	-1.666	-3.569, 0.238	0.086

Log likelihood= -136.982, n=49

Overall treatment

Variable	Coefficient	95% CI	p-value
Current MMAS	0.015	-0.001, 0.031	0.061
R10: Misunderstanding exercise	-1.241	-2.189, -0.293	0.01
Expected age of menopause	-0.110	-0.218, -0.002	0.046
D1: Not used to valuing	0.830	0.320, 1.339	0.001
R7: effect of treatment	0.512	0.054, 0.969	0.028
EQ5Damended	1.856	0.797, 2.916	0.001
EQ5D	-0.961	-2.039, 0.116	0.08
R2: Found valuation difficult	-0.740	-1.470, -0.009	0.047
R3: Nominal amount	2.019	0.096, 3.943	0.04
Prevented daily activities	0.977	0.039, 1.915	0.041
Constant	5.948	-0.036, 11.932	0.051

Log likelihood=-188.115, n=67

A5.3: Associations with EQ-5Da

	Current MMAS (rho)	Current EQ-5Da (rho)	Current EQ-5D (rho)
Current MMAS	1.0000		
Current EQ-5Da	0.5392*	1.0000	
Current EQ-5D	0.0931	0.1860	1.0000

*p<0.01

It can be seen that the amended EQ-5D has a significant correlation with MMAS scores, whilst EQ-5D has a weak non-significant correlation. The rho correlation coefficient (0.5392) for the amended EQ-5D and MMAS is considered to be a strong correlation.

REFERENCES

Abbott, J.; Hawe, J.; Hunter, D.; et al. (2003). A double-blind randomized trial comparing the Cavaterm (TM) and the NovaSure (TM) endometrial ablation systems for the treatment of dysfunctional uterine bleeding. **Fertility and Sterility**, 80 (1): 203-208

Altman, D.G. (1990). **Practical statistics for medical research**. London: Chapman and Hall.

Anand, P. & Wailoo, A. (2000) Utilities vs. rights to publicly provided goods: arguments and evidence from health-care rationing. **Economica**, 67: 543-577

Andronis, L., Barton, P., and Bryan S (2009).Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. **Health Technology Assessment,** 13, 29.

Argyrous, G. (2000). **Statistics for social and health research with a guide to SPSS**. London: SAGE publications

Arrow, K., Solow, R., Portney, P.R., et al. (1993). **Report of the NOAA panel of contingent valuation**. Federal Register; 58: 4601–4614

Bain, C.; Cooper, K.G.; and Parkin, D.E. (2002). Microwave endometrial ablation versus endometrial resection: A randomized controlled trial. **Obstetrics and Gynecology**, 99 (6): 983-987

Baker, R., Robinson, A., and Smith, R. (2008). How do respondents explain WTP responses? A review of the qualitative evidence, **Journal of Socio-Economics**, 37 (4); 1427-1442

Bannigan, K. & Watson, R. (2009). Reliability and validity in a nutshell. **Journal of Clinical Nursing**, 18: 3237-3243

Birch, S. & Donaldson, C. (2003). Valuing the benefits and the costs of healthcare programmes: where's the 'extra' in extra-welfarism. **Social Science and Medicine**. 56: 1121-1133

Blumenthal, P.D., Trussell, J., Singh, R.H., et al. (2006). Cost-effectiveness of treatments for dysfunctional uterine bleeding in women who need contraception. **Contraception**, 74 (3): 249-258

Blumenthal, P.D., Dawson, L., and Hurskainen, R. (2011). Cost-effectiveness and quality of life associated with heavy menstrual bleeding among women using the levonorgestrel-releasing intrauterine system. **International Journal of Gynaecology and Obstetrics.** 112 (3): 171-178.

Boadway, R.W. & Bruce, N. (1984) **Welfare economics**. 1st edition. Oxford: Oxford University Press

Bongers, M.Y., Bourdrez, P., Heintz, A.P. et al. (2005). Bipolar radio frequency endometrial ablation compared with balloon endometrial ablation in dysfunctional uterine bleeding: impact on patients' health-related quality of life. **Fertility and Sterility**, 83 (3): 724-734.

Brazier, J. (2005) **Current state of the art in preference based measures of health and avenues for future research**. HEDS discussion paper series. http://eprints.whiterose.ac.uk/id/eprint/10935

Brazier, J., Akehurst, H., Brennan, A. et al. (2005). Should patients have a greater role in valuing health states? **Applied Health Economics and Health Policy,** 4 (4): 201-208

Brazier, J., Devrill, M., Green, C. et al (1999). A review of the use of health status measures in economic evaluation. **Health Technology Assessment**, 3: 9

Brazier, J., Ratcliffe, J., Salomon, J.A. et al. (2010). **Measuring and Valuing Health Benefits for Economic Evaluation.** Oxford: Oxford University Press

Brazier, J., Roberts, J., and Deverill M. (2002). The estimation of a preference-based measure of health from the SF-36. **Journal of Health Economics**, 21: 271-292

Brazier, J., Roberts, J., Tsuchiya, A. et al. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. **Health Economics**, 13 (9): 873-884.

Brazier, J. & Rowen, D. (2011). NICE DSU technical support document 11. **Alternatives to EQ-5D for generating health state utilities**. Report by the Decision support unit. Sheffield: School of Health and Related Research, University of Sheffield, UK

Brazier, J., Usherwood, T.P., Harper, R. et al. (1998). Deriving a preference based single index measure for health from the SF-36. **Journal of Clinical Epidemiology** 51: 1115-1129

Brennan, A., Chick, S.E., Davies, R. (2006). A taxonomy of model structures for economic evaluation of health technologies**. Health Economics** 15 (12): 1295-1310

Bricker, L., Garcia, J., Henderson, J., et al. (2000). Ultrasound screening in pregnancy: a systematic review of the clinical effectiveness, cost effectiveness and womens' views. **Health Technology Assesment**, 4 (16): 1-193

Briggs, A.H. (2000). Handling uncertainty in cost-effectiveness models. **Pharmacoeconomics**, 17: 479–500.

