Feasibility studies for a trial of septal resection to prevent subfertility and miscarriage.

Matthew Prior

Thesis submitted to The University of Nottingham
for the degree of Doctor of Philosophy

April 2018

Abstract

Introduction

Congenital uterine anomalies result from the abnormal embryological development of the paramesonephric (Müllerian) ducts. Debate surrounding their classification is ongoing. During this PhD project both the American Society for Reproductive Medicine and The European Society of Human Reproduction and Embryology (ESHRE) with the European Society for Gynaecological Endoscopy (ESGE) have published different classification systems.

To adequately assess uterine morphology requires concurrent imaging of the external and internal contours of the uterus. Three-dimensional ultrasound can achieve this view and is more acceptable to women than more invasive tests.

In view of the inconsistent classification systems and use of various diagnostic tests data regarding prevalence and effect on reproductive outcome must be interpreted cautiously. Nonetheless, it appears that uterine septa are more prevalent in women with poor reproductive outcomes and associated with infertility, miscarriage and preterm birth.

The rationale for septal resection is to restore normal uterine anatomy with the intention of improving reproductive outcome. However all previous studies assessing the effect of septal resection are biased due to their observational design using women as their own controls. There have been no randomised controlled trials.

Aims

The aim was to investigate the hypothesis that a trial of septal resection for women with subfertility or miscarriage is feasible to conduct.

Studies

A non-invasive method for assessing uterine septa is a vital prerequisite for a trial, to allow women to be assessed properly and consented before an invasive operative hysteroscopy is done. Furthermore, a test is required to assess the efficacy of septal resection post-procedure as some women are left with a residual septum. So it was important to establish the diagnostic reliability of three-dimensional ultrasound to identify uterine septa. ESHRE recommended using three-dimensional ultrasound, but few data existed demonstrating its reliability. I conducted two reliability studies. The first found that the intra-rater reliability of experts in uterine anomalies was only moderate. This study was limited as it could not differentiate between the manipulation of uterine volumes, uterine measurements or application of classification systems. A second study assessed the intra- and inter-rater reliability of uterine measurements and then used algorithms to classify the volumes. This showed uterine measurements are highly reproducible with cavity indentation showing the highest intra- and inter-rater reliability (both ICC 0.98). Applying classification algorithms to these measurements showed almost perfect intra- and inter-rater agreement.

When designing a trial to definitively establish the efficacy of an intervention it is crucial to ensure it is adequately powered to detect a difference which is important to patients. So next I set out to calculate the minimum clinically important difference for women to consider septal resection. Using survey and interviews I calculated that the median minimum clinically important difference for women with a history of poor reproductive outcome was a change in future risk of miscarriage from 30% to 10%. The sample size required to detect this difference was 118, (59 in each arm) where α=0.05 and (1-β)=0.9.

Between August 2014 and August 2017, 6035 women with subfertility and miscarriage were screened in three settings to identify those eligible to participate. Three-dimensional ultrasound was used as initial screening in a tertiary fertility clinic (n=2846) and recurrent miscarriage clinic (n=189). Around 3000 women underwent two-dimensional ultrasound in the early pregnancy unit, of which five had suspected uterine anomalies and were referred for three-dimensional ultrasound. A uterine septum was present in 15, of whom 13 were eligible to participate in the pilot septal resection trial. So identifying one eligible participant required approximately 464 women to be screened. Divided by patient population, to identify one eligible participant the

number of women to be screened using three-dimensional ultrasound was 407 women with infertility, or 27 women with recurrent miscarriage. Approximately one septum was identified per 1500 women who had a two-dimensional scan in the early pregnancy unit.

Subsequently I undertook three studies of the willingness of patients and clinicians to participate in the pilot trial. First I undertook a survey of clinicians. Sixty-seven clinicians responded and approximately half would randomise to a septal resection trial, but only 12 clinicians (18%) showed interest in becoming involved in such a trial. The next study investigated the willingness of eligible patients to participate. In Nottingham, 2 of 13 women were randomised to the study. A survey and interview with women who declined participation found that many had strong views on their treatment, either in favour of, or against resection and did not wish to leave it to chance. The third study assessed the willingness of other units to be involved in the trial. Twelve clinicians from eleven units expressed interest in becoming a trial site, but only one unit went on to randomise 4 further patients. Reasons for non-participation included lack of funding, exclusion from the NIHR Clinical Research Network Portfolio, inadequate access to three-dimensional ultrasound and local service configuration. During the study period six women were randomised.

The final two chapters set out to establish if a septal resection trial is a research priority using two different methods. Firstly an economic evaluation and value-of-information analysis. Value of information is the amount a decision maker should be willing to pay for information prior to making a decision. As resources are finite, these methods can be used to direct future research efforts towards the most efficient studies. Septal resection, was compared with progesterone as a large study, the PROMISE trial, was recently funded. Using existing evidence, namely meta-analyses of biased studies, it predicted that there was no value in conducting further research into progesterone or septal resection as with existing evidence, they were both cost-effective interventions. Nonetheless, when more uncertainty was introduced based on expert opinion the value of information for both interventions increased. Using the most favourable assumptions for our present level of uncertainty, assuming the usual NHS value of £30,000 for a quality adjusted life year (QALY) the expected value over ten years to the NHS of the information from a progesterone trial was £1.6B compared with £80m for a septal resection trial. The main reason for this was that progesterone has a larger potential treatment population,

all women with recurrent miscarriage, rather than just those with a uterine septum. Other trials are therefore a higher priority.

Finally, I established the top research priorities for miscarriage. Miscarriage research uncertainties were identified and prioritised collaboratively by women and healthcare professionals. A priority setting partnership was conducted on the topic of miscarriage using methodology advocated by the James Lind Alliance, a subsidiary of NIHR. A total of 2402 unique questions were identified and refined to 58 summary questions. The top 10 research uncertainties were collaboratively agreed by women who have experienced miscarriage, those affected by miscarriage and healthcare professionals at a final workshop. The top question was "What are the effective interventions to prevent miscarriage, threatened miscarriage and recurrent miscarriage?". This would include septal resection to prevent miscarriage, but would also be any other potential intervention.

Discussion

The high reliability of three-dimensional ultrasound to diagnose uterine septa was demonstrated. This was dependent on taking uterine measurements, as subjective application of any classification system is less reliable.

The low prevalence of septa and the lack of willingness from patients and clinicians to participate would make a trial prohibitively time consuming or expensive. Furthermore, a uterine septum is not the condition for which resection is performed. Instead resection is to prevent infertility, miscarriage or preterm birth, meaning that ideally three separate trials should be performed.

I demonstrated that funding a trial is unlikely to be the best use of resources as the cost of conducting a trial is high compared with the value of information it could provide to decision makers. However, patients and clinicians still consider further research in this area a priority. Herein lies the dilemma of whether to pursue a trial, which will provide the highest quality evidence but is likely infeasible, or to consider alternative methodology, albeit with limitations, to help guide women and their clinicians who are considering septal resection.

Collaboration with other international groups is the only way to advance this trial further. For example, the Dutch TRUST trial. TRUST has also had difficult recruiting, but has randomised 43 women over nine years. A

collaborative approach between the two studies may lead to the completion of a trial. Nonetheless, TRUST is not adequately powered to detect the minimum clinically important difference and has less stringent criteria for diagnosis of a uterine septum.

An alternative approach could be a database registering women with uterine anomalies and three-dimensional ultrasound volumes and recording subsequent reproductive outcomes. This methodology would allow independent analysis of uterine measurements which could provide more accurate data on prevalence and effect on reproductive outcome. This could lead to the development of a classification system based on prognostic information. Furthermore, this approach would collect valuable data from women with uterine septa who do not wish to be randomised, instead choosing whether or not to undergo septal resection.

Conclusion

Presently, a randomised controlled trial of hysteroscopic septal resection for women with a history of subfertility or miscarriage and a uterine septum is not feasible. The reasons for this are that sufficient women and clinicians are not willing to participate to achieve the required sample size. Furthermore the resource required to recruit enough participants would make the trial prohibitively expensive.

Acknowledgements

Completing a PhD while working as a senior registrar in Obstetrics and Gynaecology and Chairman of the Royal College of Obstetricians and Gynaecologists Trainees' Committee has been challenging. I would not have got to this point without the help and support of others, for which I am deeply grateful.

I would like to thank Nottingham University Hospitals and Nurture Fertility for giving me the opportunity to pursue this Ph.D. and to Nottingham Hospitals Charity for funding the pilot septal resection trial.

I am enormously grateful to my PhD supervisors: Mr. Nick Raine-Fenning and Professor Jim Thornton who have provided support, guidance, and wisdom to allow me to complete this thesis.

Each chapter has relied on the contribution and collaboration of others. I am grateful to Povilas Sladkevicius, Caterina Exacoustos, Juan Luis Alcázar, Thierry Van den Bosch and Laurentiu Craciunas for their contribution to studies assessing the reliability of three-dimensional ultrasound. Identifying women eligible to participate in the pilot septum trial would not have been possible without the help from Judith Moore who referred women from Nottingham and the enthusiasm of Elissa Baskind, Adam Balen, and Tony Rutherford to run the trial in Leeds.

I would also like to thank collaborators of the the Miscarriage Priority Setting Partnership. This includes the steering group, but especially Leanne Metcalf, Arri Coomarasamy, Rachel Small, Lesley Regan and Jane Brewin, who supported, encouraged and participated in the project.

Value of information analysis was a steep learning curve introducing the theoretical and mathematical concepts from Bayesian cost-effectiveness analysis. I would not have understood without guidance from Tracey Sach and Ed Wilson. I want to especially thank Ed Wilson, for his patience and time

helping me to develop, code and run economic models.

Finally, this Ph.D. has been the main focus of my work and life for the past three years. I am ever thankful to my wife, Pooja, my son, Krishnan, and the rest of my family who have supported me to complete this work.

Abbreviations

α	Alpha (type 1 error)
β	Beta (type 2 error)
δ	Delta (minimum clinically important difference)
κ	Kappa statistic
3D	three-dimensional
AFS	American Fertility Society
ASRM	American Society for Reproductive Medicine
£B	billion pounds
BMI	body mass index
CEAC	cost-effectiveness acceptability curve
CEO	chief executive officer
CI	confidence interval
CONUTA	CONgenital UTerine Anomalies working group
CRN	clinical research network
CUA	congenital uterine anomalies
DARE	The Database of Abstracts of Reviews of Effects
EPAU	Early Pregnancy Assessment Unit
ESGE	European Society for Gynaecological Endoscopy
ESHRE	The European Society of Human Reproduction and Embryology
EU	European Union
EVPI	expected value of information
EVPPI	expected value of partial perfect information
EVSI	estimated value of sample information
HSG	hysterosalpingogram
HTA	Health Technology Assessment
ICC	intra class correlation coefficient
ICER	incremental cost-effectiveness ratio
IQR	inter quartile range

IVF	in-vitro fertilisation
JLA	James Lind Alliance
£m	million pounds
MCID	minimum clinically important difference
mm	millimetres
MRC	Medical Research Council
MRI	magnetic resonance imaging
NHS	National Health Service
NICE	The National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NNT	number needed to treat
NPV	negative predictive value
OR	odds ratio
PCOS	polycystic ovary syndrome
PPV	positive predictive value
PROMISE	Progesterone in recurrent miscarriage trial
PSA	probabilistic sensitivity analysis
PSP	priority setting partnership
QALY	quality adjusted life year
RCOG	The Royal College of Obstetricians and Gynaecologists
REC	Research Ethics Committee
RM	recurrent miscarriage
RR	risk ratio
SD	standard deviation
SEPTUM	Randomised, controlled study of hysteroscopic metroplasty for women with a septate uterus and a history of miscarriage or preterm labour
TNM	tumour node metastasis classification
TRUST	The Randomised Uterine Septum Transsection Trial
VCI	volume contrast imaging
VCUAM	Vagina Cervix Uterus Adnex-associated Malformation classification system
VOI	value of information

Publications

Prior M, Raine-Fenning N. Response to: "hysteroscopic metroplasty: reproductive outcome in relation to septum size" by Paradisi et al. *Arch Gynecol Obs*. 2014;290(2):203.

Prior M, Hibberd R, Asemota N et al. Inadvertent p-hacking among trials and systematic reviews of the effect of progestogens in pregnancy? A systematic review and meta-analysis. *BJOG*, 2017. 124: 1008—-1015. doi:10.1111/1471-0528.14506.

Prior, M, & Raine-Fenning, N. Uterine Anomalies and Early Pregnancy. In E. Kirk (Ed.), *Early Pregnancy Ultrasound: A Practical Guide* (pp. 69-74). Cambridge: Cambridge University Press, 2017. doi:10.1017/9781316481776.010.

Prior M, Bagness C, Brewin J et al. Priorities for research in miscarriage: a priority setting partnership between people affected by miscarriage and professionals following the James Lind Alliance methodology. *BMJ Open*, 2017. 7: e016571. doi:10.1136/bmjopen-2017-016571.

Prior M, Richardson A, Asif S et al. Reproductive outcome of women with congenital uterine anomalies after assisted reproduction: A prospective observational study. *Ultrasound Obstet Gynecol*, 2017. [In press]. doi:10.1002/uog.18935.

Presentations

British Society of Gynaecological Imaging. London. April 2015.

Uterine Septa. When to resect. (Oral)

South Wales and South West Infertility Group. Cardiff. November 2015.

Uterine Septa. When to resect. (Oral)

Guy's & St Thomas' Assisted Conception Unit. November 2015.

The SEPTUM trial. (Oral)

European Society of Human Reproduction and Embryology (ESHRE). Warwick. January 2016.

Pilot RCT of hysteroscopic septal resection for women with a history of infertility, miscarriage or preterm birth. (Oral)

Fertility Winter Symposium. St Anton, Austria. March 2016,

Uterine Septa. When to resect. (Oral)

Leeds Fertility. May 2016.

The SEPTUM trial. (Oral)

Royal College of Obstetricians and Gynaecologists (RCOG) World Congress. Birmingham June 2016.

Uterine Anomalies: are they as bad as we think? (Oral)

Royal College of Obstetricians and Gynaecologists (RCOG) Annual academic meeting. London. March 2017.

P-hacking among trials and systematic reviews of the e ect of progestogens in pregnancy. (Oral)

Fertility Winter Symposium. Saalbach. March 2017.

Reproductive outcome of women with congenital uterine anomalies after assisted reproduction treatment. (Oral)

Association of Early Pregnancy Units Annual Scientific Meeting. Warwick. November 2017.

The Miscarriage Priority Setting Partnership: A collaboration between people affected and professionals to prioritise research questions. (Oral)

Royal College of Obstetricians and Gynaecologists (RCOG) Annual Academic Meeting. London. November 2017.

Progestogens in pregnancy. Does the data add up? (Oral)

Contents

1 Literature review **1**

1.1 Introduction 1

1.2 Uterine septa 2

1.2.1 Embryology 2

1.2.2 Classification 3

1.2.3 Diagnosis 9

1.2.4 Prevalence 15

1.2.5 Effect on reproductive outcome 18

1.2.6 Pathogenesis 20

1.2.7 Summary 21

1.3 Septal resection 22

1.3.1 Surgical technique 22

1.3.2 Safety 23

1.3.3 Efficacy 27

1.3.4 After resection 30

1.3.5 Summary 31

1.4 Hypotheses 33

2 Diagnostic reliability of three-dimensional ultrasound **35**

2.1 Introduction 35

2.2 Study 1: Off-line Analysis of ultraSonographic Images Study (OASIS) 36

2.2.1 Introduction . . . 36
2.2.2 Methods . . . 36
2.2.3 Results . . . 41
2.2.4 Discussion . . . 43
2.3 Study 2: Intra- and inter-rater reliability of uterine measurements with three-dimensional ultrasound . . . 46
2.3.1 Introduction . . . 46
2.3.2 Methods . . . 46
2.3.3 Results . . . 54
2.3.4 Discussion . . . 59
2.4 Conclusion . . . 61

3 Sample size **63**

3.1 Introduction . . . 63
3.2 Methods . . . 68
3.2.1 Research questions . . . 68
3.2.2 Setting . . . 68
3.2.3 Design . . . 68
3.2.4 Study population . . . 69
3.2.5 Study conduct . . . 70
3.2.6 Data collection . . . 70
3.2.7 Statistical and data analysis . . . 70
3.2.8 Ethics and governance . . . 71
3.3 Results . . . 71
3.3.1 Minimum Clinically Important Difference . . . 73
3.3.2 Sample size calculations . . . 75
3.4 Discussion . . . 76
3.4.1 Main findings . . . 76
3.4.2 Strengths and limitations . . . 76

3.4.3 Interpretation . . . 78
3.4.4 Conclusions . . . 80

4 Screening **81**

4.1 Introduction . . . 81
4.2 Methods . . . 82
4.2.1 Research question . . . 82
4.2.2 Setting . . . 82
4.2.3 Design . . . 83
4.2.4 Study population . . . 85
4.2.5 Study conduct . . . 87
4.2.6 Variables, bias and study size . . . 87
4.2.7 Data collection . . . 88
4.2.8 Statistical and data analysis . . . 88
4.2.9 Ethics and governance . . . 88
4.3 Results . . . 88
4.4 Discussion . . . 94
4.4.1 Conclusion . . . 97

5 Participation of patients, clinicians and units **99**

5.1 Study 1: Clinician opinion of hysteroscopic septal resection . . . 99
5.1.1 Introduction . . . 99
5.1.2 Aims . . . 100
5.1.3 Methods . . . 100
5.1.4 Results . . . 101
5.1.5 Discussion . . . 102
5.2 Study 2: Willingness of patients to be randomised . . . 108
5.2.1 Introduction . . . 108
5.2.2 Aims . . . 108
5.2.3 Methods . . . 108

5.2.4 Results . . . 110
5.2.5 Discussion . . . 113
5.3 Study 3: Exploring the potential for multiple trial sites . . . 115
5.3.1 Introduction . . . 115
5.3.2 Aim . . . 115
5.3.3 Methods . . . 115
5.3.4 Results . . . 116
5.3.5 Discussion . . . 118
5.4 Discussion and Conclusions . . . 119
5.4.1 Conclusion . . . 120