Briggs, A.H., Ades, A.E., and Price, M.J. (2003). Probabilistic sensitivity analysis for decision trees with multiple branches: Use of the Dirichlet distribution in a Bayesian framework. **Medical Decision Making**, 23: 341-350

Briggs, A., Clark, T., Wolstenholme, J., et al. (2003). Missing… presumed at random: cost-analysis of incomplete data. **Health Economics**, 12: 377-392

Briggs, A., Sculpher, M., and Claxton, K. (2006). **Decision modelling for health economic evaluation.** Oxford: Oxford University Press

Brooks, R. (1996). EuroQol: the current state of play. **Health Policy,** 37 (1): 53–72

Brouwer, W.B.F., Culyer, A.J., van Exel, A.J.A., et al. (2008).Welfarism vs extra-welfarism. **Journal of Health Economics** 27: 325-338

Brouwer, W.B.F. & Koopmanschap, M.A. (2000). On the economic foundations of CEA. Ladies and gentleman take your positions! **Journal of Health Economics** 19: 439-459

Brouwer, W.B.F., Koopmanschap, M.A., and Rutten, F. (1997) Productivity costs in cost-effectivenessanalysis: nyumerator or denominator: a further discussion. **Health Economics**, 6: 5-11

Brown, P.M., Farquhar, C.M., Lethaby, A., et al. (2006). Cost-effectiveness analysis of levonorgestrel intrauterine system and thermal balloon ablation for heavy menstrual bleeding. **BJOG: An International Journal of Obstetrics and Gynaecology**, 113 (7): 797-803.

Bryman, A. (2012). **Social research methods**. Oxford: Oxford University Press

Bulmer, P. (2008). Menorrhagia. **Obstetrics, Gynaecology and Reproductive Medicine**, 18: 289-293

Bunkheila, A., & Powell, M. (2002). Menorrhagia and dysfunctional uterine bleeding. **Current Obstetrics and Gynaecology**, 12: 328-333.

Busfield, R.A.; Farquhar, C.M.; Sowter, M.C.; et al. (2006). A randomised trial comparing the levonorgestrel intrauterine system and thermal balloon ablation for heavy menstrual bleeding. **BJOG: An International Journal of Obstetrics and Gynaecology**, 113 (3): 257-263

Buxton, M.J., Drummond, M.F., van Hout, B.A., et al. (1997). Modelling in economic evaluation: an unavoidable fact of life. **Health Economics** 6: 212-227

Carpenter, J, R., & Kenward, M,G. (2007). Missing data in clinical trials. **Health Technology Assessment**. http://www.hta.nhs.uk/nihrmethodology/reports/1589.pdf

Centre for Reviews and Dissemination (2009). **Systematic Reviews: CRD's guidance for undertaking reviews in health care.** York: Centre for Reviews and Dissemination, University of York.

Chadha, Y.; Mollison, J.; Howie, F.; et al. (2000). Guidelines in gynaecology: evaluation in menorrhagia and in urinary incontinence**. BJOG: An International Journal of Obstetrics & Gynaecology**, 107 (4): 535-543

Clark, T.J., & Gupta, J.K. (2004). Outpatient thermal balloon ablation of the endometrium. **Fertility and Sterility**, 82 (5): 1395-1401.

Clark, T.J., Khan, K.S., Foon, R., et al. (2002). Quality of life instruments in studies of menorrhagia: a systematic review. **European Journal of Obstetrics, Gynecology, and Reproductive Biology**, 104 (2): 96-104.

Clark, T.J.; Samuel, N.; Malick, S.; et al. (2011). Bipolar Radiofrequency Compared With Thermal Balloon Endometrial Ablation in the Office A Randomized Controlled Trial. **Obstetrics and Gynecology**, 117 (1): 109-118

Clegg, J.P., Guest, J.F., and Hurskainen, R. (2007). Cost-utility of levonorgestrel intrauterine system compared with hysterectomy and second generation endometrial ablative techniques in managing patients with menorrhagia in the UK**. Current Medical Research Opinion**; 23: 1637-48.

Clinical Knowledge Summaries. (2012). **Menorrhagia**. [online] Available from: http://cks.nice.org.uk/menorrhagia#!prescribinginfo Last accessed 14.09.13

Coast, J. (2004). Is economic evaluation in touch with societys health values. **British Medical Journal,** 329: 1233-1236

Coast, J. (2009). Maximisation in extra-welfarism: a critique of the current position in health economics. **Social Science and Medicine**, 69: 786-792

Coast, J., Smith, R.D., and Lorgelly, P. (2008a). Should the capability approach be applied in health economics? **Health Economics**, 17: 667–670

Coast, J., Smith, R.D., and Lorgelly, P. (2008b). Welfarism, extra-welfarism and capability: The spread of ideas in health economics. **Social Science and Medicine**. 67; 1190-1198

Cohen, G.A. (1993). Equality of what? On welfare, goods and capabilities. In: Nussbaum, M.C., Sen, A.K. **The quality of life**. Oxford: Clarendon Press

Cooper, K.G., Bain,C., Lawrie,L., et al. (2005). A randomised comparison of microwave endometrial ablation with transcervical resection of the endometrium; follow up at a minimum of five years. **BJOG: An International Journal of Obstetrics and Gynaecology**, 112 (4): 470-475

Cooper, K.G.; Bain, C.; and Parkin, D.E. (1999a). Comparison of microwave endometrial ablation and transcervical resection of the endometrium for treatment of heavy menstrual loss a randomised trial. **Lancet**, 354 (9193): 1859-1863