6 Economic evaluation and value of information analysis 121

6.1 Introduction . . . 121
6.1.1 Miscarriage . . . 122
6.1.2 Progesterone to prevent recurrent miscarriage . . . 123
6.2 Study 1: A value of information analysis of progestogen for recurrent miscarriage - before and after PROMISE. . . . 126
6.2.1 Methods . . . 126
6.2.2 Results . . . 137
6.2.3 Discussion . . . 143
6.3 Study 2: A value of information analysis of hysteroscopic septal resection for infertility and miscarriage . . . 147
6.3.1 Methods . . . 148
6.3.2 Results . . . 152
6.3.3 Discussion . . . 157
6.4 General conclusions . . . 158

7 Miscarriage research prioritisation 161

7.1 Introduction . . . 161
7.2 Aims . . . 162

7.3 Methods 162
7.4 Results 169
7.5 Discussion 174

8 Conclusions and further work **179**

8.1 Strengths and limitations 183
8.2 Future research 184

References **186**

Appendices **207**

1 Ethical approval 209
2 Off-line Analysis of ultraSonographic Images Study (OASIS): Uterine Anomalies 210
3 Sample Size Survey 212
4 Patient Survey 216
5 Clinician Survey 221
6 Investigator Briefing 225
7 Identifying the top research priorities for miscarriage 241
8 All refined questions 248

List of Figures

1.1 Genital ducts in the sixth week. 2

1.2 Formation of uterine tube. 2

1.3 Spectrum of uterine morphology, with comparison of classification systems. 8

1.4 Measurements required to diagnose and classify a uterine septum. 9

1.5 Meta-analysis of studies of hysteroscopic septal resection for the outcome of pregnancy rate. 29

1.6 Meta-analysis of studies of hysteroscopic septal resection for the outcome of miscarriage. 29

1.7 Meta-analysis of studies of hysteroscopic septal resection for the outcome of term birth. 30

2.1 Three-dimensional multiplanar display of the uterus. 39

2.2 Classification systems used by expert raters. 41

2.3 Measurements required to classify uterine morphology. 49

2.4 Algorithm for Salim classification of uterine anomalies. 51

2.5 Algorithm for ASRM classification of uterine anomalies. 51

2.6 Algorithm for ESHRE-ESGE classification of uterine anomalies. 51

2.7 Bland–Altman plots of intra-rater reliability. 57

2.8 Bland–Altman plots of inter-rater reliability. 58

3.1 Minimum clinically important difference. 74

4.1 Screening flow diagram. 89

4.2 Uterine measurements of eligible septa. 93

5.1 What is the gold standard test used in your unit to diagnose and define the presence and type of uterine anomaly? 103

5.2 Availability of three-dimensional ultrasound and hysteroscopic septal resection. 103

5.3 Which classification systems do you use to diagnose uterine anomalies? 103

5.4 Frequency of septal resection. 104

5.5 How confident are you in your response to the number of septa you have seen and managed in the last year? 104

5.6 Septal resection surgical technique. 104

5.7 For which conditions have you performed septal resection? . . . 105

5.8 For which conditions would you be willing to randomise a patient to either septal resection or control? 105

6.1 Decision tree model structure of progestogen to prevent miscarriage. 126

6.2 Meta-analysis of trials of progestogen in recurrent miscarriage for the outcome of miscarriage. 128

6.3 Cost-effectiveness plane for progestogen to treat recurrent miscarriage using meta-analysis prior. 138

6.4 Cost-effectiveness plane for progestogen to treat recurrent miscarriage using a prior derived from expert opinion. 138

6.5 Cost-effectiveness acceptability curve for progestogen to prevent recurrent miscarriage using the meta-analysis prior. . . 139

6.6 Cost-effectiveness acceptability curve for progestogen to prevent recurrent miscarriage using the prior derived from expert opinion. 140

6.7 Expected value of perfect information (EVPI) for progestogen to prevent recurrent miscarriage using the meta-analysis prior. . 140

6.8 Expected value of perfect information (EVPI) for progestogen to prevent recurrent miscarriage using the prior derived from expert opinion. 141

6.9 Estimated Value of Sample Information (EVSI) for progestogen to prevent recurrent miscarriage. 142

6.10 Progestogen versus placebo: primary registered dichotomous outcome. 146

6.11 Progestogen versus placebo: primary registered continuous outcome. 147

6.12 Decision tree model structure of septal resection to prevent miscarriage. 148

6.13 Meta-analysis of studies of hysteroscopic septal resection for the outcome of term birth. 150

6.14 Cost-effectiveness plane for septal resection to prevent miscarriage using meta-analysis prior. 153

6.15 Cost-effectiveness plane for septal resection to prevent miscarriage using prior derived from expert opinion. 153

6.16 Cost-effectiveness acceptability curve and expected value of perfect information for septal resection to prevent miscarriage using meta-analysis prior. 154

6.17 Cost-effectiveness acceptability curve and expected value of perfect information for septal resection to prevent miscarriage using expert opinion prior. 154

6.18 Expected value of perfect information (EVPI) for septal resection to prevent miscarriage using the meta-analysis prior. . 155

6.19 Expected value of perfect information (EVPI) for septal resection to prevent miscarriage using the prior derived from expert opinion. 156

6.20 Estimated Value of Sample Information (EVSI) of a trial of septal resection to prevent miscarriage. 157

7.1 Summary of the James Lind Alliance Priority Setting process. . 165

7.2 Flow diagram of results. 170

List of Tables

1.1 Salim criteria for the classification of congenital uterine anomalies. 5

1.2 Complications after hysteroscopic septal resection. 26

1.3 Summary of systematic reviews and meta-analyses of congenital uterine anomalies. 32

2.1 The Kappa statistic. 40

2.2 Diagnosis of uterine volumes for each rater. 42

2.3 Fleiss-Kappa (κ) inter-rater agreement for diagnosis of uterine morphology by classification system used. 43

2.4 Uterine measurements. 49

2.5 Descriptive statistics of uterine measurements. 54

2.6 The intra-rater reliability of data measurement for uterine morphology. 55

2.7 The inter-rater reliability of data measurement for uterine morphology. 56

2.8 Intra-rater reliability of classification of uterine morphology. . . 56

2.9 Inter-rater reliability of classification of uterine morphology. . . 56

3.1 Participant demographics, before and after exclusions. 72

3.2 Reasons for exclusion. 73

3.3 Sample sizes for a trial of septal resection 76

3.4 Number needed to treat (NNT) for interventions commonly used in obstetrics and gynaecology. 79

4.1 Characteristics of patients with a uterine septum. 91

4.2 Demographics and uterine measurements of eligible women. . . 93

5.1 Reasons for patient non-participation in a septal resection trial. 111

5.2 Reasons for site non-participation in a septal resection trial. . . 117

6.1 Probability values of progestogen to prevent miscarriage. 129

6.2 Cost of alternative progestogen preparations. 130

6.3 Care pathway costs and activity levels. 131

6.4 Cost values. 133

6.5 Probability values of miscarriage with a uterine septum. 149

7.1 Miscarriage Priority Setting Partnership steering group members. 164

7.2 Taxonomy for miscarriage questions. 168

7.3 Analysis by demographic. 172

7.4 Top 10 ranked research questions for miscarriage. 173

7.5 Response rate to James Lind Alliance priority setting partnerships. 176

7.6 Themes from pregnancy priority setting partnerships. 177

8.1 Summary of research findings: Chapter 2 180

8.2 Summary of research findings: Chapter 3 180

8.3 Summary of research findings: Chapter 4 181

8.4 Summary of research findings: Chapter 5 181

8.5 Summary of research findings: Chapter 6 182

8.6 Summary of research findings: Chapter 7 182

Chapter 1

Literature review

1.1 Introduction

Practising evidenced-based medicine relies on the availability of high quality studies to demonstrate interventions are safe and effective. Randomised controlled trials are considered one of the most powerful tools used in research to establish new knowledge claims [1].

Recently, the field of reproductive medicine has been criticised for offering interventions and products to patients without evidence to demonstrate their safety or effectiveness [2]. Examples of such add-ons are intralipid infusions, endometrial scratch and pre-implantation genetic screening. History also warns us of the disastrous consequence from using inadequately tested interventions such as thalidomide and diethylstilbestrol.

Congenital uterine anomalies are more prevalent in women with a history of miscarriage and evidence suggests these women are more likely to have higher rates of pregnancy loss than those with a normal uterus. Presently, a uterine septum is the only congenital uterine anomaly with a potentially effective treatment—surgical resection.

No randomised controlled trials have been done to investigate the efficacy of septal resection. Therefore a definitive randomised controlled trial is required to demonstrate septal resection is safe and effective. This thesis assesses the feasibility of conducting such a trial.

This literature review focusses on the highest levels of evidence. Namely systematic reviews and meta-analyses of observational studies. Nonetheless, where appropriate I also review the original studies to provide depth.

1.2 Uterine septa

Congenital uterine anomalies are a spectrum of abnormal configurations arising from the abnormal embryological development of the paramesonephric (Müllerian) ducts. A uterine septum is one such anomaly and has been associated with adverse reproductive outcomes. Presently, a uterine septum is the only congenital uterine anomaly with a potentially effective treatment—surgical resection.

1.2.1 Embryology

The internal reproductive system begins its development in the sixth week, originating from the common mesodermal ridge along the posterior wall of the abdominal cavity, shown in Figure 1.1. Initial development is indifferent between male and female embryos. Both have two pairs of genital ducts: mesonephric (Wolffian) ducts and paramesonephric (Müllerian) ducts [3].

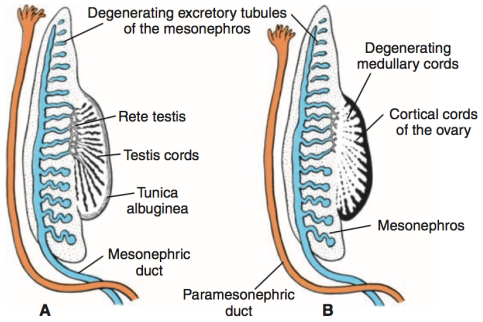


Figure 1.1: Genital ducts in the sixth week in the male (A) and female (B). The mesonephric and paramesonephric ducts are present in both. From Langman's Medical Embryology [3].

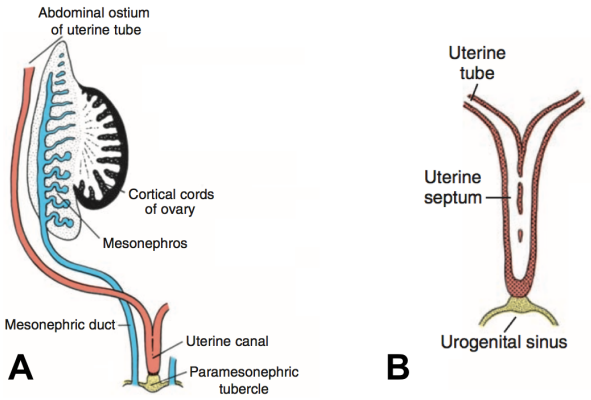


Figure 1.2: A: Genital ducts at the end of the eight week. Note the paramesonephric tubercle and the formation of the uterine canal. B: Septal resorption forms of uterine tube by the end of the ninth week. From Langman's Medical Embryology [3].

The paramesonephric duct has a funnel like opening in the abdominal cavity and runs caudolaterally before growing towards the midline. In females, the absence of testosterone and anti-Müllerian hormone results in atresia of the mesonephric ducts which act as a guide for further development of the paramesonephric ducts [4].

By the end of the eighth week the paramesonephric ducts fuse to form the uterine tube, initially divided by a septum ending at the paramesonephric tubercle. Fusion and resorption begin at the level of the uterine isthmus tubes and progress in both the cranial and caudal direction [5]. By nine weeks gestation, the uterine septum has resorbed resulting in a single uterine cavity shown in Figure 1.2.

Failure of normal embryological uterine development leads to a range of anomalies dependent on where the failure occurs [6]. Failure or inadequate of fusion of the paramesonephric ducts results in a unicornuate or bicornuate uterus respectively. Inadequate septal resorption in its mildest form, leads to an arcuate uterus, where there is a mild indentation of the uterine wall into the uterine cavity. When septal resorption is more profound a uterine septum remains which can extend as far as the internal cervical os.

1.2.2 Classification

Categorisation of pathology is common in medicine to standardise and stratify diagnosis, and provide a common language for capturing, sharing and collecting data for clinicians and researchers. An ideal classification system should be based on aetiology, pathogenesis and should incorporate prognostic information to suggest appropiate treatments. It must also be simple to apply, but have high diagnostic accuracy and inter-rater and intra-rater reliability [7]. At least 26 classification systems for uterine anomalies have been published [7–16].So I will discuss those most widely adopted and show there is no universally accepted standard definition of a septate uterus.

American Fertility Society (AFS) [9]

The American Fertility Society Classification, now known as the American Society for Reproductive Medicine, is currently the most widely used classification system for Müllerian anomalies. Published in 1988, this system

encompasses all anomalies of the female reproductive tract. It does not attempt to strictly classify anomalies, instead publishing a series of diagrams to aid categorisation. Benefits include its simplicity and flexibility to allow for many permutations of anomalies. Nevertheless, the normal uterus is not defined and morphometric criteria are not provided, making it possible for clinicians to assign the same anomaly to different categories. For example, a uterus with an internal fundal indentation may be categorised as normal, arcuate or subseptate, depending on the rater.

Modified AFS classification of congenital anomalies by Salim criteria [10]

Salim added additional supplementary criteria to the AFS classification to enable discrimination between normal, arcuate, subseptate, septate and bicornuate uteri. Using this classification and three-dimensional ultrasound showed a good level of agreement between two different operators when diagnosing congenital uterine anomalies (kappa=0.97) [10]. Combining this classification with a single diagnostic test may improve standardisation, but there is no reason why this classification cannot be used with other diagnostic tests. These criteria are shown in Table 1.1.

The European Society of Human Reproduction and Embryology and the European Society for Gynaecological Endoscopy (ESHRE-ESGE) [15]

Between publication of Salim's modified criteria in 2003 and the ESHRE-ESGE classification, two systems based on anatomy and embryology were proposed. Firstly, the VCUAM (Vagina Cervix Uterus Adnex-associated Malformation) classification is similar to the tumour, node, metastasis (TNM) classification for tumours, and a second based on embryology [12, 14]. These classifications were not widely adopted, so are not further discussed.

Recognising the limitations of previous classifications in terms of effective categorisation, clinical usefulness, simplicity and friendliness the European Society of Human Reproduction and Embryology (ESHRE) and the European Society for Gynaecological Endoscopy (ESGE) established a common working group under the name CONUTA (CONgenital UTerine Anomalies) with the aim of devising a new classification system.

Table 1.1: Salim criteria for the classification of congenital uterine anomalies [10].

Uterine morphology	*Fundal contour*	*External contour*
Normal	Straight or convex	Uniformly convex or with indentation $<$10 mm
Arcuate	Concave fundal indentation with central point of indentation at obtuse angle ($>90^\circ$)	Uniformly convex or with indentation $<$10 mm
Subseptate	Presence of septum, which does not extend to cervix, with central point of septum at an acute angle ($<90^\circ$)	Uniformly convex or with indentation $<$10 mm
Septate	Presence of uterine septum that completely divides cavity from fundus to cervix	Uniformly convex or with indentation $<$10 mm
Bicornuate	Two well-formed uterine cornua	Fundal indentation $>$10 mm dividing the two cornua
Unicornuate with or without rudimentary horn	Single well-formed uterine cavity with a single interstitial portion of Fallopian tube and concave fundal contour	Fundal indentation $>$10 mm dividing the two cornua if rudimentary horn present

Developed by consensus using a DELPHI procedure involving 118 experts in round 1 and 71 in round 2, the final system returned the focus to the uterus, based on anatomy rather than embryology. Recognising normal uteri vary in size between one patient and another, ESHRE-ESGE introduced the principle of ratios between uterine measurements, rather than discrete limits used in previous systems. For example, a partial septate uterus is defined as an internal indentation greater than 50% of the uterine wall thickness. However, no further rationale was given for this parameter, its reliability had not been assessed and the initial publication failed to specify how to measure the uterine wall thickness, until clarified in a further publication [17].

Advantages include being able to categorise complex anomalies, such as the bicorporeal septate uterus (U3c), or co-existing vaginal and cervical anomalies. However, despite being developed by consensus, it is questionable whether this classification meets the prespecified aims of clinical usefulness, simplicity and friendliness. None of the categories are based on evidence for the prevalence, or effect of anomaly on reproductive outcome and the initial confusion over how to measure uterine wall thickness confused some clinicians [18]. Furthermore, the removal of the arcuate category has resulted in a considerable overlap with partial septate uteri (U2a) in the ESHRE-ESGE, with normal and arcuate uteri in other classifications as shown in Figure 1.3. Applying the ESHRE-ESGE classification results in uterine septa being diagnosed more frequently and concerns have been raised that variation between different diagnostic classifications will result in a higher incidence of septal resection despite limited evidence to support it [19].

American Society for Reproductive Medicine [16]

The society's practice committee uterine septum guideline uterine septum proposed definitions for normal / arcuate, septate, and bicornuate uteri. No differentiation exists between normal and arcuate uteri, and the subseptate category has been dropped. A uterine septum is defined as an internal indentation depth greater than 1.5cm, a normal / arcuate uterus can have an indentation up to 1cm, however there is no guidance on where an indentation between 1 and 1.5cm should be categorised. Indeed despite stating these limits were based on assessment of available literature, the 1.5cm value has not been used in any previous classification system reported in the literature.

Other than being devised by the ASRM practice committee, no wider

consultation of clinicians occurred. Similar limitations to the ESHRE-ESGE classification are that none of the limits are based on data regarding prevalence or effect on reproductive outcome. In contrast to the ESHRE-ESGE classification, this system may classify significant fundal indentations as normal / arcuate, despite limited evidence to demonstrate the arcuate uterus behaves benignly.

Comparison

All four classifications are summarised in Figure 1.3, showing the spectrum of uterine anomalies. Uterine measurements for various classification systems are shown in Figure 1.4. Categories used in all four classifications use similar descriptors, but each define these terms differently, easily leading to confusion of both clinicians and researchers.

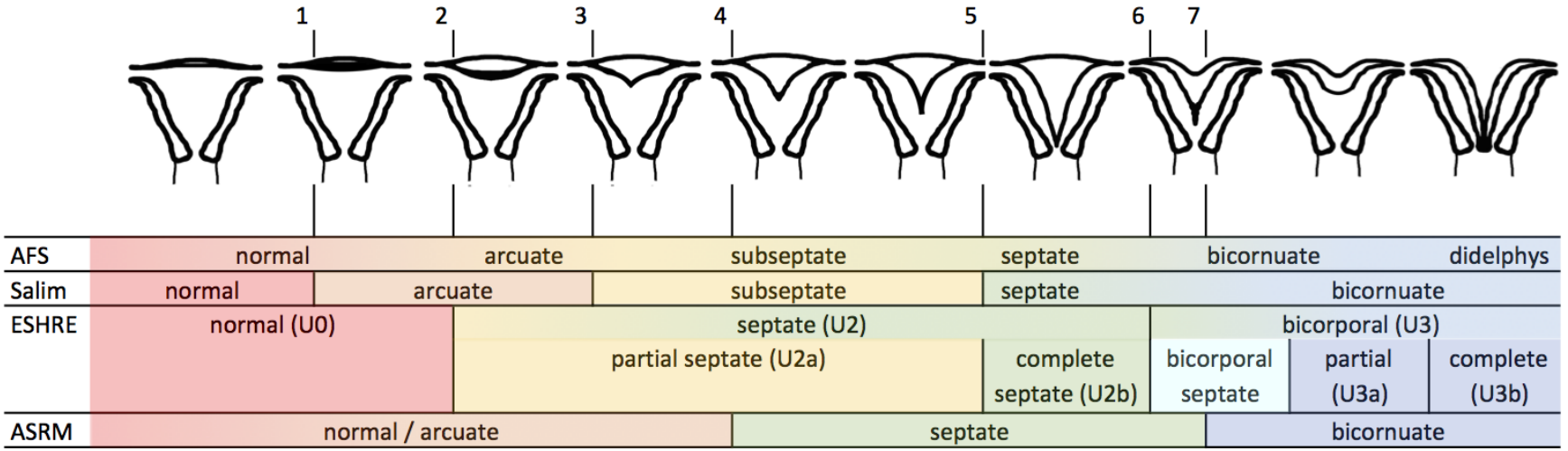


Figure 1.3: Spectrum of uterine morphology, with comparison of classification systems.

1 Internal fundal indentation
2 Internal indentation greater than 50% of the uterine wall thickness
3 Internal indentation greater than 10mm
4 Internal indentation greater than 15mm
5 Septum completely divides the uterine cavity
6 External indentation greater than 50% of the uterine wall thickness
7 External indentation greater than 10mm

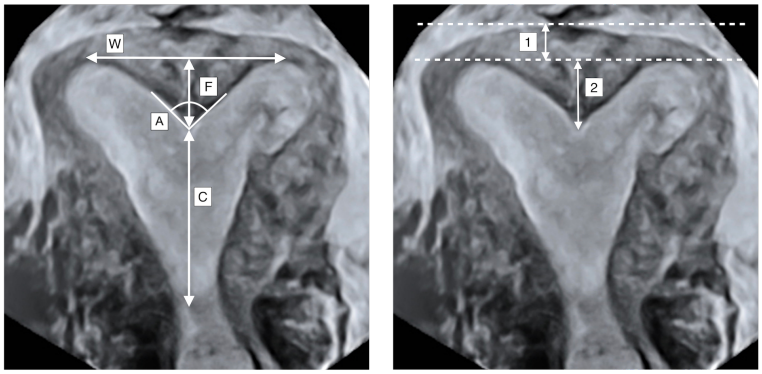


Figure 1.4: Measurements required to diagnose and classify a uterine septum. The left image illustrates uterine cavity width (W), fundal distortion (F), angle of cavity indentation (A), and the length of the unaffected uterine cavity (C) used by the AFS, Salim and ASRM classification. The right image illustrates (1) uterine wall thickness: distance between the line joining tubal ostia (interostial line) and a parallel line on the top of uterine fundus, and (2) internal midline indentation: distance between the interostial line and a parallel line on the top of midline indentation. These measurements are used for ESHRE-ESGE classification.

1.2.3 Diagnosis

Numerous tests can be used to assess the shape of the uterus, including: clinical examination, often at the time of surgery, endoscopic techniques, hysterosalpingogram (HSG), ultrasound and magnetic resonance imaging. First principles dictate that to accurately diagnose a uterine anomaly it is necessary to simultaneously see the external contour of the uterus and the fundal aspect of the endometrial cavity. Tests not meeting these criteria must be considered suboptimal. Suboptimal tests may be able to identify most types of anomaly but cannot accurately differentiate between anomalies, instead being more suitable as screening tools prior to performing an optimal test for accurate diagnosis [20]. Furthermore an optimal test must be reproducible between clinicians, providing a reliable result whoever performs the test.

Clinical Examination

Assessment of the location, shape, and size of the uterus is a routine part of bimanual examination [21]. Comparing clinical examination with surgical findings showed sensitivities of 0.64 and 0.62 with specificities of 0.8 and 0.78 for uterine size and uterine contour respectively [22]. Postgraduate experience improves the overall accuracy of pelvic examination showing higher accuracy for doctors specialising in gynaecology compared with medical students. Nonetheless, studies assessing the accuracy of clinical examination

of uterine size and shape have not differentiated between fibroids, uterine anomalies and other uterine masses [23]. Therefore it must be assumed that the accuracy of physical examination in detecting uterine anomalies is much lower. For example, it would be impossible to differentiate between a normal, arcuate, subseptate and septate uterus by bimanual examination because the endometrial cavity cannot be assessed.

Simultaneous examination of both the external contour and fundal aspect of the endometrial cavity is possible at the time of caesarean section. However this technique is limited due to the massive transformation of the uterus due to pregnancy, resulting in distorted morphology compared to the non pregnant state. The relevance uterine morphology during pregnancy must also be considered and, despite stating the obvious, it is not possible to use this technique to identify a uterine septum in women who do not achieve this gestation of pregnancy.

Endoscopy

Laparoscopy and hysteroscopy are playing an increasingly important role in the management of women's health problems and have not only replaced many open surgical procedures but also revolutionised how many conditions are diagnosed and treated. Direct visualisation of the uterus obviously allows for assessment of uterine morphology and modern optics allow for high-resolution images through smaller scopes. It is now possible to see high quality images of the outside of the uterus through a 5mm laparoscope and endometrial cavity using a 2mm diagnostic hysteroscope. Nonetheless, endoscopy is associated with some limitations. Firstly, both laparoscopy and hysteroscopy are invasive techniques and carry associated risks. Secondly, using either technique alone cannot adequately assess morphology, as in isolation they only assess the inner, or outer, aspect of the uterus. Optimal assessment required the procedures to be performed concurrently [20]. Thirdly, viewing the coronal plane is not possible with either technique, so measurements such as interostial distance, septal length and angle, or the fundal thickness cannot be made. These measurements are important to assess diagnostic reliability and comparison between studies. Finally, endoscopy because of its nature, only provides two-dimensional views of a three-dimensional organ, and even if both procedures are video recorded it is difficult to interpret these at a later date [24].

Hysterosalpingography

Using x-rays, a hysterosalpingogram (HSG) looks at the inside of the uterus and fallopian tubes and the areas around them. The National Institute for Health and Care Excellence guidelines recommend HSG to screen for tubal occlusion in the absence of comorbidities [25]. HSG is performed between 7th and 10th days of a menstrual cycle without anaesthesia and is it is considered to be very painful by about 70%–80% of women undergoing it [26]. Furthermore, HSG exposes the woman to radiation and contrast material. Coronal images of the endometrial cavity and fallopian tubes are obtained by injecting radio-opaque contrast into the cervical canal and taking a series of radiographs of the pelvis. It is therefore a good test in demonstrating that the uterus is normal and that the fallopian tubes are patent, however HSG cannot distinguish between a uterine septum and a bicornuate uterus as the external contour of the uterus cannot be assessed [11].

Ultrasound

Transvaginal ultrasonography to assess the uterus has become commonplace in gynaecology and is regarded by some clinicians as an extension of physical examination. Its popularity is due to its simple, quick and relatively non-invasive nature and is readily accessible in clinical practice. Ultrasonographic assessment of the uterus typically involves determining the version and flexion of the uterus, measuring its size and endometrial thickness. Most sonographers use two-dimensional ultrasonography to assess these characteristics. The introduction of three-dimensional ultrasonography provides further information regarding uterine morphology and allows the storage of volumes for future off-line analysis.

Two-dimensional ultrasound

Ultrasound has been used to assess uterine morphology since the 1960s. Two-dimensional transvaginal ultrasound is commonly preferred over the transabdominal approach because it allows for the use of higher frequency probes and therefore higher resolution images without the need for a full bladder. Transvaginal ultrasound has high acceptance among women although it does have the limitation of not being suitable for women who have not been previously sexually active and is unable to assess larger pelvic masses such

as large fibroids. Two-dimensional ultrasound is well established as a reliable diagnostic modality in evaluation of the pelvic organs. Nevertheless, the planes of the pelvic organs it captures are limited to transverse and sagittal sections. As previously described to accurately assess uterine morphology it is necessary to simultaneously assess the external fundal contour of the uterus and upper aspect of the endometrial cavity, this is only possible in the coronal plane. Used as a screening test, transvaginal ultrasonography has provided sensitivity rates up to 100% [27], but the distinction between different types of anomalies is often impossible [28]. Consequently a two-dimensional sonographer may suspect a uterine anomaly but not be able to confirm it and is unable to differentiate a bicornuate from a septate uterus.

Three-dimensional ultrasound

Three-dimensional ultrasound provides views of the uterus in the coronal plane with the added benefit of being low cost and having high patient tolerability [29]. Acquisition involves a series of two-dimensional images taken from a pre-selected region of interest. These images have relative positional coordinates and are reconstructed by the machine in three planes. Subsequently the volume can be be manipulated to enable the clinician to view the uterus in the coronal plane. In Section 2.3.2 of the next chapter, I discuss the three-dimensional ultrasound technique used to acquire, manipulate and interpret uterine volumes to diagnose a uterine septum.

The reliability of three-dimensional ultrasound for the diagnosis of uterine anomalies has been tested in three published studies. Firstly, in the paper which also presented the modified AFS classification by Salim et al [10]. Two raters assessed 83 uterine volumes and categorised them using the aforementioned classification. Measurements were also taken for the uterine cavity width, fundal distortion and the length of the unaffected uterine cavity, shown in Figure 1.4. Reliability was assessed using the Kappa statistic for categorical variables and intra-class correlation co-efficient for continuous variables. These statistical tests are explained in Chapter 2. Combining three-dimensional ultrasound with a single classification to categorise anomalies showed almost perfect agreement (κ=0.97). Intra-class correlation coefficients were also satisfactory with limits of agreement $\pm$1.43 to $\pm$2.51 mm. Limitations of this study are that the two raters came from the same unit, and used criteria derived themselves. It it likely they were trained together, or by one another, questioning the application of this method outside of their unit.

Secondly, a study where several observers assessed 60 static three-dimensional ultrasound images in the mid-coronal plane [30]. Limited information regarding methodology for this study is available as it was published as an abstract only. However, no description of the classification system used by each rater was presented. Results showed moderate inter-rater reliability for the diagnosis of uterine anomalies with a Fleiss' Kappa of κ=0.54. While this study has limitations, including using static images, obtained by a single rater manipulating the dataset, it is true to clinical practice, where many clinicians use the AFS classification without morphometric criteria, or no classification whatsoever.

The third study was the first to test intra-, alongside inter-rater reliability, where one rater, re-rates the volumes at a later date. Uterine measurements, proposed by various classification systems were taken from the three-dimensional volumes of twenty consecutive women [31]. Results showed high reliability of standardised measurements with intra-class correlation coefficients over 0.95, except for lateral uterine wall thickness [31]. While reliability of measurements has been shown to be high, this does not necessarily translate to reliable classification. Furthermore, like the first study, both raters worked in the same unit, so generalisability to other units was not demonstrated.

Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) applies a strong oscillating magnetic field, resulting in excited hydrogen atoms emitting a radio frequency signal, which is then measured by a receiving coil. The contrast between tissues is determined by the rate at which the hydrogen atoms return to their normal state [32]. MRI is a powerful imaging tool with the ability to accurately characterise and display the anatomical relationship of the female reproductive organs to surrounding structures. High-resolution images in both axial and coronal planes make it easy to diagnose and characterise uterine anomalies using MRI. Uterine measurements, such as those previously described can also be taken and it is also possible to assess tissue composition differentiating between fibrous and muscular uterine septa [33]. The reliability of MRI diagnosis of congenital uterine anomalies has not been studied, but given its increased resolution one would expect it to be similar to three-dimensional ultrasound.

Optimal diagnostic tests

An ideal diagnostic test must be accurate as well as reliable. Diagnostic accuracy relates to the potential of test to make the correct diagnosis. Accuracy can be assessed by sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), or estimated by combining these values of each imaging technique according to the Altman formula [34].

Existing diagnostic accuracy studies compare three-dimensional ultrasound with the reference standard of hysteroscopy and laparoscopy. Nonetheless, it is not possible to take uterine measurements with hysteroscopy and laparoscopy and so limits the comparison. For example, a study of 117 women with suspected anomalies on two-dimensional ultrasound, compared a three-dimensional scan with hysteroscopy in conjunction with laparoscopy. The authors found no difference in diagnosis and questioned the need for endoscopy, which is more invasive to diagnose the most common congenital uterine anomalies [35].

A similar study by Moini et al, evaluated 204 women with fertility problems who had three-dimensional ultrasound and hysteroscopy and laparoscopy. They found a sensitivity of three-dimensional ultrasound of 86.6%, and specificity was 96.9% [36].

The ESHRE have proposed that three-dimensional ultrasound is the gold standard test for diagnosing uterine anomalies based on a systematic review of literature from 1988 to 2014. It showed that the highest degrees of overall diagnostic accuracy were in decreasing order: 97.6% (95% CI, 95.6–100) for three-dimensional ultrasound, 96.5% (95% CI, 91.1–100) for HSG, and 86.6% (95% CI, 51.0–83.7) for two-dimensional ultrasound. There were no studies found reporting on the use of MRI as a screening tool. MRI was shown to be able to correctly sub-classify 85.8% of anomalies, which implies that the accuracy of identifying the presence of a malformation is well above 90%. Overall it appears that three-dimensional ultrasound may be more accurate than MRI in sub-classifying malformations [17].

This guidance seems reasonable. To accurately diagnose a uterine anomaly it is necessary to simultaneously assess the external contour of the uterus and the fundal aspect of the endometrial cavity. To classify the anomaly measurements need to be taken of any internal or external fundal indentation. Of all the diagnostic tests, only three-dimensional ultrasound and MRI are capable of achieving this. It has been suggested that performing hysteroscopy

and laparoscopy simultaneously can be considered optimal [20]. However it is not possible to measure the fundal indentation or uterine wall thickness with endoscopic tests, therefore I believe this is also suboptimal.

Considering the diagnostic accuracy of the test alone is misleading. Diagnostic accuracy must be a combination of test and classification. Almost all previous diagnostic accuracy studies have used classification systems which containing inconsistencies and limitations as previously described.

1.2.4 Prevalence

The reported prevalence of congenital uterine anomalies, including uterine septa, varies widely between studies. Reasons for this include the population studied, diagnostic test used and application of different, or sometimes no, classification system. Therefore when discussing prevalence, and subsequently effect on reproductive outcome, it is important to consider anomalies adjacent to septa on the anomaly spectrum shown in Figure 1.3, namely arcuate and bicornuate uteri. A narrative literature review found many investigators do not clearly distinguish between bicornuate and septate uteri [37]. This is a reasonable assumption as using sub-optimal tests, not capable of viewing the uterus in the coronal plane, makes distinguishing between the two anomalies difficult. Implications are that the prevalence of uterine septa may be under-represented, or they may have been misclassified as arcuate or bicornuate uteri, or vice versa.

Early studies used a suboptimal test, laparoscopy or laparotomy with HSG to assess morphology. One retrospective study published in 1991, used this diagnostic technique on 679 women undergoing sterilisation [38]. Uterine anomalies were grouped according the classification that preceded the AFS system proposed by Buttram and Gibbons [8]. A uterus was considered septate when the HSG scan showed a septum occupying more than one third of the uterine cavity in length. Prevalence of all anomalies was 3.2% (22/679), of which 20 (2.9%) were septate. A similar retrospective study published six years later, using the same diagnostic technique studied 3181 women who were either fertile, infertile or sterile [39], although the authors did not specify the classification used. The overall prevalence of uterine anomalies was similar (4.0%). Nevertheless, prevalence was higher in women who were infertile (6.3%) compared with those who were fertile (3.8%) or sterile (2.4%).

The most common anomalies were septate (33.6%) and arcuate (32.8%) uteri. Variation in prevalence of septa between these studies is wide, despite prevalence of total anomalies being similar. This provides evidence to support the view that classification by subtype is poorly reproduced. Nonetheless, both studies are limited by the suboptimal diagnostic test used, as it is not possible to perform laparoscopy and HSG concurrently.

Recently a study recruited 622 women from the general population and performed saline contrast sonohysterography, a two-dimensional technique, before classification using the AFS system. The overall prevalence of uterine anomalies was 9.8% (61/622), 42 had arcuate uteri (6.8%), 17 partial septate (2.7%) and 1 a complete septate uterus (0.16%) [40].

Studies of women with a history of recurrent miscarriage also report wide variation. A prospective study of 500 consecutive women presenting with a history of recurrent miscarriages who had a pelvic ultrasound identified 9 anomalies (1.8%), of which six were subseptate (1.2%) and three bicornuate (0.6%) [41]. The authors did not state which classification system was used. This is a much lower prevalence than previous studies discussed. Limitations include the use of, presumably, two-dimensional ultrasound, as the authors did not state this in their methods, which has been demonstrated to be an inferior test. In comparison another study of 500 women, requesting contraception (n=241), or with a history of recurrent miscarriage or subfertility (n=259) also used two-dimensional ultrasound and the AFS classification and found much higher prevalence. Researchers identified 76 anomalies (15.2%), of which 33 (6.6%) were arcuate, 6 (1.2%) septate, 7 (1.4%) bicornuate and 27 others [42]. A Japanese study of 1000 women who had miscarried and underwent HSG again found a higher number of 147 congenital uterine anomalies (14.7%), of which a higher number, fifty (5.0%), were septate [43]. Again the classification was not specified. Finally, a study of 60 women with recurrent miscarriage used hysteroscopy, but did not specifying which classification was used, found anomalies in 23 women (38.3%), but did not classify anomalies by subtype [44]. These studies demonstrate how even when studying a similar population, recurrent miscarriage, the prevalence of all uterine anomalies varies widely from 1.8% to 38.3%, and uterine septa from 1.2% to 5.0%.