Cooper, K.G., Grant, A.M., and Garratt,A.M. (1997a). The impact of using a partially randomised patient preference design when evaluating alternative managements for heavy menstrual bleeding. **British Journal of Obstetrics and Gynaecology**, 104 (12): 1367-1373

Cooper, K.G.; Jack, S.A.; Parkin, D.E.; et al. (2001). Five-year follow up of women randomised to medical management or transcervical resection of the endometrium for heavy menstrual loss: clinical and quality of life outcomes. **BJOG: An International Journal of Obstetrics and Gynaecology**, 108 (12): 1222-1228

Cooper, K.G.; Parkin, D.E.; Garratt, A.M.; et al. (1997b). A randomised comparison of medical and hysteroscopic management in women consulting a gynaecologist for treatment of heavy menstrual loss**. British Journal of Obstetrics and Gynaecology**, 104 (12): 1360-1366

Cooper, K.G.; Parkin, D.E.; Garratt, A.M.; et al. (1999b). Two-year follow up of women randomised to medical management or transcervical resection of the endometrium for heavy menstrual loss: clinical and quality of life outcomes. **British Journal of Obstetrics and Gynaecology**, 106 (3): 258-265

Coulter, A., Bradlow, J., Agass, M., et al. (1991). Outcomes of referrals to gynaecology outpatient clinics for menstrual problems: an audit of general practice records. **British Journal of Obstetrics and Gynaecology**, 98: 789-796

Coulter, A., Long, A., Kelland, J., et al. (1995). Managing Menorrhagia. **Quality in Health care**, 4: 218-226

Coulter, A., Peto,V., and Jenkinson,C. (1994). Quality of life and patient satisfaction following treatment for menorrhagia. **Family Practice**, 11 (4): 394-401

Cribb, J., Joyce, R., and Philips, D. (2012). **Living standards, poverty and inequality in the UK: 2012**. London: Institute for Fiscal studies

Crosignani, P.G.; Vercellini, P.; Apolone, G.; et al. (1997a). Endometrial resection versus vaginal hysterectomy for menorrhagia: long-term clinical and quality-of-life outcomes. **American Journal of Obstetrics and Gynecology**, 177 (1): 95-101

Crosignani, P.G.; Vercellini, P.; Mosconi, P.; et al. (1997b). Levonorgestrel-releasing intrauterine device versus hysteroscopic endometrial resection in the treatment of dysfunctional uterine bleeding. **Obstetrics and Gynecology**, 90 (2): 257-263

Culyer, A. (1991). The normative economics of health care finance and provision. In A. McGuire, P. Fenn, and K.Mayhew (Eds.), Providing health care: **The economics of alternative systems of finance and delivery**. Oxford: Oxford University Press

Currie, G.R., Donaldson, C., and McIntosh, E. (1999). **Cost-effectiveness analysis and 'the societal approach' should the twain ever meet**? Institute of Health Economics Working paper 00-3

Curtis L. (2011). **Unit costs of health and social care: 2011**. Canterbury: Personal Social Service Research Unit

Dalmau-Matarrodona, E. (2001). Alternative approaches to obtain optimal bid values in contingent valuation studies and to model protest zeros. Estimating the determinants of individuals' willingness to pay for home care services in day case surgeries. **Health Economics**, 10; 101-118

Daniels, J.P., Middleton, L.J., Champaneria, R., et al. (2012). Second generation endometrial ablation techniques for heavy menstrual bleeding: a network meta-analysis. **British Medical Journal**, 23 (344): e2564

Davis, A.; Godwin, A.; Lippman, J.; et al. (2000). Triphasic norgestimate-ethinyl estradiol for treating dysfunctional uterine bleeding. **Obstetrics and Gynecology**, 96 (6): 913-920

Davis, J.C., Liu-Ambrose, T., Khan, K.M., et al. (2012) SF-6D and EQ-5D result in widely divergent incremental cost-effectiveness ratios in a clinical trial of older women: implications for health policy decisions. **Osteoporosis International**, 23 (7): 1849-1857

De Souza, S.S., Camargos, A.F., de Rezende,C.P., et al. (2010). A randomized prospective trial comparing the levonorgestrel-releasing intrauterine system with thermal balloon ablation for the treatment of heavy menstrual bleeding. **Contraception**, 81 (3): 226-231.

Dickersin, K., Munro, M.G., Clark, M., et al. (2007). Hysterectomy compared with endometrial ablation for dysfunctional uterine bleeding - A randomized controlled trial. **Obstetrics and Gynecology**, 110 (6): 1279-1289.

Diener, A., O'Brien, B., and Gafni, A. (1998). 'Health care contingent valuation studies: a review and classification of the literature'. **Health Economics**, 7: 313–326.

Dolan, P. (1997). Aggregating health state valuations. **Journal of Health Services and Research Policy**, 2 (3):160-165.

Dolan, P. (2000). The Measurement of health related quality of life. In Culyer, A.J, and Newhouse, J.P (Eds.). **Handbook of health economics.** North-Holland: Amsterdam

Dolan, P., Gudex, C., Kind P., et al. (1995). **A social tariff for Euroqol: results from a UK general population survey**. Working paper, University of York

Dolan, P., Shaw, R., Tsuchiya, A., and Williams, A. (2005). QALY maximisation and people's preferences: a methodological review of the literature. **Health Economics**, 14, 197–208

Donaldson, C. (1999). Valuing publicly provided health care: Does ability to pay preclude the use of willingness to pay? **Social Science and Medicine**, 49: 551–564.