Three-dimensional ultrasound and magnetic resonance imaging are considered the best tests, but have rarely been applied to the general population as they are invasive or resource intensive. Studies using three-dimensional ultrasound

mostly include women with infertility, undergoing assisted reproduction treatment or with a history of recurrent miscarriage.

A meta-analysis identified 94 observational studies comprising 89,861 women [20]. Surprisingly it found no difference in prevalence using suboptimal and optimal tests, but did demonstrate that uterine anomalies are more common in women with a history of miscarriage, recurrent miscarriage or infertility. No studies were identified investigating the prevalence of uterine anomalies in women with a history of preterm birth. Using optimal tests, the overall prevalence in women with infertility was 8.0% (95% CI 5.3–12.0), with history of miscarriage 13.3% (95% CI 8.9–20.0) and mixed infertility and recurrent miscarriage 24.5% (95% CI 18.3–32.8). Prevalence of uterine septa using optimal tests in women with infertility was 3.0% (95% CI 1.3–6.7), with history of miscarriage 5.3% (95% CI 1.7–16.8) and mixed infertility and recurrent miscarriage 15.4% (95% CI 12.5–19.0).

In the unit I have been based for this doctoral research, an observational study investigated the prevalence of uterine anomalies in 1385 women undergoing assisted reproduction treatment [45]. All women had three-dimensional ultrasound, and uterine morphology classified using the Salim system. Uterine anomalies were identified in 142 (13.3%) women. Arcuate uteri were the commonest anomaly in 164 (11.8%) women, and there were 7 (0.5%) septate and 5 (0.4%) subseptate uteri. This study is significant, because it studied the population where the septal resection trial pilot will take place. A rough estimate can be that approximately 1% of women with subfertility, undergoing assisted reproduction treatment will be eligible to participate in the trial. Limitations include selection bias, as it only studied women requiring assisted reproduction. Those with subfertility not requiring assisted reproduction, recurrent miscarriage or preterm birth were not studied. Furthermore, the inter and intra-rater reliability of three-dimensional ultrasound and the Salim classification system have not been validated.

In summary, the reported prevalence of uterine septa varies widely. The arcuate uterus is the most common anomaly, but its prevalence is not increased in women with adverse pregnancy history. Conversely a septate uterus is more common in women with infertility or miscarriage compared to the general population. Different populations have been studied, but the best diagnostic tests are invasive, meaning they are rarely applied to the general population. Furthermore, even when populations are well defined, such as recurrent

miscarriage, and single tests used, there remains debate over classification resulting in difficulties in combining studies for meta-analysis. Nonetheless, from the best available evidence it can be estimated that a uterine septum will be present in 1% of women with infertility [45], 5% of those with a history of miscarriage patients [20], but uncertain for women with a history of preterm birth.

1.2.5 Effect on reproductive outcome

Evidence from prevalence studies suggests uterine anomalies adversely impact reproductive outcome as they are found more frequently in these populations. This is useful, but remains indirect, more reliable conclusions can be drawn from comparative studies.

Two comparative studies are the extension of prevalence studies previously discussed [39, 45]. In 1997, Raga's retrospective study of 3181 women only compared outcome between subtypes of anomalies and did not compare these with normal uteri. A total of 43 women with a septate uterus, had 145 pregnancies of which 37 (25.5%) ended in early miscarriage, 3 (2.1%) ectopic, 9 (6.2%) late miscarriage, 21 (14.5%) preterm birth and 75 (51.7%) term birth. Forty-two women with arcuate uteri had better reproductive outcomes. From 110 pregnancies, 14 (12.7%) ended in early miscarriage, 3 (2.7%) ectopic, 2 (1.8%) late miscarriage, 5 (4.5%) preterm birth and 86 (78.3%) term birth. Twenty-six women with bicornuate uteri had similar early miscarriage outcomes to those with a uterine septum. Possible explanations include inaccurate diagnosis between the two anomalies or similar pathology for bicornuate and septate uteri. Reproductive outcomes were worst for unicornuate and uterus didelphys where term births occurred in only 8 out of 31 pregnancies (25.8%). The second study was a further part to the prevalence study carried out in Nottingham [45]. A prospective cohort study of women with infertility underwent assisted reproduction treatment. This included 364 women with a normal uterus and 76 with anomalies sub-grouped by type; 66 arcuate, 3 septate, 2 unicornuate, 4 subseptate and 1 bicornuate. Pregnancy rates were 43.3% for normal uteri, 54.5% for arcuate and 71.4% for septate and subseptate uteri. Three of five women (60%) with a septate uterus had a first trimester miscarriage compared with 5 (13.9%) with an arcuate and 20 (12.7%) with a normal uterus. The study finished before the end of pregnancy and so second trimester loss and term birth rates were not reported.

Acien compared normal uteri with anomalies finding live births in 42 of 47 (89%) versus 202 of 383 (53%) pregnancies respectively [46]. Women with a subseptate or septate uterus had live births in 41 of 65 (63.1%) pregnancies. Obstetric outcomes were also analysed including intrauterine growth restriction, premature rupture of membranes and breech delivery. Women with a uterine septum were statistically more likely to have a breech presentation 16 of 21 (76.2%) versus 4 of 40 (10%). The incidence of breech presentation in the normal group is much higher than found in most other studies, but this is not discussed by the authors do not discuss this. No differences were found in other obstetric outcomes.

Three-dimensional ultrasound to diagnose uterine anomalies has only been used in two published studies reporting reproductive outcome. Conveniently, both use the same modified AFS classification system. First, is a retrospective study of 1089 women attending gynaecology outpatients, with no history of infertility or recurrent miscarriage [24]. It included 983 women with normal uteri, 72 arcuate and 29 subseptate. Analysis of previous pregnancy histories showed similar numbers of previous pregnancies in 818 (83.2%) normal, 64 (88.9%) arcuate and 22 (75.8%) septate uteri. First trimester miscarriage was higher in women with uteri that were subseptate 24 (42%), compared with arcuate 14 (15.9%) and normal 235 (12%). However, women with arcuate uteri had higher rates of second trimester miscarriage 7 (7.9%) compared with septate 2 (3.6%) and 69 (3.5%) normal. Preterm birth was also higher in women with 11 (12.5%) arcuate and 6 (10.5%) septate, compared with 114 (6.2%) normal. In women with a subseptate uterus, the first trimester miscarriage rate appeared to decrease as the uterine septal length increases. This finding did not reach statistical significance, and intuitively one would expect the opposite to be true, that the greater the anomaly the higher the risk of miscarriage. These data would be an argument against septal resection. The authors found no correlation between septal length and second trimester miscarriage or preterm birth rates. The second study is the Nottingham study already discussed [45].

A meta-analysis of nine studies [47], included three of studies already discussed in this section [24, 45, 46]. Five of the six remaining studies were retrospective in nature [48–52]. One prospective study of 309 women undergoing hysteroscopy for abnormal uterine bleeding did not differentiate between septate and bicornuate uteri and identified only 12 of these anomalies [53]. Pregnancy rates at 24 months were similar in women with and without

uterine anomalies, but the 36-month cumulative live birth rate was significantly lower in women with a septate/bicornuate uterus 18 of 31 (58.1%), 540 of 648 (83.3%). Only the two previously described studies diagnosed anomalies with three-dimensional ultrasound and none used MRI.

Meta-analysis of the three studies reporting clinical pregnancy rate for septate and subseptate uteri revealed a reduction in these women (RR, 0.86; 95% CI, 0.77–0.96;P = 0.009; I^2 = 0%). Four studies in which septal defects were described [24, 46, 50, 51] revealed a significant increase in first trimester miscarriage (RR, 2.89; 95% CI, 2.02–4.14, P <0.001; I2 = 40%). Nevertheless, the definition of first trimester varied, from pregnancy loss before 12 weeks [24, 50] or 14 weeks [46, 51]. Meta-analysis of the four studies that considered septal defects revealed no effect on second trimester miscarriage rates (RR, 2.22; 95% CI, 0.74–6.65; P = 0.15, however there was large heterogeneity between studies I^2=50%.

In the most recent meta-analysis [54] the natural conception rate reported in three studies women with a uterine septum [24, 46, 50] was reduced (RR 0.86, 95% CI 0.77–0.96;P = 0.009; I^2 = 0%).

Putting the limitations of these studies to one side, meta-analyses show that women with a uterine septum have lower clinical pregnancy rates (RR, 0.86; 95% CI, 0.77–0.96; P = 0.009) and increased rates of first trimester miscarriage (RR, 2.89;95% CI; 2.02–4.14; P < 0.001), preterm birth (RR, 2.14; 95% CI, 1.48–3.11; P < 0.001) and fetal malpresentation (RR, 6.24; 95% CI, 4.05–9.62; P < 0.001) [47]. All of these associations meet conventional levels of significance and 95% confidence intervals do not cross 1.

1.2.6 Pathogenesis

The pathological mechanism from a uterine septum is unclear, although several theories have been proposed. Early studies showed septa consist of fibromuscular tissue, with poor underlying vasculature, which was believed to lead to poor placentation [55]. However, this theory has been disputed by studies showing, in some septa, no difference in vascularity to normal myometrium [56, 57]. Others suggest that fibrous tissue, which has a reduced sensitivity to steroid hormones, may be unsuitable for implantation [58], but not all septa are fibrous. Some consist of myometrial tissue which may lead to increased contractility of the uterus, resulting in miscarriage or preterm birth

[27, 59]. Defects in vascular endothelial growth factor in the endometrial tissue covering the septum may have a role to play in septal pathology [60]. Reduced uterine capacity, impaired hypertrophy or stretch of the myometrium have all been proposed as potential mechanisms. Nevertheless the only evidence to support this is the conclusions from a study of 509 women with recurrent miscarriage [61]. Investigators found women with major uterine anomalies, including septa, had poorer reproductive outcome in subsequent pregnancies with pregnancy loss at an earlier gestation. But there is no further published literature to support or refute this theory.

No research has considered that uterine anomalies are a spectrum of conditions, and even within one subgroup, uterine septa, no two are the same. Smaller uterine septa, or large broad based arcuate indentations are more likely to consist of muscular tissue, whereas a septum extending to the internal os is likely to taper into a fibrous band. It may be possible that multiple pathologies co-exist, or previous studies, limited by catch all classification systems, have wrongly grouped together two discrete anomalies. Nonetheless, no previous research has been conducted to explore this theory.

1.2.7 Summary

Uterine septa arise due to failure of canalisation of the uterine tube during embryological development. There remains disagreement regarding the definition of a uterine septum and the ESHRE-ESGE and ASRM have conflicting definitions. Evidence suggests uterine septa are more prevalent in women with a past history of pregnancy loss, but not infertility alone, and are associated with poor reproductive outcome. Nonetheless, this evidence is limited by studies with heterogeneous populations, the use of non-standardised classification systems and various diagnostic tests. Optimal assessment of uterine morphology must simultaneously view the fundal contour and endometrial cavity of the uterus in the coronal plane. Traditionally the gold standard has been combined hysteroscopy and laparoscopy, but this does not provide a coronal view or enable measurements to be taken. Therefore, three-dimensional ultrasound and MRI are now the recommended diagnostic tests by both the ESHRE and ASRM.

1.3 Septal resection

The rationale for septal resection is to attempt to correct the failure of canalisation by removing septal tissue to restore normal uterine anatomy.

1.3.1 Surgical technique

Traditionally surgical correction of the septate uterus was performed by transabdominal metroplasty. Two techniques have been described, both involve opening the uterus and then removing the septum by wedge resection, the Jones technique; or incision, the Tomkins technique [62, 63]. These types of operation are now rarely performed due to the high morbidity and risk of complications. Laparotomy is associated with prolonged hospital stay and increased need for pain relief. A reduction in uterine volume occurs with both techniques and there is also a risk of developing adhesions which could lead to tubal occlusion. After full thickness breach of the myometrium there is a significant risk of uterine scar rupture during a subsequent pregnancy. So women must wait longer before trying to conceive and are advised to deliver by caesarean section [64].

The first reported attempt at transcervical removal of a uterine septum was in 1884 [65]. Endoscopy had not been invented and so this was done blindly. The hysteroscopic approach has now been widely adopted over open metroplasty to reduce morbidity, shorten postoperative recovery and remove the need for caesarean section in future deliveries.

The septum is assessed using a hysteroscope with an operative channel, before it is divided. Micro-scissors, electrodiathermy or laser can be used to restore the uterus to its normal anatomy by dividing the septum and removing excess tissue. Surgery is often performed during the proliferative phase of the menstrual cycle to reduce bleeding, which can otherwise obscure the operative field.

Intra-uterine adhesions can form after surgery within the uterine cavity. Various strategies have been suggested to reduce this risk including leaving an intrauterine device, or Foley's catheter within the cavity; prophylactic antibiotics, anti-adhesion agents or oestrogen therapy with the aim of promoting rapid epithelialisation [37].

1.3.2 Safety

Operative hysteroscopy, as with all surgery, is associated with risk. Complications include bleeding, infection, cervical laceration, uterine perforation, pulmonary oedema, intrauterine adhesions and remaining septal remnant. Furthermore, subsequently when the woman hopefully falls pregnant, there is a increased risk of uterine rupture. These complications are summarised in Table 1.2. Data were derived from seven studies; one narrative review of 37 studies by Valle et al [66], a prospective study [67] and five case series [68–72].

Bleeding

The incidence of bleeding is reported to be between 0 to 1.2%. Valle et al suggested causes of bleeding could include inadequate uterine distension, or drift towards the uterine wall, resulting in exposure of blood vessels [66]. Nonetheless, this review did not compare surgical technique, as one would expect more bleeding with cold scissors compared with electro-diathermy, which has a coagulation effect. Many surgeons advocate the use of scissors to prevent ablative injury to surrounding endometrium and myometrial vessels [37]. The phase of the menstrual cycle and the nature of the septum were not discussed. A narrow avascular fibrous septum is less likely to bleed than a wide muscular septum, however, the study with the highest incidence of bleeding used a monopolar electrode in the early proliferative phase of the cycle [69]. Explanations for this may include the study design, patient population and assessment of bleeding. All 170 resections were performed by one surgeon over eight years but only 2 (1.2%) had excessive bleeding, managed using an intrauterine Foley's catheter to provide tamponade. Performing on average over twenty resections a year, is a high number, and may suggest the patient population included women with broader indentations classified by some as arcuate uteri. Indeed, the author did not state in their methodology how they classified anomalies. Finally, evidence suggests estimation bleeding is subjective and inaccurate [73]. No woman from all 43 studies required a blood transfusion, indicating that even when blood loss is described as excessive, the clinical implications are minimal.

Infection

No studies reported infection rates after surgery. Nonetheless, the RCOG state infection is a frequent risk after diagnostic hysteroscopy [74].

Cervical laceration

Three cases were reported. One a difficult cervical dilatation resulted in a laceration [69], two other cases were reported in another study [75], but neither manuscripts described the complication in detail.

Pulmonary oedema

Hysteroscopy uses pressurised fluid to distend the endometrial cavity which can be absorbed into the systemic circulation resulting in fluid overload. Prolonged surgery or ablative techniques increase this risk. Nonetheless, the risk of fluid is believed to be low as only two cases have been reported.

Intrauterine adhesions

Valle et al reported one case of post operative adhesions from 2167 procedures and claim they are not a problem when the procedure is performed systematically and completely without leaving any septal remnant [66]. Nevertheless, Roy et al found 11 (6%) of women had adhesions at repeat hysteroscopy requiring adhesiolysis. Reasons for this discrepancy in rates of adhesions may be that most clinicians do not perform repeat hysteroscopy or that the single surgeon in Roy's study has a high complication rate. As previously discussed, this surgeon performed high numbers of resections, and may include those with wider anomalies leaving an larger post operative surface area which may give rise to adhesions.

Septal remnant

Valle et al did not report rates of septal remnant, but in three remaining studies, the incidence ranges from 5 to 8% [69, 71, 72]. Further analysis of studies within Valle et al's review found a study of 72 septal resections, where on second look hysteroscopy 15 (21%) women had a normal uterus, another 40 women (55%) had a small fundal notch, defined as less than 1cm [76].

A repeat procedure was carried out in 17 (24%) women, where hysteroscopy revealed a septal remnant more than one centimetre.

Uterine perforation

Instrumentation and resection in the uterine cavity can result in uterine perforation. Rates of perforation range from 0 to 2.1%. Valle et al advise concurrent laparoscopy or ultrasound guidance to reduce this risk. However, no uterine perforation resulted in injury to any surrounding viscera.

Uterine rupture

Valle et al identified 18 case reports of uterine rupture during pregnancy or delivery follow septum resection [66]. All these reports describe a complication during the original resection, such a overzealous resection, excessive use of diathermy or uterine perforation. Indeed, eight of these cases had a recognised uterine perforation during initial surgery. Further to these, I found three additional case reports, two of which report recurrent uterine rupture after hysteroscopic septal resection [77–79]. None of these three cases report any complication during the initial resection.

Table 1.2: Complications after hysteroscopic septal resection.