Donaldson, C., Farrar, S., Mapp, T., et al. (1997a) Assessing community values in health care: is the willingness to pay method feasible? **Health Care Analysis**, 5 (1): 7-29

Donaldson, C., Jones, A.M., Mapp, T.J., et al. (1998). Limited dependent variables in willingness to pay studies: applications in health care. **Applied Economics**, 30: 667-677

Donaldson, C., Shackley, P., Adbdalla, M. (1997b). Using willingness to pay to value close substitutes: carrier screening for cystic fibrosis revisited. **Health Economics**, 6: 145-159

Draper, N., & Smith, H. (1998). **Applied regression analysis**. (3rd ed.) New York: Wiley

Drummond, M.F., Sculpher, M.J., Torrance, G.W., et al. (2005). **Methods for the economic evaluation of health care programmes**. 3rd edition. Oxford: Oxford University Press

Duckitt, K., & Shaw, R.W. (1998). Is medical management of menorrhagia obsolete? **British Journal of Obstetrics and Gynaecology**, 105 (6): 569-572

Edwards, A.; Elwyn, G.; Hood, K.; et al. (2004). Patient-based outcome results from a cluster randomized trial of shared decision making skill development and use of risk communication aids in general practice. **Family Practice**, 21 (4): 347-354

Effective health care (1995). The management of menorrhagia. **Effective Health Care Bulletin**; 9: 1–14

Endrikat, J., Vilos, G., Muysers, C., et al. (2012). The levonorgestrel-releasing intrauterine system provides a reliable, long-term treatment option for women with idiopathic menorrhagia. **Archives of Gynecology and Obstetrics**; 285 (1):117–121.

Fraser, I.S., McCarron, G., Markham, R., (1984). A preliminary study of factors influencing the perception of menstrual blood loss volume. **American Journal of Obstetrics and Gynaecology**, 149: 788-793.

Frew, E., Wolstenholme, J., and Whynes, D. (2001). Willingness to pay for colorectal cancer screening. **European Journal of Cancer**, 37: 1746.

Frew, E.J., Whynes, D.K., Wolstenholme, J.L. (2003). Eliciting willingness to pay: comparing closed-ended with open-ended and payment scale formats. **Medical Decision Making,** 23: 150-159

Frick, K.D., Clark, M.A., Steinwachs, D.M., et al. (2009). Financial and quality-of-life burden of dysfunctional uterine bleeding among women agreeing to obtain surgical treatment. **Womens Health Issues**, 19 (1): 70-78.

Gafni, A. (2006). Economic evaluation of health-care programmes: Is CEA better than CBA? **Environmental and Resource Economics,** 34: 407-418

Ganz, M.I., Shah, D., Gidwani, R., et al. (2013). The cost-effectiveness of the levonorgestrel-releasing intrauterine system for the treatment of idiopathic heavy menstrual bleeding in the United States. **Value in Health,** 16: 325-333

Garratt, A.M., Ruta, D.A., Abdalla, M.I., et al. (1993). The Sf-36 health survey questionnaire - an outcome measure suitable for routine use within the NHS. **British Medical Journal**, 306 (6890): 1440-1444.

Garratt, A.M., Ruta, D.A., Abdalla, M.I., et al. (1994). SF 36 health survey questionnaire: II. Responsiveness to changes in health status in four common clinical conditions. **Quality in Health Care**, 3 (4): 186-192.

Gemzell-Danielson, K., Inci, P., and Heikinheimo, O. (2013). Safety and efficacy of the levonorgestrel relasing intrauterine system: recent insights. **Expert Review of Obstetrics and Gynecology**, 8 (3): 235-247

Glick, H.A., Doshi, J.A., Sonnad, S.S., et al. (2007). **Economic Evaluation in Clinical Trials**. Oxford: Oxford University Press

Gokyildiz, S., Aslan, E., Beji, N.K., et al. (2013). The effects of menorrhagia on women's quality of life: A case-control study. **ISRN Obstetrics and Gynaecology**, doi.org/10.1155/2013/918179

Gold, M., Siegel, J., Russell, L., et al. (1996). **Cost effectiveness in health and medicine**. New York: Oxford University Press

Gorgen, H.; Api, M.; Akca, A.; et al. (2009). Use of the Levonorgestrel-IUS in the treatment of menorrhagia: assessment of quality of life in Turkish users. **Archives of Gynecology and Obstetrics**, 279 (6): 835-840

Gralapp, C. (1996). **Insert omission**. Agency for health research and quality [online] http://webmm.ahrq.gov/case.aspx?caseID=124 [Last accessed 05.09.13]

Grant, C., Gallier, L., Fahey, T., et al. (2000). Management of menorrhagia in primary care impact on referral and hysterectomy: data from the Somerset Morbidity Project. **Journal of Epidemiology and Community Health**, 54 (9): 709-713

Gray, A.M., Clarke, P.M., Wolstenholme, J.L., et al. (2010). **Applied methods of cost-effectiveness analysis in health care**. Oxford: Oxford University Press

Gupta, J., Kai, J., Middleton, L., et al. (2013). Levonorgestrel intrauterine system vs medical therapy for menorrhagia. **New England Journal of Medicine**, 368: 128-137

Habiba, M., Julian, S., Taub, N., et al. (2010). Limited role of multi-attribute utility scale and SF-36 in predicting management outcome of heavy menstrual bleeding. **European Journal of Obstetrics and Gynecology and Reproductive Biology**, 148 (1): 81-85

Haefeli, M., Elfering, A., Mcintosh, E., et al. (2008). A Cost-Benefit Analysis Using Contingent Valuation Techniques: A Feasibility Study in Spinal Surgery. **Value in Health**, 11 (4): 575-588

Hallberg, L., Hogdahl, A.M., Nilsson, L., et al. (1966). Menstrual blood loss - a population study. **Acta Obstetrica et Gynaecologica Scandinavica,** 45: 320-351

Halstead, J.M., Luloff, A.E., and Stevens, T.H. (1992), 'Protest Bidders in Contingent Valuation', **Northeastern Journal of Agricultural Economics** 21, 160–169.