Study		*Complication*						
Name	*Design and participants*	*Perforation*	*Uterine rupture*	*Bleeding*	*Cervical laceration*	*Pulmonary oedema*	*Adhesions*	*Septal remnant*
Valle[66], 2013	Review 37 studies n = 2167	17 (0.8%)	18* (0.8%)	9 (0.4%)	2 (0.1%)	1 (0.05%)	1 (0.05%)	-
Valli [67], 2004	Prospective study n = 48	1 (2.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (%)
Colacurci [68], 1998	Retrospective case series n = 923	8 (0.9%)	0 (0%)	1 (0.1%)	0 (0%)	1 (0.1%)	0 (0%)	-
Roy [69], 2011	Retrospective case series n = 170	2 (1.2%)	-	2 (1.2%)	1 (0.6%)	0 (0%)	11 (6%)	8†(5%)
Paradisi [70], 2011	Case series n = 246	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (%)
Tehraninejad [71], 2013	Retrospective case series n = 263	3 (1.1%)	-	1 (0.4%)	-	-	0 (0%)	21 (8%)
Ayas [72], 2011	Retrospective case series n = 181	3 (1.7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (2.2%)	9 (5%)

- not reported
*10 women had a uterine perforation at septal resection
†5 women <1cm, 3 women >1cm remnant

1.3.3 Efficacy

Evidence of efficacy for septal resection is limited to observational studies as there have been no randomised controlled trials [80]. Two meta-analyses of observational studies have been published [54, 66].

The first review, identified 29 studies, including 2074 women. Analyses was divided into two groups: group 1 included 19 "clean" studies after excluding studies with inconsistent definition of pregnancy rate (i.e. all post-resection pregnancies reported, not just index pregnancy) and studies with a substantial number of subjects lost to follow-up or with ongoing pregnancies; group 2 included all studies. Results showed increased pregnancy (event rate, 6.78; 95% CI, 6.25–7.28), for all studies and (event rate, 6.35; 95% CI 5.66–6.99) for "clean" studies. Live birth rates also increased across all studies (event rate, 5.35; 95% CI 4.78—5.91) and when limited to "clean" studies (event rate 5.02, 95% CI 4.34—-5.71). The authors concluded that hysteroscopic resection should be the treatment of choice for a uterine septum. I question how appropriate it is to label the 19 remaining observational studies as "clean". The main limitation of these studies are that outcomes were compared before and after resection, so that women acted as their own controls. This comparison is not fair, as the indication for resection was previous poor reproductive outcome, either unexplained subfertility or a miscarriage. These women could have miscarried by chance and still have a high probability of conceiving and having a live birth. These comparisons will always favour the intervention. The studies analysed are not correctly designed, or of high enough quality to justify the conclusion that hysteroscopic resection should be the treatment of choice for a uterine septum. The only way to effectively evaluate septal resection is by a randomised controlled trial.

Venetis et al, presented a review of the effect of uterine anomalies on reproductive outcome. Towards the end of the paper, a meta-analysis of studies which had a comparison group consisting of women with septa who did not undergo resection was presented. Five studies were included. However one of these studies included women with the more complex anomaly of a complete uterine septum with double cervix and vagina, which is quite different to a septum in isolation [81]. Therefore I will critically appraise each of the other four studies in turn [67, 82–84].

Firstly, Heinonen et al conducted a study at one unit in Finland between 1986 and 1995 [82]. The population included 64 women who underwent septal

resection, 32 open and 32 hysteroscopic. The hysteroscopic resection group included women with a history or previous miscarriage or preterm labour (n = 24), primary infertility (n = 8) or other causes of infertility e.g tubal, male factor, endometriosis, PCOS (n = 7). Three women went onto undergo IVF. Limitations include not stating reasons for offering intervention or not, and the control group population not specified.

Pang 2011 et al, divided women into two study groups: those with recurrent miscarriage, defined as two or more miscarriages, and women with no adverse pregnancy history [83]. All uterine septa were diagnosed using three-dimensional ultrasound, classified according to the AFS classification. Women were followed up for 15 months.

The third study included women with primary infertility used hysterosalpingogram and the AFS classification [84]. Women were followed up for 14 months, from diagnosis, or for those who had resection, 14 months after normal outpatient hysteroscopy. The number of those who conceived naturally or underwent assisted reproduction was not reported.

Finally, a study of women with recurrent miscarriage, with a septum on hysteroscopy also using the AFS classification [67]. This study had the longest follow up period of 36 months. Two patients were lost to follow up. Both the initial pregnancy after treatment and all pregnancies in the follow up period were reported.

As part of this literature review, I conducted a meta-analysis of these four studies for the outcome rates of pregnancy rate, miscarriage and live birth, shown in Figures 1.5, 1.6 and 1.7 respectively. Septal resection had no effect on pregnancy rates (odds ratio, 0.96; 95% CI 0.34–2.72), but had lower incidence of miscarriage (odds ratio, 0.20; 95% CI 0.09–0.47) and higher live birth rates (odds ratio, 4.36; 95% CI 2.15–8.87).

These four studies have major limitations. The diagnosis of a uterine septum is inconsistent within and between studies, the populations are heterogeneous and none of the studies explained how the decision was made for not intervening. Reasons could include women not consenting to surgery or clinicians not offering resection. Randomisation is important as it excludes allocation, confounding and ascertainment bias, of which these studies fall foul. In view of the bias in existing studies I believe it is challenging to interpret the data regarding outcomes after resection, especially as follow up periods and outcome measures are not standardised. Therefore I believe hysteroscopic

septal resection is still an unproven intervention.

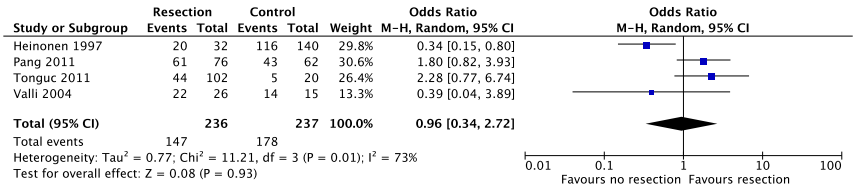


Figure 1.5: Meta-analysis of studies of hysteroscopic septal resection for the outcome of pregnancy rate.
The figure shows the odds ratio for each study as a square (with size proportional to the amount of information) and the horizontal line depicts the 95% CIs. The open diamond indicates the odds ratio (OR) with 95% CI overall. The vertical line, at the odds ratio of unity, corresponds to the line of no effect.

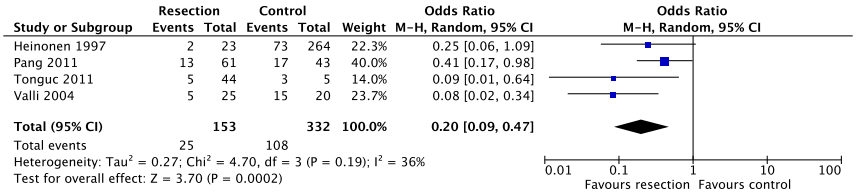


Figure 1.6: Meta-analysis of studies of hysteroscopic septal resection for the outcome of miscarriage.
The figure shows the odds ratio for each study as a square (with size proportional to the amount of information) and the horizontal line depicts the 95% CIs. The open diamond indicates the odds ratio (OR) with 95% CI overall. The vertical line, at the odds ratio of unity, corresponds to the line of no effect.

Comparing techniques

While no randomised controlled trials of septal resection versus no treatment have been completed, it is surprising that three studies comparing different surgical techniques have been conducted [75, 85, 86].

Firstly, a study including 81 women, compared the pregnancy outcome of 70 women following septal resection using micro-scissors or unipolar electro-diathermy [85]. Retrospective in design, this study was not controlled, with 17 and 53 women in each group, and 11 lost to follow up. Indications included infertility and recurrent miscarriage, but also non reproductive indications such as abnormal uterine bleeding and dysmenorrhoea. Pregnancy and live-birth outcomes were significantly better in the micro-scissors group.

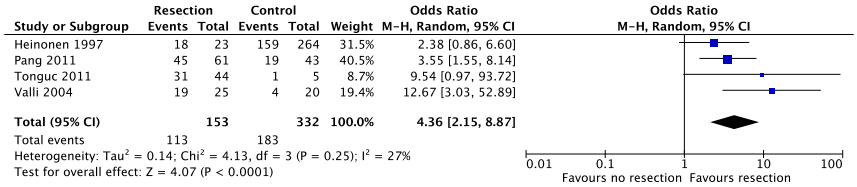


Figure 1.7: Meta-analysis of studies of hysteroscopic septal resection for the outcome of term birth.
The figure shows the odds ratio for each study as a square (with size proportional to the amount of information) and the horizontal line depicts the 95% CIs. The open diamond indicates the odds ratio (OR) with 95% CI overall. The vertical line, at the odds ratio of unity, corresponds to the line of no effect.

However the mean follow up in the micro-scissors group was longer than the resectoscope group.

The second study compared resection using monopolar knife versus a VersaPoint device [86]. A prospective randomised study was conducted including 160 women, 80 in each group, with a history of either primary infertility or recurrent miscarriage. Women were followed up for 1 year. Reproductive outcomes were no different between the two groups, but VersaPoint resection was quicker and associated with less fluid overload or cervical trauma. Thirdly, a retrospective study again comparing loop electro-diathermy with VersaPoint included 63 women [75]. Findings were similar to the previous study, showing no difference in pregnancy, miscarriage or live-birth rates between the two groups.

1.3.4 After resection

Trying to conceive after septal resection is dependent on the time it takes for the endometrium to heal. Evidence from two studies suggest this is about two months. Firstly, a study investigating the repair process randomised 19 women post-operatively to follow-up hysteroscopy at 1, 2, 4 or 8 weeks after hysteroscopic septal resection [87]. By 8 weeks the uterine cavity appeared normal and the covering endometrium was regular. Another study of 16 women who had repeat hysteroscopy until wound healing appeared complete found that at 1 month 19% had complete wound healing and all had healed by 2 months. [88].

The effect on pregnancy outcome was assessed by a retrospective cohort

study of 282 women who underwent assisted reproduction treatment after septal resection. Embryo transfer was performed at various time periods after resection, between 9 and 17 weeks. Results showed pregnancy and miscarriage rates were the same between the three groups.

Residual septum can be present after resection, either due to incomplete resection, adhesions or regeneration [89]. A prospective, observational cohort study was performed with 96 women who underwent septal resection [90]. Predictors of residual septum were pre-operative septal width and septal surface area. Women who were treated with an anti-adhesion gel barrier had less residual septum, but this was introduced part way through the study and women were not randomised to this intervention. Implications of these findings are that partially septate uteri with a broad base, perhaps categorised as arcuate under another classification, are more likely to have residual septa and adhesions. Therefore one may question septal resection in this group of women, who may have less benefit, as their septum is less pronounced, but have higher risk of complications.

1.3.5 Summary

This literature review has focused on the highest quality evidence available, namely meta-analyses summarised in Table 1.3.

Table 1.3: Summary of systematic reviews and meta-analyses of congenital uterine anomalies (CUA).

Author / Topic	*Search date*	*Population*	*Outcome*	*Number of studies*	*Number of women*	*Key findings*	*Conclusions*
Chan 2011 [20], Prevalence	2011	All CUA	Prevalence	94	89861		CUA are more common in women with poor reproductive outcome
		Canalisation defects					
		Unselected	Prevalence of septa		5163	2.3% (1.8—2.9)	When using optimal diagnostic tests uterine septa are no more prevalent in women with infertility, but are more prevalent in those with a previous miscarriage.
		Infertility	Prevalence of septa		10303	3.0% (1.3—6.7)	
		Miscarriage	Prevalence of septa		2082	5.3% (1.7—16.8)	
		Mixed infertility and recurrent miscarriage	Prevalence of septa		7053	15.4% (12.5—19)	
Chan 2011 [47], Effect on reproductive outcome	2011	All CUA		9			
		Canalisation defects	Clinical pregnancy	3	case 90 control 1035	RR 0.86 (0.77—0.96)	Uterine septa reduce fertility and increase the rate of first trimester miscarriage.
			First trimester miscarriage	4	case 271 control 2424	RR 2.89 (2.02—4.14)	
Venetis 2014 [54], Effect on reproductive outcome	2012	All CUA		25			As a group, all CUA reduce fertility and increase the rate of miscarriage.
			Conception	8	case 972 control 8690	RR 0.85 (0.73—1.00)	
			First trimester miscarriage	9	case 2284 control 72472	RR 1.56 (1.17—2.08)	
			Second trimester miscarriage	6	case 1295 control 3265	RR 2.31 (1.63—3.27)	
Efficacy of septal resection		Post septal resection	Conception	4	resection 206 no resection 102	RR 1.14 (0.79—1.65)	Septal resection reduces the risk of future miscarriage, but has no effect on rates of conception or preterm birth.
			Miscarriage	6	resection 167 no resection 124	RR 0.37 (0.25—0.55)	
			Preterm labour (<37 weeks)	6	resection 182 no resection 143	RR 0.66 (0.29—1.49)	
Valle 2013 [66], Efficacy of septal resection	2011	Post septal resection	Reproductive outcome	29*	2074		Septal resection significantly improves rates of conception and live birth.
			Conception all	29		Event rate 6.78 (6.25—7.28)	
			Conception clean	19		Event rate 5.02 (4.34—5.71)	
			Live birth all	29		Event rate 5.35 (4.78—5.91)	
			Live birth clean	19		Event rate 5.02 (4.34—5.71)	

* 19 remaining "clean" studies after excluding studies with inconsistent definition of pregnancy rate (i.e., all post-resection pregnancies reported, not just index pregnancy) and studies with a substantial number of subjects lost to follow-up or with ongoing pregnancies,

Uterine septa are no more prevalent in women with a history of infertility, but more common in women with previous miscarriage or preterm birth and its presence is detrimental to reproductive outcome. Observational studies suggest septal resection improves reproductive outcome.

The rationale for septal resection is to restore normal uterine anatomy with the intention of improving reproductive outcome. Transabdominal open metroplasty has been superseded by hysteroscopic septal resection which is less invasive and associated with fewer complications.

The efficacy of septal resection on reproductive outcomes varies between studies. All studies are observational and the majority are limited by the absence of control groups. These studies are limited by inconsistent follow up data and often do not report live birth outcomes. Only four controlled studies involving 473 women have been published and none are of high quality. Studies fail to adequately define or classify the characteristics of septa and do not clearly state how the control group was defined.

A medical intervention should be supported by the highest level of evidence before its widespread adoption by clinicians. Nonetheless, no randomised controlled trials of hysteroscopic septal resection have been conducted. The TRUST trial is ongoing in the Netherlands, but this is limited by the use of suboptimal tests to diagnose uterine septa. The National Institute for Health and Care Excellence consider septal resection an unproven intervention.

1.4 Hypotheses

The overall aim of my research is to investigate the hypothesis that a trial of septal resection for women with subfertility or miscarriage is feasible to conduct. If this hypothesis is proven, a definitive randomised controlled trial will be recommended.

In order to test my hypothesis, the following specific objectives will be addressed:

1. To assess the intra- and inter-rater reliability of three-dimensional ultrasound for diagnosis of congenital uterine anomalies, specifically uterine septa (Chapter 2).

2. To calculate the sample size required to demonstrate a clinically important difference for patients and clinicians (Chapter 3).

3. To determine how many women require screening with three-dimensional ultrasound to identify eligible participants (Chapter 4).

4. To determine the willingness of clinicians to support a septal resection trial (Chapter 5).

5. To assess the reasons why eligible women chose to participate, or not, in the pilot septal resection trial (Chapter 5).

6. To explore the potential for multi-centre recruitment to a septal resection trial (Chapter 5).

7. To calculate the value of information from a septal resection trial and compare this with a recently funded study (Chapter 6).

8. To identify whether a septal resection trial is a research priority for patients and clincians (Chapter 7).

Chapter 2

Diagnostic reliability of three-dimensional ultrasound

2.1 Introduction

Reliability refers to the error resulting from inconsistent measurement of the same attribute recorded by another rater, or recorded a second time by the same rater [91]. Fundamental to the design of a septal resection trial is the ability to reliably diagnose a uterine septum. A multi-centre trial will be reliant on multiple sonographers making consistent diagnoses, and for results to be applicable to clinical practice, diagnosis of a uterine septum must be reliable and reproducible.

Current evidence, reviewed in Chapter 1, shows wide variation in the reported prevalence of uterine anomalies. Explanations for this are the use of different diagnostic techniques with variable, and yet to be determined test accuracy, as well as non-standardised classification systems [20].

Ultrasound to assess uterine morphology is well established in clinical practice. The uterus can be easily identified and the appearance of the endometrium and myometrium are readily seen. Three-dimensional ultrasound has the advantage that it allows for reconstruction of the uterus in the coronal plane [29]. Furthermore, a three-dimensional ultrasound volume can be easily stored, allowing another operator to review the scan at a later date, or remotely. If three-dimensional ultrasound is shown to be a reliable test, using it as a single diagnostic test in the septal resection trial study protocol will improve consistency. The use of multiple diagnostic modalities is a limitation of

The Randomised Uterine Septum Transsection Trial (TRUST) ongoing in the Netherlands [92].

This chapter reports two studies assessing the reliability of three-dimensional ultrasound. Firstly, a study where five expert raters classified uterine volumes according to their own practice. The main limitation of this study was it could not differentiate between the reliability of uterine measurement, or the application of classification system. Therefore a second study was conducted to separately test the reliability of ultrasonographic measurements, and application of classification system.

To assess the intra- and inter-rater reliability of three-dimensional ultrasound for diagnosis of congenital uterine anomalies, specifically uterine septa.

2.2 Study 1: Off-line Analysis of ultraSonographic Images Study (OASIS)

2.2.1 Introduction

Various classifications have been published to categorise uterine anomalies, but many clinicians categorise uterine morphology based on subjective appearance. Reliability of diagnosis of a uterine septum is fundamental to the design of a septal resection trial, so I wanted to assess current clinical practice between raters who had not been trained in the same centre.

2.2.2 Methods

Research question

What is the inter-rater reliability of the classification of uterine anomalies, when raters use their preferred classification system?

Setting

The study was coordinated by myself using uterine volumes acquired at Nurture Fertility which were classified off-line by raters at five centres in five countries.

Design

A reliability study to assess inter-rater reliability in the ultrasound diagnosis of congenital uterine anomalies. Three-dimensional ultrasound uterine volumes were acquired prospectively as part of screening of women with recurrent miscarriage for recruitment to the pilot septal resection trial described in Section 4.2.3. Twenty three-dimensional ultrasound uterine volumes were classified by five experts.

Study population

Classification of twenty three-dimensional uterine volumes by five experts. Experts who participated in the study were, Povilas Sladkeviciu (Sweden), Caterina Exacoustos (Italy), Juan Zambrano (Brazil) , Thierry Van den Bosch (Holland) and Nick Raine-Fenning (United Kingdom). Experts were academics who have published widely on congenital uterine anomalies and use three-dimensional ultrasound in their everyday practice. Each expert had at least ten years previous experience using three-dimensional ultrasound to assess uterine morphology.

Study conduct

Twenty uterine volumes were included which I believed represented normal uteri and others with known congenital anomalies conforming to the American Fertility Society's classification system. I wanted to ensure a range of anomalies, including borderline cases, were incorporated in the study. This would not have been possible by using consecutive cases. For practical purposes only a sample size of twenty volumes was considered to be an acceptable number for raters to assess.

Three-dimensional ultrasound acquisition

Ultrasonography was performed using the same Voluson E8 scan machine and a RIC 5 – 9 MHz endocavity transducer (GE Healthcare, Austria) for all ultrasound studies within this thesis.

Scans were performed at various stages throughout the menstrual cycle for the convenience of the women involved. Before commencing the scan each woman was asked to empty her bladder to limit discomfort. Women were scanned in

a modified Lloyd-Davies position with their legs supported by stirrups to allow free manipulation of the transvaginal transducer.

Initially, a detailed two-dimensional ultrasound assessment of the pelvis was made to exclude obvious pathology and a mid-line longitudinal view of the uterus was obtained. Machine depth and zoom settings were adjusted so the longitudinal view of the uterus included the fundus and cervix filled the screen. In this view the endometrial thickness was measured as the maximum distance between the endometrial–myometrial interface of the anterior to the posterior wall of the uterus to the nearest tenth of a millimetre.

The difference between two and three-dimensional ultrasonography is the addition a third axis of data resulting in a volume as opposed to a two-dimensional area. Acquiring a three-dimensional volume of the uterus was achieved by pre-selecting a region of interest defined as the two dimensions of volume to be acquired (length and height). The region of interest was selected from the two dimensional mid-line longitudinal view of the uterus to ensure that the whole uterus and upper cervix were included. The third dimension of the volume dataset (width) was determined by the angle of acquisition, this was set to 120° which is the maximum angle of the machine.

When acquiring a three-dimensional volume, a compromise between image quality and speed of acquisition must be decided. If the subject moves during the volume acquisition the resultant image becomes disrupted by artefact. This is particularly important if the subject of interest is moving, such as blood vessels or the heart. As the uterus is a relatively immobile organ the advantage of acquiring a volume at maximum quality outweighed the potential for movement artefact as the time for acquisition was between 10 and 15 seconds. Therefore all volumes were acquired using the maximum image quality setting to achieve the highest quality volume.

Once these parameters were set, the woman was asked to remain still while the transducer automatically acquired multiple two-dimensional image planes which were then used to reconstruct the three-dimensional uterine volume.

The volume was displayed in a multiplanar view, which simultaneously displays the longitudinal, transverse and coronal planes of the acquired data in an orthogonal manner so that each image is at 90° to the two other images. A multiplanar view is demonstrated in Figure 2.1. The view was examined to ensure that the whole uterus and upper cervix had been captured. The acquisition was repeated if there was any doubt about the quality or

completeness of the volume. Once an adequate volume had been acquired it was saved to the hard drive of the ultrasound machine using a patient-specific coding system that did not include the name of the patient in order to comply with the UK Data Protection Act. Volumes were then exported to secure cloud computer storage (Dropbox Inc, San Francisco, USA). Each rater was provided with an electronic link where they could download the unprocessed volumes, a standardised assessment form and the latest edition of 4D View software (GE Healthcare Version 10.5). Independently, each examiner performed off-line analysis of the uterine volumes.



Figure 2.1: Three-dimensional multiplanar display of the uterus. The volume was acquired using transvaginal ultrasonography and displayed using 4D View.

Data collection

Each rater completed the standardised assessment form shown in Appendix 2. For each of the 20 cases, the raters were asked 8 questions: (1) is the uterus normal?; (2) what is your diagnosis?; (3) is the endometrial cavity shape normal?; (4) is the internal contour of the fundal aspect of the endometrial cavity concave or convex?; (5) can you see two tubal ostea?; (6) is the external outer contour of the uterus normal?; (7) is there a uterine septum?; and if so, (8) what is the septal length and septal width? Raters were not asked to adhere to a specific classification system to ensure they performed the analysis as they would in everyday practice. However, all choices of classification were allowed. The assessment form was returned by email.

Statistical and data analysis

For the determination of inter-rater reliability, statistical analysis was performed using Mini Tab 17 (Coventry, UK). Inter-rater reliability for diagnosis was

Table 2.1: The Kappa (κ) statistic with interpretation labels applied to corresponding ranges of kappa.

Kappa statistic (κ)	*Strength of agreement*
$<$0.00	None
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost perfect

analysed. To describe inter-rater reliability, the Fleiss-Kappa (κ) statistic was used. Fleiss-Kappa is a generic term for several similar measures of agreement used with categorical data, which reflect the classification of objects into different groups or categories. Typically it is used in assessing the degree to which two or more raters, examining the same data, agree when it comes to assigning the data into categories. In this study, the Fleiss-Kappa statistic was used to assess the extent by which the experts vary in their assessment and categorisation of uterine morphology. Complete agreement corresponds to κ=1, and lack of agreement corresponds to κ=0. A negative value of kappa would mean negative agreement, usually caused by the rater's tendency toward avoiding a grade assigned to an object by others. Recommended descriptions of numerical κ values are illustrated in Table 2.1 [93]. These values are arbitrary, but provide benchmarks for discussion.

Ethics and governance

This study was conducted as part of the study titled: "Randomised, controlled study of hysteroscopic metroplasty for women with a septate uterus and a history of miscarriage or preterm labour". This was approved by the National Research Ethics Service in October 2013 (REC reference:13/EM/0362; protocol number: 13GY007).

All patient identifiers including name, date of birth and hospital number were removed from the uterine volume datasets before they were uploaded to Dropbox.

2.2.3 Results

All five experts analysed the 20 datasets providing a diagnosis for each case. The classification systems used by the raters are shown in Figure 2.2. Two experts (B and E) used the original American Fertility Society's classification (AFS) classification. The remaining three (A, C and D) used Salim's modified AFS classification. One expert (D) provided their diagnosis using the ESHRE-ESGE classification system in addition to the Salim system. Unanimous agreement between the five raters on diagnosis was present in 7 of 20 (35%) cases shown in Table 2.2.

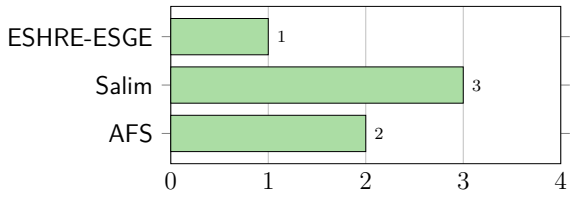


Figure 2.2: Classification systems used by expert raters.

Inter-rater agreement assessed by Fleiss-Kappa is shown in Table 2.3. The overall level of agreement between the five raters was 'moderate' as demonstrated by a kappa score of 0.59. Agreement was highest for identification of the normal uterus (κ=0.63) and common anomalies (arcuate κ=0.62; subseptate κ=0.73) all of which shows a 'substantial' level of agreement. There was greater variability when assigning the diagnosis of fusion defects (bicornuate κ=0.20; unicornuate κ=0.56). Analysis shows the level of agreement for the shape of the endometrial cavity was 'substantial' (κ=0.63). However, overall agreement on the external contour of the uterus was only 'fair' (κ=0.36).

Subgroup analysis was performed using three experts (A, C and D) who all used the Salim classification system. Unanimous agreement on diagnosis was present in 16 of 20 (80%) cases. Inter-rater reliability for the whole 20 cases was 'substantial' (κ=0.81). The diagnosis of normal was κ=0.78 and diagnosis of arcuate (κ=0.93) and subseptate (κ=1.00) uterus showing 'almost perfect' agreement. Nonetheless, there was still less agreement when assigning the diagnosis of fusion defects (bicornuate κ=-0.02; unicornuate κ=0.29). When comparing the two raters who used the original AFS classification, overall agreement was κ=0.21, showing only 'fair' agreement.

Addition of the ESHRE-ESGE rating from expert D, to the analysis reduced inter-rater agreement to κ=0.53: κ=0.6 for normal uteri, κ=0.46 for arcuate,

Table 2.2: Diagnosis of uterine volumes for each rater. Rows highlighted in bold font show unanimous agreement between raters.

Case	*Rater and classification*				
	A *Salim*	*B* *AFS*	*C* *Salim*	*D* *Salim*	*E* *AFS*
1	**Normal**	**Normal**	**Normal**	**Normal**	**Normal**
2	Arcuate	Arcuate	Normal	Arcuate	Arcuate
3	Subseptate	Subseptate	Bicornuate	Subseptate	Septate
4	**Normal**	**Normal**	**Normal**	**Normal**	**Normal**
5	Arcuate	Arcuate	Normal	Arcuate	Arcuate
6	Subseptate	Subseptate	Bicornuate	Subseptate	Septate
7	**Subseptate**	**Subseptate**	**Septate**	**Subseptate**	**Septate**
8	Normal	Unicornuate	Normal	Normal	Unicornuate
9	**Normal**	**Normal**	**Normal**	**Normal**	**Normal**
10	Arcuate	Arcuate	Normal	Arcuate	Arcuate
11	Arcuate	Arcuate	Arcuate	Arcuate	Subseptate
12	Unicornuate	Unicornuate	Unicornuate	Normal	Unicornuate
13	Arcuate	Arcuate	Normal	Normal	Arcuate
14	**Normal**	**Normal**	**Normal**	**Normal**	**Normal**
15	Arcuate	Arcuate	Normal	Arcuate	Arcuate
16	**Normal**	**Normal**	**Normal**	**Normal**	**Normal**
17	Arcuate	Arcuate	Normal	Arcuate	Arcuate
18	**Normal**	**Normal**	**Normal**	**Normal**	**Normal**
19	Subseptate	Subseptate	Bicornuate	Subseptate	Septate
20	Uncertain	Unicornuate	Uncertain	Bicornuate	Bicornuate

Table 2.3: Fleiss-Kappa (κ) inter-rater agreement for diagnosis of uterine morphology by classification system used.

Morphology	*Overall (n=5)*	*AFS (n=2)*	*Salim (n=3)*	*Overall inc ESHRE (n=6*)*
Normal	0.63	0.30	0.78	0.60
Arcuate	0.62	-0.21	0.93	0.46
Subseptate	0.73	0.22	1.00	0.67
Bicornuate	0.20	0.31	-0.02	0.30
Unicornuate	0.56	0.64	0.29	0.44
Overall	0.59	0.21	0.81	0.53

* One rater used both the Salim and ESHRE classification systems.

κ=0.67 for septum; κ=0.30 for bicornuate; and κ=0.4 for unicornuate.

2.2.4 Discussion

To my knowledge this is the first study to assess the inter-rater reliability of three-dimensional ultrasound in the diagnosis of congenital uterine anomalies using off-line analysis of uterine volumes by multiple experts across the world.

Reliability in diagnosis between expert raters using multiple classification systems can be described as moderate (κ=0.53), using the recommended descriptions for interpretation of Kappa values. Inter-rater reliability is fair (κ=0.21) between expert raters when using the AFS classification system, but this increases to almost perfect with the addition of morphometric criteria suggested by Salim.

With the aim of replicating clinical practice I purposefully did not specify which classification system the experts should use. The results of this study show that international experts do not always agree on the diagnosis of uterine anomalies and that disagreement is greater when different classifications are used. When experts do use the same system, it is unsurprising that the inter-rater reliability is higher, especially in differentiating between, normal, arcuate and subseptate uteri as limits vary between different systems.

The main limitation of this study is that it was not designed to assess the steps involved in making the diagnosis. These are: (1) acquisition of the volume; (2) volume manipulation to identify the mid-coronal plane; (3) identifying and measuring the correct uterine dimensions; and (4) applying a standardised

classification system. Reliability could be affected by any one, or all of these steps.

The sample size was based on practical purposes rather than using a power calculation to determine the number of cases. Cases used for the study included those with complex anomalies and some volumes of poor image quality. The intention was to assess whether it was possible to classify difficult cases using three-dimensional ultrasound. However, this may have inadvertently reduced reliability compared with using consecutive cases. It would have been preferable to have had each rater perform each scan independently, however due to the international nature of this study, it would not have been practical.

Fleiss-Kappa is the appropriate statistical test to use for categorical data, but there are inherent problems with its use and the recommended arbitrary descriptions of numerical κ values may lead to an under or over estimation of reliability. In practice when ratings are compared between clinicians, agreement will usually be better than that expected by chance. It is unlikely a test will have a zero value for kappa showing no agreement, so these descriptors are not very meaningful. Interpretation of kappa values should be based on what would be acceptable, or unacceptable, in clinical practice. This is dependent on clinical context. In the context of diagnosing a uterine septum for a clinical trial, it can be argued that the minimal acceptable level of agreement should be set very high, for example, greater than 0.80. This will ensure that septal resection, with its associated risks, is only performed on true septa.

This study demonstrates the effect of different classification systems on the inter-rater reliability and highlights the importance of consensus between clinicians when agreeing a classification system. Only three experts used the same classification system–the Salim system. The other two using the original AFS system. One expert used both the Salim and the newer ESHRE-ESGE classification. This raises questions about the validity of all previous studies investigating prevalence and the association with reproductive outcome as reliability of diagnosis is unlikely to be high within and between studies, especially when using different systems, or those without morphometric criteria. These implications are far reaching, suggesting that wrong conclusions may have been drawn regarding the effect of uterine anomalies on reproductive outcome, moreover the effectiveness of septal resection.

A previous study using static three-dimensional ultrasound images, found moderate inter-rater reliability for the diagnosis of uterine anomalies with a

Fleiss' kappa of κ=0.54 [30]. This is similar to the results of my study, suggesting that volume manipulation does not increase reliability. However, both studies are limited by not restricting raters to one classification system. A more recent study assessed the intra- and inter-rater reliability of uterine measurements from three-dimensional volumes [31]. Evidence suggested high reliability (intra- and inter-class correlation coefficients $\geq$0.95) of standardised measurements except for lateral uterine wall thickness. Combined, these studies imply that measuring uterine dimensions is reliable between raters, but reliability is lost when classification systems are applied. To address this question I conducted a further study to analyse these two steps separately.

2.3 Study 2: Intra- and inter-rater reliability of uterine measurements with three-dimensional ultrasound

2.3.1 Introduction

Study 1, the off-line analysis of ultrasonographic images study, found a moderate inter-rater reliability (κ=0.53) between experts when using three-dimensional ultrasound to diagnose uterine anomalies. The main limitation was that it did not analyse the steps of measuring uterine dimensions and applying the classification separately. Previous studies suggested high reliability when making uterine measurements [31]. So I hypothesised that reliability is reduced by the subjective application of various classification systems.

2.3.2 Methods

Research question

What is the intra- and inter-rater reliability of taking uterine measurements and applying different classification systems?

Setting

This study was performed at Nurture Fertility.

Design

A reliability study to assess intra- and inter-rater reliability of taking uterine measurements and applying different classification systems. Three-dimensional ultrasound uterine volumes were acquired prospectively as part of screening of women with recurrent miscarriage for recruitment to the pilot septal resection trial described in Section 4.2.3.

Study population

Uterine measurements taken by two raters from women with a history of recurrent miscarriage.

Study conduct

For the assessment of reliability, analysis was performed by two raters on volumes acquired from consecutive women screened. To reduce bias, uterine volumes were saved immediately after acquisition, before either rater manipulated the dataset to enhance the view independently. The acquisition technique was previously described in Section 2.2.2.

Both raters were experienced in three-dimensional ultrasound. Rater 1, myself (MP), was a clinical research fellow, taking time out of the Obstetrics and Gynaecology speciality training program at year 6, to pursue this PhD. Rater 1 had achieved competency for the RCOG intermediate ultrasound modules in gynaecology and early pregnancy scanning. Before this study rater 1 had undertaken a course on three-dimensional pelvic ultrasonography. Rater 2, Laurentiu Craciunas, was an Academic Clinical Fellow in Obstetrics and Gynaecology and in year 3 of the Obstetrics and Gynaecology training program and had attended a course on three-dimensional pelvic ultrasonography.

Each rater independently performed off-line analysis of the uterine volumes using 4D view computer software (Version 10.5.0, GE Healthcare, USA). Volume manipulation, analysis and measurements used the standard operating procedure as follows.

Volume manipulation

Manipulations were made in the multiplanar view on either the ultrasound machine or using 4D view (Version 10.5.0, GE Healthcare, USA).

Minimal data manipulation was required due to the data acquisition technique, which meant that the longitudinal, transverse and coronal planes of the uterus were readily identified in the majority of cases. Nonetheless, the multiplanar display was adjusted using a standard operating procedure to ensure a standardised mid-coronal view. In this view the endometrium was evident in all three planes, both the fundal aspect of the uterus and upper endometrial cavity were demonstrable in the coronal plane, and the interstitial portions of

both Fallopian tubes were displayed simultaneously. The uterine morphology and shape of the endometrial cavity were then assessed in the coronal plane. The standard operating procedure is as follows:

1. Open the uterine volume in multiplanar view showing all three planes as demonstrated in Figure 2.1.
2. Focus on the longitudinal view and optimise the endometrial line using x, y, z and slice dials.
3. Focus on the transverse view and optimise the endometrial line using x, y, z and slice dials.
4. Focus on the coronal view and optimise the endometrial line using x, y, z and slice dials.
5. Apply volume contrast imaging (VCI).
6. Fill the screen with the coronal view.
7. Rotate 90° to obtain the image in Figure 2.3.
8. Measure (1) interostial distance, (2) cavity indentation, (3) angle of the indentation and (4) fundal wall thickness.
9. Return to the longitudinal view and measure (5) endometrial thickness.