Hanley, N., Mourato, S., and Wright, R.E. (2001). Choice modelling approaches: a superior alternative for environmental evaluation? **Journal of Economic Surveys,** 15 (3): 435-462

Harless, D.W, & Camerer C.F. (1994). The predictive utility of generalized expected utility theories. **Econometrica**, 62 (6): 1251-1289

Hawe, J.; Abbott, J.; Hunter, D.; et al. (2003). A randomised controlled trial comparing the Cavaterm endometrial ablation system with the Nd:YAG laser for the treatment of dysfunctional uterine bleeding. **BJOG: An International Journal of Obstetrics and Gynaecology**, 110 (4): 350-357

Hehenkamp, W.J.; Volkers, N.A.; Birnie, E.; et al. (2008) Symptomatic uterine fibroids: treatment with uterine artery embolization or hysterectomy--results from the randomized clinical Embolisation versus Hysterectomy (EMMY) Trial. **Radiology,** 246 (3): 823-832

Helal, A.S.; Abdel-Hady, E.; Mashaly, A.E.; et al. (2011). Modified thermal balloon endometrial ablation in low resource settings: a cost-effective method using Foley's catheter. **Archives of Gynecology and Obstetrics**, 284 (3): 671-675

Henshaw, R.; Coyle, C.; Low, S.; et al. (2002). A retrospective cohort study comparing microwave endometrial ablation with levonorgestrel-releasing intrauterine device in the management of heavy menstrual bleeding. **Australian & New Zealand Journal of Obstetrics and Gynaecology**, 42 (2): 205-209

Herdman, M., Gudex, C., Lloyd, A., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). **Quality of Life Research**, 20 (10); 1727-1736

Hertwig, R., Hoffrage, U., & Martignon, L. (1999). Quick estimation: Letting the environment do the work. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), **Simple heuristics that make us smart**. New York: Oxford University Press.

Hicks, J.R. (1939). The foundation of welfare economics. **Economic Journal**, 49; 696-712

Higgins, J.P.T., & Green, S. (2009). **Cochrane Handbook for Systematic Reviews of Interventions**. The Cochrane Collaboration and John Wiley & Sons Ltd

Hurley, J., (1998). Welfarism, extra-welfarism and evaluative economic analysis in the health care sector. In: Barer, M.L., Getzen, T.E., Stoddard,G.L. (Eds.), **Health, Health Care and Health Economics: Perspectives on Distribution**. John Wiley & Sons Ltd., Chichester

Hurskainen, R., Teperi, J., Rissanen, P., et al. (2001). Quality of life and cost-effectiveness of levonorgestrel-releasing intrauterine system versus hysterectomy for treatment of menorrhagia: a randomised trial. **Lancet**, 357 (9252): 273-277

Hurskainen, R., Teperi, J., Rissanen, P., et al. (2004). Clinical outcomes and costs with the levonorgestrel-releasing intrauterine system or hysterectomy for treatment of menorrhagia - Randomized trial 5-year follow-up. **JAMA: Journal of the American Medical Association**, 291 (12): 1456-1463

Husereau, D., Drummond, M., Petrou,S., et al. (2013). Consolidated health economic evaluation reporting standards (CHEERS) statement. **British Medical Journal**, 346: f1049

James, A., Matchar, D.B., and Myers, E.R. (2004). Testing for von Willebrand disease in women with menorrhagia: a systematic review. **Obstetrics and Gynecology**, 104 (2): 381-388

Janssen, C.A.H., Scholten, P.C., and Heintz, P.M., (1998). Reconsidering menorrhagia in gynaecological practice is a 30-year old definition still valid? **European Journal of Obstetrics and Gynaecology and Reproductive Biology,** 78: 69-72.

Jenkinson, C., Peto, V., and Coulter, A. (1996). Making sense of ambiguity: Evaluation of internal reliability and face validity of the SF 36 questionnaire in women presenting with menorrhagia. **Quality in Health Care**, 5 (1): 9-12.

Johannesson, M., & Jonsson, B. (1991) Economic evaluation in healthcare: is there a role for cost-benefit analysis. **Health Policy,** 14 (1): 1-23

Joint formulary committee. (2011). **British National Formulary 62**. BMJ Group & Pharmaceutical Press

Jorgensen, B.S., Syme, G.J., Bishop, B.J., et al. (1999). Protest responses in contingent valuation. **Environmental and Resource Economics,** 14: 131-150

Kadir, R.A. (2009). Menorrhagia: treatment options. **Thrombosis Research**, 123 (Suppl 2): S21-S29

Kahnemann, D., & Knetsch, J.L. (1992). Valuing public goods: the purchase of moral satisfaction. **Journal of Environmental Economics and Management,** 22: 57-70

Kaldor, N. (1939) Welfare propositions of economics and interpersonal comparisons of utility. **Economic Journal,** 49 (195): 549-552

Kennedy, A.D.M., Sculpher, M.J., Coulter, A., et al (2003). A multicentre randomised controlled trial assessing the costs and benefits of using structured information and analysis of women's preferences in the management of menorrhagia. **Health Technology Assessment,**.7 (8)

Kilonzo, M.M., Sambrook, A.M., Cook, J.A., et al. (2010). A cost-utility analysis of microwave endometrial ablation versus thermal balloon endometrial ablation. **Value in Health**, 13 (5): 528-534

Kleijn, J.H.; Engels, R.; Bourdrez, P.; et al. (2008). Five-year follow up of a randomised controlled trial comparing NovaSure and ThermaChoice endometrial ablation. **BJOG-An International Journal of Obstetrics and Gynaecology**, 115 (2): 193-198