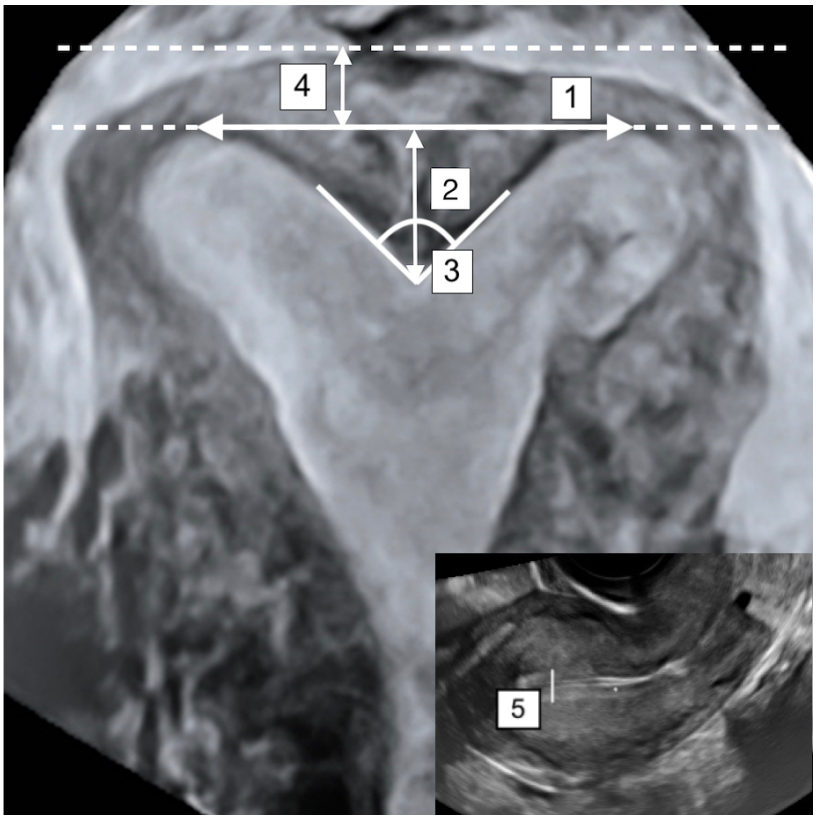


Figure 2.3: A coronal view of the uterus acquired using three-dimensional after manipulation using a standard operating procedure to obtain the following measurements required to classify uterine morphology. 1: interostial distance, 2: cavity indentation, 3: angle of the indentation, 4: fundal wall thickness, and in the longitudinal view 5: endometrial thickness.

Table 2.4: Uterine measurements.

Measurement	*Method of measurement*
1. Interostial distance	Distance between the two ostia at the transition point between the endometrial cavity and the isthmic part of the fallopian tubes
2. Cavity Indentation	Perpendicular distance between the point of maximum cavity indentation and the mid-interostial line. If the uterine cavity outline is concave, then the cavity indentation represented a septum and is measured as a positive number; if the cavity outline was convex, then the cavity indentation is measured as a negative number.
3. Angle of cavity indentation	The angle between two lines originating from the two ostia and intersecting at the point of maximum cavity indentation. When the uterine cavity outline is concave then the angle is less than 180°, whereas when it is convex, the angle is more than 180°.
4. Fundal wall thickness	Measured from the cavity indentation to the fundal wall outline in the coronal view, crossing the mid-point of the interostial line perpendicularly.
5. Endometrial thickness	Measured from echogenic border to echogenic border across the endometrial cavity on a sagittal midline plane

Reliability

Intra-rater reliability was assessed by rater 1 (MP) who assessed and measured all uterine volumes on two occasions, six weeks apart. A period of six weeks was chosen for repeat measurement to reduce the likelihood of remembering the previous measurements. Inter-rater reliability was assessed by comparing the initial measurements of rater 1 with those from rater 2. Raters were blinded to each others results.

To compare the reliability of three classifications systems, Salim, ASRM and ESHRE-ESGE, each raters measurements were analysed using algorithms, shown in Figures 2.4, 2.5 and 2.6, to categorise each uterine volume accordingly. The AFS classification was not assessed in this study as it does not provide morphometric criteria to classify uteri from measurements. The aim of this study was to specifically assess the reliability of three-dimensional ultrasound to diagnose uterine septa. The Salim classification has the additional category of arcuate uterus, which is combined with normal in the ASRM classification and does not exist in the ESHRE-ESGE classification. To account for this additional category, a further analysis classifying the cases with the dichotomous outcome of whether a septum is present of not was performed. I devised each algorithm based on the respective academic paper each classification system was presented [10, 15, 16].

Data collection

Each rater completed a standard data collection proforma. This included the date of completion and the five uterine measurements described in Table 2.4 for each uterine volume.

Statistical and data analysis

All data were entered and analysed in the Statistical Package for Social Sciences (SPSS) for Windows (Version 22, SPSS Inc, IL, USA). Intra- and inter-rater analysis were performed using the same statistical methods.

One sample t-tests demonstrated significance levels for each measurement were not statistically significant and therefore data were suitable to be analysed using the intraclass correlation coefficient and Bland-Altman charts.

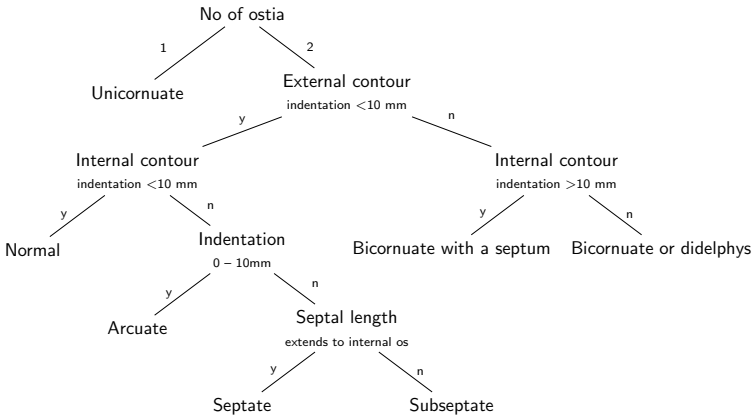


Figure 2.4: Algorithm for Salim classification of uterine anomalies.

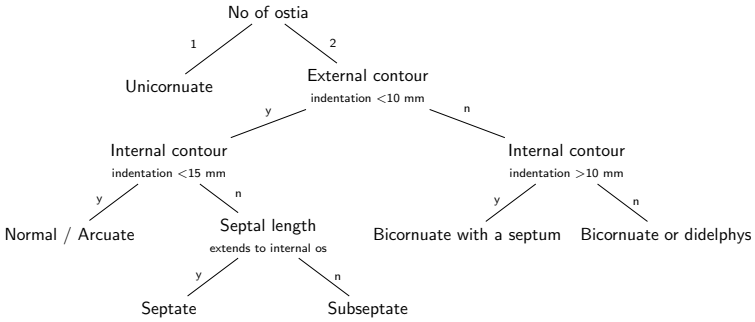


Figure 2.5: Algorithm for ASRM classification of uterine anomalies.

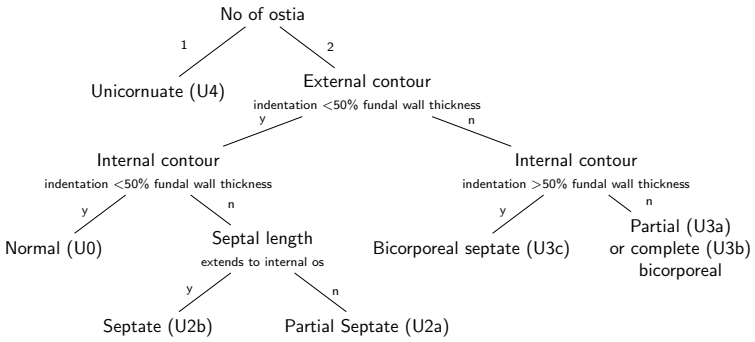


Figure 2.6: Algorithm for ESHRE-ESGE classification of uterine anomalies.

Intra-class correlation coefficient

Reliability was assessed by calculating two-way mixed intraclass correlation coefficients (ICC), with absolute agreement, and their 95% confidence intervals. Variation in measurements within a sample are of two sorts. Firstly, the variation between uterine morphology aggregated across the sample; secondly, the variation in measurements between raters. Mathematically, the intraclass correlation co-efficient is the ratio of the variance of all individual uterine volume measurements to the total variation in rater measurements. This is represented by Equation 2.3.1. As measurements between uteri show greater agreement, the numerator and denominator of the intraclass correlation coefficient will become similar in value and the intraclass correlation coefficient will approach 1. Conversely, if the agreement between raters measurements were to decrease, the denominator would increase in value and the intraclass correlation coefficient would approach 0 in value. An intra-class correlation coefficient of 1 indicates that the total variance is due solely to the variation between the uterine volumes, there being no contribution to the total variance from variation between the raters; while an intra-class correlation coefficient of 0 indicates that none of the total variance is due to variation between volumes and all the total variance being attributed to variation between raters. It is generally accepted that an intra-class correlation coefficient over 0.75 implies acceptable reliability [94]. The intra-class correlation coefficient test assumes normal distribution of the data which was confirmed if the Shapiro-Wilk significance value was $P<0.05$.

$$\frac{(\text{SD of subjects true values})^2}{(\text{SD of subjects true values})^2 + (\text{SD measurement error})^2} \tag{2.3.1}$$

Bland-Altman plot and limits of agreement

Limits of agreement were calculated and Bland-Altman charts prepared [95]. This is a scatter plot of the percentage difference between the measurements obtained by the two raters against the mean of the measurements for each uterine volume in the study. The repeatability coefficient was calculated as 2 times the standard deviation of the percentage differences between the two measurements. Limits of agreement were placed at two data intervals of the mean plus or minus the repeatability coefficient, and therefore encompassed 95% of the differences. Bland–Altman plots allow for a visual inspection of the inter-rater percentage difference in measurements versus the magnitude of the

measurements.

Assessment of bias, confidence interval of mean difference

Systematic bias between the first and second measurement was assessed by calculating the 95% confidence interval for the mean difference of measurements. When zero lay within this interval, no bias was assumed to exist.

Cohen's Kappa

Classification derived from uterine measurements using algorithms result in categorical data. Cohen's Kappa (κ) is the appropriate statistic to use for categorical ratings from two raters. Cohen's kappa is similar to the Fleiss-kappa statistic used in Study 1, but more appropriate statistic when there are no more than two raters. The recommended descriptions of numerical κ values apply equally to both Cohen's and Fleiss' Kappa and are shown in Table 2.1 [93].

Sample size calculation

A sample size calculation showed that a minimum of 46 participants were required to observe an expected intra-class correlation coefficient (ICC) of 0.90 and an acceptable ICC of at least 0.90 with a power of 80% and a significance threshold of 5% [96].

Ethics and governance

This study was conducted as part of the study titled: "Randomised, controlled study of hysteroscopic metroplasty for women with a septate uterus and a history of miscarriage or preterm labour". This was approved by the National Research Ethics Service in October 2013 (REC reference:13/EM/0362; protocol number: 13GY007).

All patient identifiers including name, date of birth and hospital number were removed from the uterine volume datasets before they were uploaded to Dropbox.

Table 2.5: Descriptive statistics of uterine measurements by each rater. Measurements in mm unless otherwise stated. *Statistics are calculated on rater 1's first measurements.

		*Rater 1**				*Rater 2*			
	N	*Mean*	*Median*	*Range*	*SD*	*Mean*	*Median*	*Range*	*SD*
Interostial distance	44	29.4	28.7	17.8 to 52.8	6.2	32.0	31.1	17.3 to 56.1	6.7
Cavity Indentation	44	4.7	1.9	-4.5 to 37.1	8.3	4.0	2.0	-3.6 to 28.6	7.2
Angle of cavity indentation (°)	44	153	164	29 to 211	39	159	166	66 to 216	33
Fundal wall thickness	45	10.4	10.6	3.8 to 20.6	2.8	11.2	11.1	5.9 to 21.6	3.2
Endometrial thickness	43	7.0	6.3	2.1 to 14.5	3.4	7.1	5.8	2.6 to 20.8	3.6

2.3.3 Results

Participants

Forty six women were included in the study and measurements were made on forty five volumes. One woman was excluded as she was seven weeks pregnant, distorting uterine anatomy. In a further case both raters independently diagnosed the same volume as a unicornuate uterus, therefore measurements of interostial distance, cavity indentation and angle of cavity indentation were not available. Finally, two women had fluid within the endometrial cavity and so endometrial thickness was not measured in these cases. Rater 1 repeated the measurements six weeks apart. The mean, median, range and standard deviation of all measurements for both raters are shown in Table 2.5.

Intra-rater reliability

Intra-rater reliability measurements are shown in Table 2.6. All measurements were normally distributed with a Shapiro-Wilk significance level less than 0.05. All measurements had an ICC higher than 0.75, considered to be good agreement. The highest ICCs were in decreasing order: cavity indentation, angle of cavity indentation, interostial distance, fundal wall thickness, and endometrial thickness. All point estimates of the ICC were greater than 0.9, except for endometrial thickness which was 0.89.

Systematic bias was present in fundal wall thickness, as the 95% confidence interval was from -1.25 to -0.01. When this interval does not cross 1, it suggests the second measurement of fundal wall thickness was systematically lower than the first. There was no systematic bias for other measurements.

Bland-Altman plots for for the intra-rater reliability for all measurements are shown in Figure 2.7.

Table 2.6: The intra-rater reliability of data measurement for uterine morphology by rater 1. Reliability is assessed by ICC using absolute agreement and 95% CIs.

	N	*Inter-class correlation coefficient (95% CI)*	*Mean percentage difference*	*Limits of agreement*	*95% confidence interval*	*Repeatability coefficient*
Interostial distance	44	0.95 (0.91 to 0.97)	2.3	-15.7 to 20.2	-3.56 to -1.72	4.5
Cavity indentation	41	0.98 (0.97 to 0.99)	-54.2	-287.6 to 179.3	-0.13 to 1.41	-106.2
Angle of cavity indentation	44	0.98 (0.96 to 0.99)	2.1	-18.5 to 22.7	-10.69 to -2.16	4.2
Fundal Wall thickness	45	0.92 (0.87 to 0.96)	3.8	-27.7 to 35.2	-1.43 to 0.00	7.4
Endometrial thickness	43	0.89 (0.79 to 0.94)	7.7	-67.1 to 82.4	-0.98 to 0.87	15.0

Inter-rater reliability

Inter-rater reliability measurements are shown in Table 2.7. All measurements were normally distributed with a Shapiro-Wilk significance level less than 0.05. All measurements had an ICC higher than 0.75, considered to be good agreement. ICCs in decreasing order are: cavity indentation, angle of cavity indentation, interostial distance, fundal wall thickness, and endometrial thickness.

Systematic bias was present in measurements of interostial distance and angle of cavity indentation where rater 2 measured systematically higher than rater 1. There was no systematic bias for other measurements.

Bland-Altman plots for the inter-rater reliability for all measurements are shown in Figure 2.8.

Reliability of classification system

Results for intra- and inter-rater reliability of uterine morphology based on each raters measurements using the Salim, ASRM and ESHRE-ESGE systems are shown in Tables 2.8 and 2.9. Both the ASRM and ESHRE-ESGE showed almost perfect agreement, however the Salim classification showed substantial agreement. When used purely to identify uterine septa, the Salim classification also demonstrated almost perfect agreement.

Table 2.7: The inter-rater reliability of data measurement for uterine morphology between two raters. Reliability is assessed by ICC using absolute agreement and 95% CIs.

	N	*Inter-class correlation coefficient (95% CI)*	*Mean percentage difference*	*Limits of agreement*	*95% confidence interval*	*Repeatability coefficient*
Interostial distance	44	0.90 (0.56 to 0.96)	8.5	-12.0 to 29.1	-1.37 to 0.15	16.8
Cavity indentation	41	0.98 (0.95 to 0.99)	-29.0	-344.7 to 286.6	-0.52 to 0.85	-56.9
Angle of cavity indentation	44	0.95 (0.90 to 0.98)	6.6	-23.8 to 37.1	-5.70 to 1.61	13.0
Fundal Wall thickness	45	0.81 (0.65 to 0.90)	6.3	-42.6 to 55.2	-0.92 to -0.01	12.3
Endometrial thickness	43	0.78 (0.60 to 0.88)	2.4	-73.1 to 77.9	-1.25 to 0.10	4.7

Table 2.8: Intra-rater reliability of classification of uterine morphology.

Classification	κ	*Level of agreement*	*SE*	*p-value*
Salim	0.622	Substantial	0.102	$<$0.01
Salim (septum/no septum)	0.877	Almost perfect	0.121	$<$0.01
ASRM	0.912	Almost perfect	0.087	$<$0.01
ESHRE-ESGE	0.889	Almost perfect	0.077	$<$0.01

Table 2.9: Inter-rater reliability of classification of uterine morphology.