Klose, T. (1999), 'The contingent valuation method in health care', **Health Policy**, 47: 97–123

Knies, S., Evers, S.M.A.A., Candel, M.J.J.M., et al. (2009). Utilities of the EQ-5D. **PharmacoEconomics**, 27 (9); 767-779

Kujovich, J.L. (2005). von Willebrand's disease and menorrhagia: prevalence, diagnosis, and management. **American Journal of Hematology**, 79 (3): 220-228

Kuppermann, M.; Varner, R.E.; Summitt, R.L.; et al. (2004). Effect of hysterectomy vs medical treatment on health-related quality of life and sexual functioning - The medicine or surgery (Ms) randomized trial. **JAMA-Journal of the American Medical Association**, 291 (12): 1447-1455

Lete, I., Cristobal, I., Febrer, L., et al. (2011). Economic evaluation of the levonorgestrel-releasing intrauterine system for the treatment of dysfunctional uterine bleeding in Spain. **European Journal of Obstetrics and Gynecology and Reproductive Biology,** 154: 71–80

Little, I.M.D. (1949). The foundation of welfare economics. **Oxford Economic Papers**. 1 (2); 227-246

Longworth, L., & Rowen, D. (2013). Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. **Value in Health,** 16 (1): 202-210

Malak, A.K., & Shawki, O. (2006). Management of menorrhagia with the levonorgestrel intrauterine system versus endometrial resection. **Gynecological Surgery**. 3 (4): 275-280

Matsumoto, Y.; Yamabe, S.; Ideta, K.; et al. (2007). Impact of use of combined oral contraceptive pill on the quality of life of Japanese women. **Journal of Obstetrics and Gynaecology Research**, 33 (4): 529-535

McIntosh, E., Clarke, P.M., Frew, E.J., et al. (2010). **Applied methods of cost-benefit anlaysis in health care**. Oxford: Oxford University Press

Mishan, E.J. (1972) **Elements of cost-benefit analysis**. London: George Allen and Unwin Ltd

Mitchell, R,C., & Carson, R.T. (1989). **Using surveys to value public goods**. 3rd edition. Washington DC: Resources for the future

Morris, S., Devlin, N., and Parkin, D. (2007) **Economic analysis in health care**. Oxford: Oxford University Press

Mugford, M. (1996). **How does the method of cost estimation affect the assessment of cost effectiveness in health care?** Oxford: Oxford University Press

Musgrave, R.A., 1959. **The economics of public finance**. New York: McGraw Hill

National Collaborating Centre for Women's and Children's Health (2007). **Heavy Menstrual Bleeding**. London: Royal College of Obstetricians and Gynaecologists.

National Health Service Choices. (2012). **Periods**. [online] http://www.nhs.uk/conditions/periods/Pages/Introduction.aspx [Last accessed 05.09.13]

National Institute of Health and Clinical Excellence (2004). **Guide to the methods of technology appraisal**. London: National Institute of Health and Clinical Excellence

National Institute of Health and Clinical Excellence (2007). **Heavy menstrual bleeding: clinical guideline**. London: National Institute of Health and Clinical Excellence.

National Institute of Health and Clinical Excellence (2008). **Guide to the methods of technology appraisal**. London: National Institute of Health and Clinical Excellence

National Institute of Health and Clinical Excellence (2013). **Guide to the methods of technology appraisal 2013**. London: National Institute of Health and Clinical Excellence

O'Brien, B., & Gafni, A. (1996). When do the dollars make sense?: Toward a conceptual framework for contingent valuation studies in healthcare. **Medical Decision Making**, 16: 288-299

Office of Population Census and Surveys (1995). **Morbidity Statistics from General Practice**. London: HMSO

O'Flynn, N., & Britten, N. (2000). Menorrhagia in general practice disease or illness. **Social Science and Medicine**, 50 (5): 651-661.

Olah, K.S.; Alliston, J.; Jones, J.; et al. (2005). Thermal ablation performed in a primary care setting: the South Warwickshire Experience. **BJOG-An International Journal of Obstetrics and Gynaecology**, 112 (8): 1117-1120

Olsen, J.A., & Smith, R.D. (2001). Theory versus practice: a review of willingness to pay in health and health care. **Health Economics**, 10: 39-52

Parkin, D., & Devlin, N. (2006) Is there a case for using visual analogue scale valuations in CUA? **Health Economics**, 15 (7): 653-664

Pattison, H., Daniels, J.P., Kai, J., et al. (2011). The measurement properties of the menorrhagia multi-attribute quality-of-life scale: a psychometric analysis. **BJOG-An International Journal of Obstetrics and Gynaecology,** 118 (12): 1528-1531

Philips, Z., Ginnelly, L., Sculpher, M., et al. (2004). Review of Guidelines for Good Practice in Decision-Analytic Modelling in Health Technology Assessment. **Health Technology Assessment**, 8: 36

Pirkis, J.E., Burgess, P.M., Kirk, P.K., et al. (2005). A review of the psychometric properties of the Health of the Nation Outcome Scales (HoNOS) family of measures. **Health and Quality of life outcomes,** 3: 76

Protheroe, J. (2004). Modern Management of Menorrhagia. **Journal of Family Planning and Reproductive Health Care**, 30 (2): 118-122

Protheroe, J.; Bower, P.; Chew-Graham, C.; et al. (2007). Effectiveness of a computerized decision aid in primary care on decision making and quality of life in menorrhagia: results of the MENTIP randomized controlled trial. **Medical Decision Making**, 27 (5): 575-584

Ramazanzadeh, F., Tavakolianfar, T., Shariat, M., et al. (2012). Levonorgestrel-releasing IUD versus copper IUD in control of dysmenorrhea, satisfaction and quality of life in women using IUD. **Iranian Journal of Reproductive Medicine**, 10 (1): 41-46

Reaves, D.X., Kramer, R.A., and Holmes, T.P. (1999). Does question format matter? Valuing an endangered species. **Environmental and Resource Economics**, 14: 365-383

Rees, M.C. (1991) Role of menstrual blood loss measurements in management of complaints of excessive menstrual bleeding. **British Journal of Obstetrics and Gynaecology**, 98 (3): 327-328

Ritchie, J., & Lewis, J. (2003) **Qualitative Research Practice: a guide for social science students and researchers**. London: SAGE Publications

Roberts, T.E., Henderson, J., Mugford, M., et al. (2002) Antenatal ultrasound screening for fetal abnormalities: a systematic review of cost and cost effectiveness studies. **British Journal of Obstetrics and Gynaecology**, 109: 44-56

Roberts, T.E., Tsourapas, A., Middleton, L.J., et al (2011). Hysterectomy, endometrial ablation, and levonorgestrel releasing intrauterine system (Mirena) for treatment of heavy menstrual bleeding: cost effectiveness analysis. **British Medical Journal**; 342:d2202

Royal College of Obstetricians and Gynaecologists (1998). **The Initial Management of Menorrhagia**. London: Royal College of Obstetricians and Gynaecologists

Ruta, D.A., Abdalla, M.I., Garratt, A.M., et al. (1994). SF 36 health survey questionnaire: I. Reliability in two patient based studies. **Quality in Health Care**, 3 (4): 180-185

Ryan, M. (1999) Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. **Social Science and Medicine**, 45 (4): 535-546

Ryan, M., & Gerard, K. (2003). Using discrete choice experiments to value healthcare programmes: current practice and future research recommendations. **Applied Health Economics and Health Policy,** 2 (1): 55-64

Ryan, M., & San Miguel, F. (2000). Testing for consistency in willingness to pay experiments. **Journal of Economic Psychology**, 21: 305-317.

Sach, T.H., Barton, G.R., Jenkinson, C., et al. (2009). Comparing cost-utility estimates: Does the choice of EQ-5D or SF-6D matter? **Medical Care,** 47 (8): 889-894

Sach, T., Smith, R.D., and Whynes, D. (2007), ‘A ‘‘league table’’ of contingent valuation results for pharmaceutical interventions: a hard pill to swallow?’**PharmacoEconomics**, 25 (2): 107-127

Sambrook, A.M., Bain, C., Parkin, D.E., et al. (2009). A randomised comparison of microwave endometrial ablation with transcervical resection of the endometrium: follow up at a minimum of 10 years. **BJOG: An International Journal of Obstetrics and Gynaecology**, 116, (8): 1033-1037

Sambrook, A. & Cooper, K. (2005). RCOG Guidelines on Menorrhagia - Time for an Update. **Current Obstetrics and Gynaecology**, 15: 382-386

Sambrook, A.M.; Cooper, K.G.; Campbell, M.K.; et al. (2009a). Clinical outcomes from a randomised comparison of Microwave Endometrial Ablation with Thermal Balloon endometrial ablation for the treatment of heavy menstrual bleeding**. BJOG-An International Journal of Obstetrics and Gynaecology**, 116 (8): 1038-1045

Samuel, N.C.; Karragianniadou, E.; and Clark, T.J. (2009b). Outpatient versus day-case endometrial ablation using the NovaSure impedance-controlled ablative system. **Gynecological Surgery**, 6 (1): 3-9

Sanghera, S., Frew, E., Kai, J., et al. (2013). An Assessment of Economic Measures used in Menorrhagia: A Systematic Review. **Social Science and Medicine,** 98: 149-153

San Miguel, F., Ryan, M., McIntosh, E (2000). Applying conjoint analysis in economic evaluations: an application to menorrhagia. **Applied Economics,** 32: 823-833

Sculpher, M. (1998). A cost-utility analysis of abdominal hysterectomy versus transcervical endometrial resection for the surgical treatment of menorrhagia. **International Journal of Technology Assessment in Health Care**; 14: 302-319

Sculpher, M. (2001). The role and estimation of productivity costs in economic evaluation. In: (Eds.), M.F Drummond and A.McGuire). **Economic evaluation in healthcare: merging theory with practice.** Oxford: Oxford University Press

Sculpher, M.J.; 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 Obstetrics and Gynaecology**, 100 (3): 244-252

Sculpher, M.J.; Dwyer, N.; Byford, S.; et al. (1996). 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**, 103 (2): 142-149

Sen, A. (1977) Social choice theory: a re-examination. **Econometrica**, 45: 53–89.

Sen, A.K., (1980). Equality of what? In: McMurrin, S. **The Tanner Lectures on Human Values.** Cambridge: Cambridge University Press.

Shackley, P., & Donaldson, C. (2000). Willingness to pay for publicly financed healthcare: how should we use the numbers? **Applied Economics,** 32: 2015-2021

Shackley, P., & Donaldson, C. (2002). Should we use willingness to pay to elicit community preferences for healthcare? New evidence from using a ‘marginal’ approach. **Journal of Health Economics**, 21: 971-991

Shapley, M., Jordan, K., and Croft, P.R. (2002). Why women consult with increased vaginal bleeding: a case-control study. **British Journal of General Practice**; 52 (475):108–113

Shaw, R.W., Brickley, M.R., Evans, L., et al. (1998). Perceptions of women on the impact of menorrhagia on their health using multi-attribute utility assessment. **British Journal of Obstetrics and Gynaecology**, 105 (11); 1155-1159

Shawki, O.; El-Sherbiny, W.; Saber, W.; et al. (2009). Treatment of heavy menstrual bleeding associated with uterine leiomyoma with the levonorgestrel-releasing intrauterine system. **Gynecological Surgery**, 6 (4): 331-337

Sibbald, B., & Roland, M. (1998). Understanding Controlled Trials: why are randomised controlled trials important. **British Medical Journal,** 316: 201.

Sloman, J., & Wride, A. (2009) **Economics**. London: Financial Times, Prentice Hall

Smith, R.D. (2003). Construction of the contingent valuation market in healthcare: a critical assessment. **Health Economics**, 12: 609-628

Smith, R.D. (2006). It’s not just what you do, it’s the way that you do it: the effect of different payment card formats and survey administration on willingness to pay for health gain. **Health Economics,** 15: 281-293

Smith, R.D. (2007). The relationship between reliability and size of willingness-to-pay values: a qualitative insight. **Health Economics**, 16: 211-216

Smith, R.D. (2008). Contingent valuation in healthcare: does it matter how the ‘good’ is described? **Health Economics**, 17: 607-617

Stewart, A., Cummins, C., Gold, L., et al. (2001). The effectiveness of the levonorgestrel-releasing intrauterine system in menorrhagia: a systematic review. **BJOG: British Journal of Obstetrics and Gyanecology,** 108 (1): 74–86

Stewart, J.M., O’Shea, E., Donaldson, C., et al. (2002). Do ordering effects matter in willingness-to-pay studies of health care? **Journal of Health Economics,** 21: 585-599

Streiner, D.L., & Norman, G.R. (1995) **Health Measurement Scales: A Practical Guide to their Development and Use**. Oxford: Oxford University Press

Taipale, K.; Leminen, A.; Rasanen,P.; et al. (2009). Costs and health-related quality of life effects of hysterectomy in patients with benign uterine disorders. **Acta Obstetricia et Gynecologica Scandinavica**, 88 (12): 1402-1410

Thompson, S.G., & Barber, J.A. (2000). How should cost data in pragmatic randomised trials be analysed? **British Medical Journal,** 320: 1197-1120

Torrance, G.W. (1976). Social preferences for health states: an empirical evaluation of three measurement techniques. **Socio-economic Planning Science**, 10: 129-136

Torrance, G.W. (1986). Measurement of health state utilities for economic appraisal. **Journal of Health Economics**, 5: 1-30

Tsuchiya, A., Brazier, J., and Roberts, J. (2006). Comparison of valuation methods used to generate the EQ5D and the SF6D value sets in the UK. **Journal of Health Economics**; 25: 334-346.

Tsuchiya, A., & Dolan, P. (2005). The QALY model and individual preferences for health states and health profiles over time: A systematic review of the literature. **Medical Decision Making,** 25: 460-467

Tsuchiya, A., & Williams, A. (2001). Welfare economics and economic evaluation. In: Drummond, M., McGuire, A. (Eds.), **Economic Evaluation in Health Care: Merging theory with practice**. Oxford: Oxford University Press

Van der Kooij, S.M.; Hehenkamp, W.J.; Volkers, N.A.; et al. (2010). Uterine artery embolization vs hysterectomy in the treatment of symptomatic uterine fibroids: 5-year outcome from the randomized EMMY trial. **American Journal of Obstetrics & Gynecology**, 203 (2): 105-113

Van der Wilt, G.J.; Meulendijks, C.F.; and Thijssen, R.F. (2005). Experience with shared decision making in gynaecological practice: treatment decisions in patients with dysfunctional uterine blood loss. **European Journal of Obstetrics, Gynecology, and Reproductive Biology,** 120 (2): 210-216

Von Neumann, J., & Morgenstern, O. (1944). **Theory of games and economic behaviour.** New York: Princeton University press

Vuorma, S.; Teperi, J.; Aalto, A.M.; et al. (2004). A randomized trial among women with heavy menstruation -- impact of a decision aid on treatment outcomes and costs. **Health Expectations,** 7 (4): 327-337

Whitehead, J.C., Groothuis, P.A., and Blomquist, G.C. (1993), 'Testing for Non-response and Sample Selection Bias in Contingent Valuation', **Economic Letters** 41, 215–220.

Whitehurst, D.G., Bryan, S., and Lewis, M. (2011) Systematic Review and Empirical Comparison of Contemporaneous EQ-5D and SF-6D Group Mean score. **Medical Decision Making,** 31 (6): E34-44

Whitehurst, D.G., & Bryan, S. (2011). Another Study Showing that Two Preference-Based Measures of Health-Related Quality of Life (EQ-5D and SF-6D) are not Interchangeable. But why Should we Expect Them to be? **Value in Health**, 14 (4): 531-538

Whynes, D.K., Frew, E., and Wolstenholme, J.L. (2003) A comparison of two methods for eliciting contingent valuations of colorectal cancer screening. **Health Economics**, 22 (4): 555-574

Whynes, D.K., Frew, E.J., Philips, Z.N., et al. (2007). On the numerical forms of contingent valuation responses. **Journal of Economic Psychology**, 28: 462-476

Woman's Health Medicine (2005). Causes of menorrhagia.**Woman's Health Medicine,** 2 (3): 14

Yang, Y., Brazier, J., Tsuchiya, A., et al. (2011). Estimating a Preference-Based Index for a 5-Dimensional Health State Classification for Asthma Derived From the Asthma Quality of Life Questionnaire. **Medical Decision Making**; 31(2): 281-291

You, J.H., Sahota, D.S., and MoYuen, P. (2006). A cost-utility analysis of hysterectomy, endometrial resection and ablation and medical therapy for menorrhagia. **Human Reproduction**; 21 (7): 1878–1883

Zacchariah, M., & Fender, G. (2005). Managing menorrhagia in primary care. **Woman's Health Medicine**, 2 (3): 17-20