Classification	κ	*Level of agreement*	*SE*	*p-value*
Salim	0.637	Substantial	0.106	$<$0.01
Salim (septum/no septum)	0.845	Almost perfect	0.151	$<$0.01
ASRM	0.812	Almost perfect	0.13	$<$0.01
ESHRE	0.883	Almost perfect	0.081	$<$0.01

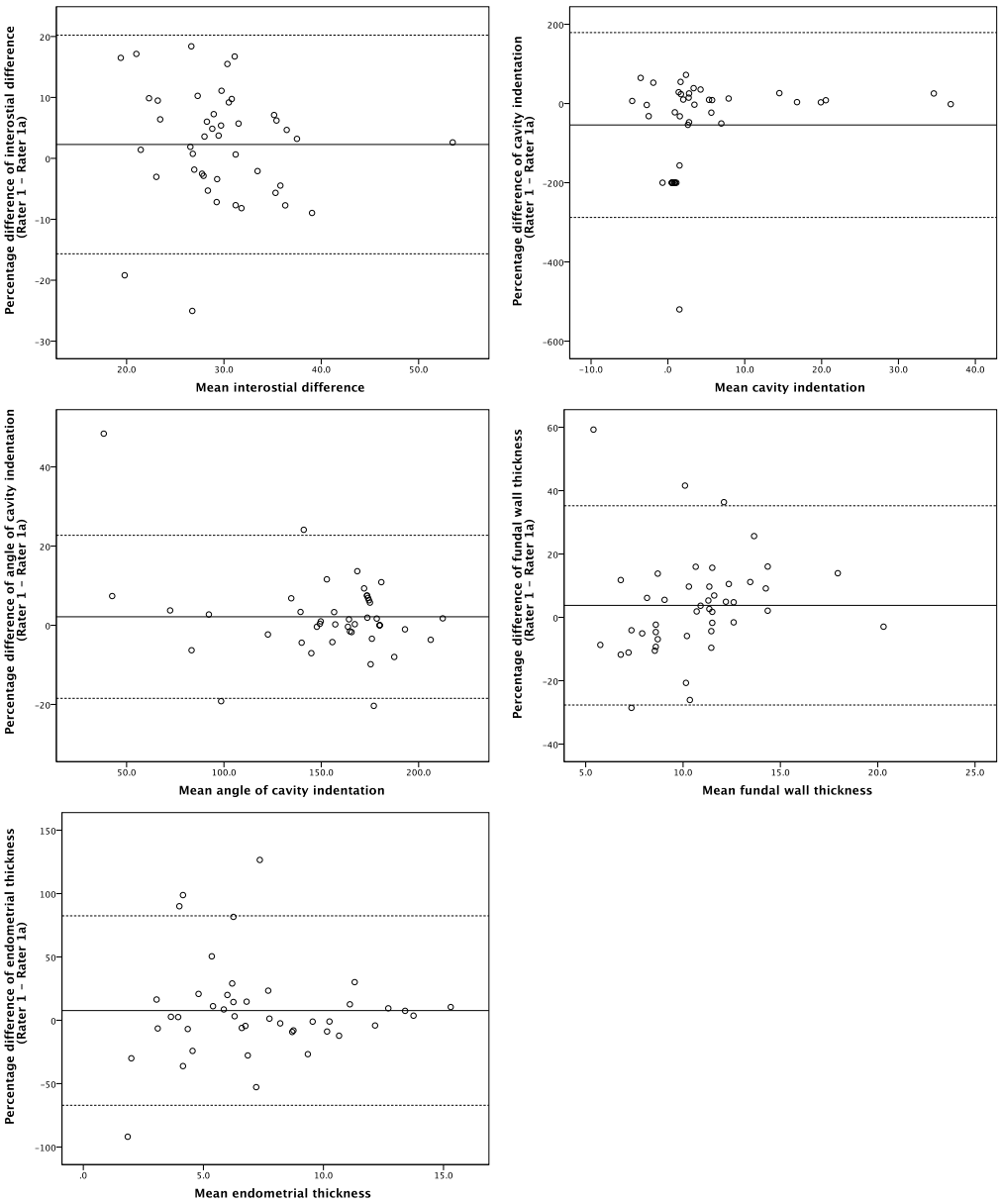


Figure 2.7: Bland–Altman plots of intra-rater reliability. The percentage difference of the measurements between the same rater six weeks apart is plotted against the mean of these measurements. Bold line shows the mean difference of the measurements. Dashed lines show limits of agreement at two standard deviations above and below the mean difference.

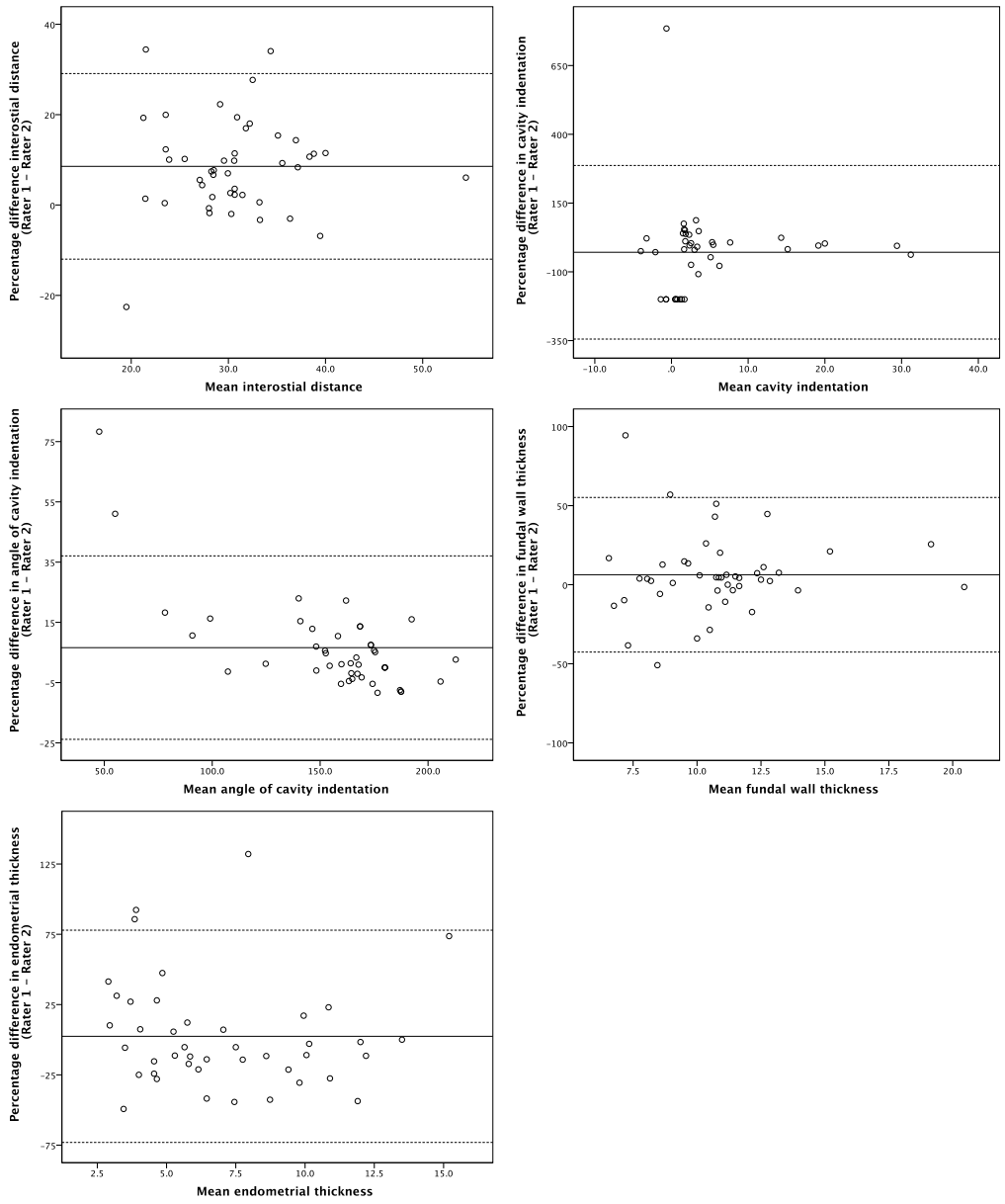


Figure 2.8: Bland–Altman plots of inter-rater reliability. The percentage difference of the measurements between two raters plotted against the mean of these measurements. Bold line shows the mean difference of the measurements. Dashed lines show limits of agreement at two standard deviations above and below the mean difference.

2.3.4 Discussion

This study has found that uterine measurements are highly reproducible. Cavity indentation shows the highest intra- and inter-rater reliability of all uterine measurements, both with intra class correlation coefficient of 0.98. Other measurements show good but decreasing reliability in the order of: angle of cavity indentation, interostial distance, fundal wall thickness, and endometrial thickness. When classification is correctly applied to uterine measurements, all three systems show high reliability. This is improved for the Salim classification when the arcuate category is removed.

These results are consistent with current literature. For example, a study of 34 women found intra class correlation coefficient for intra and inter-rater endometrial thickness measurement were 0.742 and 0.702 respectively [97]. Furthermore, reliability in this study is also similar for uterine measurements in a recent study (intra- and inter-class correlation coefficients $\geq$0.95) [31].

For the purposes of a clinical trial, it is important that cavity indentation measurements remain reliable with larger septa. Despite showing high intra and inter-rater reliability for cavity indentation (ICC 0.98 for both). The Bland-Altman plots show a greater percentage difference in values to the mean compared with other uterine measurements. This is demonstrated by the scale being larger for this measurement. This demonstrates one the limitation of thr ICC statistic where there is better agreement when measurements are smaller. So in this study, cavity indentation is on average much smaller than other distances, as demonstrated in Figure 2.3. Therefore, a small difference in measurement, say 0.1mm, becomes a high percentage difference when the average measurement is small such as a mean cavity indentation of 4mm compared with a mean interostial difference of 30mm.

The second most reliable measurement is the angle of cavity indentation. It is important to bear in mind that this measurement is a function of interostial distance and cavity indentation. This is because angle of cavity indentation is measured as the angle between the tip of the indentation and the interostial distance as shown in Figure 2.3. So it would be expected that reliability of angle of cavity indentation would lie between the two.

In practice, clinicians need to be able to distinguish between categories of uterine morphology. The ESHRE-ESGE classification categorises partial septate uteri as those with a cavity indentation greater than 50% of the

uterine wall thickness. This is less reliable than cavity indentation alone, as the reliability of cavity indentation is higher than fundal wall thickness, so combining measurement using a 50% cut off will make the diagnosis less reliable. Both the Salim and ASRM classifications use discrete cut offs of the more reliable measurement, cavity indentation, to define a uterine septum of 10mm and 15mm respectively so these classifications are more reliable.

Interestingly, the most commonly used clinical measurement, endometrial thickness was also the least reliable.

The interpretation of whether or not there is acceptable agreement between the two raters depends on subjective comparison of the limits of agreement with the range of the measurements normally encountered in clinical practice. As long as the range of the limits of agreement is considered not to be clinically important, then agreement is acceptable [95]. The majority of the measurements in this study represent normal or arcuate uteri (over 30, depending on which classification is applied). However, the measurement of most interest is cavity indentation, as this is the primary measurement used to diagnose a septum. Both intra- and inter-rater limits of agreement were approximately $\pm$5mm. While this may be significant for a single uterus, crossing the threshold between normal and septate categories, but when assessed using algorithms, the level of agreement was almost perfect.

The main limitation of this study is that the uterine volume was already acquired and so did not assess reliability and reproducibility of acquisition. The volumes were acquired by one sonographer and only one volume presented to each rater in this study. The sonographer saved the volume they deemed to be of the highest quality, so this study has not determined the reliability of acquisition itself or the reliability of measurements in the same patient, but using volumes acquired on separate occasions. However, the advantage of this study over previous reliability studies is that by using three-dimensional ultrasound, raters may manipulate the volume, rather than rely on static images. Indeed, three-dimensional ultrasound provides the basis of ultrasound simulators.

Classification should be based on prognosis and therefore it may be helpful to record measurements in a trial to see if there are any features which improve outcome. For example, resection of a large septum extending towards the internal os may be more beneficial than resecting a smaller septum of say 11mm, which only just qualifies as a septum.

2.4 Conclusion

Applying multiple classification systems, some of which may include a degree of subjectivity reduces reliability of diagnosis of uterine anomalies, including uterine septa. However, uterine measurements show high intra- and inter-rater reliability. Algorithms using uterine measurements as inputs can reliably classify uterine anomalies.

These studies, along with others raise the question of whether there is a need for a new classification system. Prospective studies, which report reliable uterine measurements, are needed to provide data to guide where to draw these limits. Although the most important specification is that clinicians need to agree to adopt it.

In conclusion, three-dimensional ultrasound can be a reliable tool for diagnosing a uterine septum, providing raters use a consistent classification system based on standardised measurements of uterine morphology. This is the case for both beginner and expert raters and between raters themselves. So I can conclude that three-dimensional ultrasound is a suitable tool to diagnose uterine septa for inclusion in a randomised controlled trial providing that acquisition of uterine volumes is also reliable.

Chapter 3

Sample size

3.1 Introduction

The sample size for a clinical trial needs to be planned carefully, with a balance between medical and statistical considerations. Funders of trials also want reassurance that they will finance research that will produce clinically meaningful results. It is not possible for researchers to investigate the effect of an intervention on the whole population. Moreover, even if it was possible, there is no identical population to act as a control group, and therefore a representative sample population is used.

By sampling a larger group, variances are reduced, as increasing the size moves closer to the ideal sample of the whole population. Larger trials are also more precise, because the more people tested, the greater certainty that any observations of difference between interventions is due to a true underlying treatment effect, not just random fluctuations in the outcome variable. Precision is also determined by the event rate, if an outcome occurs more frequently it is more likely that it can be observed in any given sample. The problem with large trials is that they are time consuming, requiring more participants and resources. Ethical concerns can also be raised. It may be possible to prove a statistical benefit using a smaller sample, causing a delay in offering a superior treatment to patients as the trial is still on-going. Conversely, a trial with fewer participants has the advantage of being quicker and cheaper to carry out, but becomes more likely to miss a clinically important difference as it is not adequately powered. It is not only doomed to failure before it has begun, but also serves to confuse the issue of determining the most effective therapy for a given condition [98].

To determine the correct sample size, a power calculation is performed. Different mathematical formulas are used depending on the type outcome variable. To simplify, in this introduction, I will discuss trials with dichotomous outcomes, where there are only two possible outcomes-such as life or death, as the primary outcome for the septal resection trial is live birth, a dichotomous outcome. Continuous outcomes are variables that can have any value between its minimum and maximum value. Sample size and power calculations can be performed by hand, through the use of tables or through the use of specialised computer software. Regardless of which tool is used, three pieces of information are required to calculate the sample size [99]: (1) the statistical chance deemed appropriate by the investigators of getting a false positive (alpha) or (2) false negative (beta) result, and (3) the minimum clinically important difference which if true would make the patient, or clinician judge the treatment as worthwhile.

Type 1 and Type 2 Error

The Type 1 error or alpha (α), is the chance of getting a false positive result. For example, if α is 0.025, it means that one in forty trials using this alpha level will demonstrate a significant result when there is actually no difference. Increasing the sample size decreases the chance of showing an effect when there is no true effect.

The Type 2 error or beta (β), is the likelihood of getting a negative result when it is actually positive. For example if β is 0.2, then one in 5 trials will show no significant result when actually there is a real difference.

Power, is the term commonly used with reference to all sample size calculations. Strictly speaking "power" refers to the number of patients required to avoid a type 2 error and is calculated as one minus beta. Sample size estimation is a more comprehensive term that looks beyond type 2 error. Nonetheless, the terms are often, but incorrectly used interchangeably.

When deciding values for alpha and beta there is no correct answer. The choice of alpha value depends on how worried you are about accepting a result as significant even though it's not. So why should the investigator settle for a value of 0.05 risking a one in twenty chance they accept a false positive? Why not be more careful and set the alpha level at 0.01 or lower? The downside of making alpha smaller is that it increases the likelihood of making a type

2 error. Reducing the likelihood of accepting a positive increases the chance of falsely accepting a negative result by default. It is therefore important to balance the two values.

It is common for investigators to use an alpha value of 0.05 and a beta value of 0.2. This suggests the research community are more cautious about type 1 versus type 2 error, being four times more willing to accept a traditional treatment by falsely missing a true difference with a new treatment, rather than incorrectly risking adopting a newer one. Nonetheless, researchers should not always accept α=0.05 and β=0.2 as default values; instead they should consider their research question and implication of making a type 1 or type 2 error. For example a researcher may be searching for new cancer treatments; the compounds developed are unstable and the tests are difficult and costly to perform. In this scenario the researcher might not mind making a type 1 errors as further research will demonstrate the drug to be ineffective, but they would want to minimise type 2 error so that a promising treatment is not falsely disregarded at an early stage of development.

The Minimum Clinically Important Difference

The third variable required to calculate sample size is delta (δ), also known as the minimum clinically important difference (MCID). Delta is the smallest clinical effect deemed important by patients and / or healthcare professionals. An intervention that makes a big clinical difference requires fewer participants to demonstrate efficacy compared with an intervention with a smaller effect.

Work using the minimum clinically important difference has almost explicitly been around clinical trials using patient reported outcomes [100–102]. However, it is infrequent for patients and healthcare professionals to ignore the side-effects, cost and risk of potential treatments. The minimum clinically important difference required for an intervention fundamentally depends on the nature of the treatment. For example, it is likely patients would find it acceptable to take daily folic acid for a reduction in the risk of neural tube defects. A tablet is easy to take and being a supplement, folic acid has no side effects. Compare this with a more invasive intervention, for instance bilateral oophorectomy to reduce the risk of ovarian cancer. Oophorectomy in a premenopausal woman is an operation with significant surgical risk resulting surgical menopause and its associated morbidity. Therefore for oophorectomy to be acceptable, a much larger clinical difference is required and so it is only

acceptable to women at high risk of ovarian cancer, such as carriers of the BRCA gene.

Alpha and beta values are predetermined, precise numbers set by the investigator; however, in theory there is only one value for the MCID and therefore only one correct sample size. Despite this, the methods used to derive the MCID can never be perfectly precise. Healthcare professionals and patients have different expectations from treatment, hypothetical questions in surveys will never be able to replicate real life decisions and humans are inherently bad at understanding risk. Therefore the sample size calculation merely provides an estimate for the number of subjects required rather than a definitive number. It can be argued that if the MCID is correctly estimated, the risk of failing to implement an effective treatment should be equal to the risk of implementing an ineffective one, so alpha (type 1) and beta (type 2) should be set equal.

Operational aspects of clinical trials

Theoretically there is only one correct sample size, but in practice constraints of finance, time, logistics and ethics enter in. It is likely that some participants will drop-out from the study either due to changing their mind, being unhappy with which group they are assigned, or lack of adherence to the study protocol. Non-compliance can also arise on the part of the investigator, especially if they are not blinded to what group the participant is assigned. Therefore after the ideal sample size has been calculated, trialists negotiate the actual number of participants required for the study in light of the these factors.

Endpoints

Many trials use primary composite endpoints to reduce the sample size. For example, a randomised controlled trial was planned in women with intrahepatic cholestasis of pregnancy, comparing ursodeoxycholic acid with placebo, and early term delivery versus expectant management [103, 104]. Intrauterine fetal death, the endpoint of most clinical significance was not used because it is a rare occurrence and would require a larger sample size. Instead the investigators used a composite outcome of fetal death or severe morbidity. In reality, only a pilot trial was performed due to lack of funding. The trial steering committee decided to prepare a new statistical analysis plan before un-blinding, a new sample size was calculated using a different primary outcome measure

of maternal itch score, calculated using the clinically meaningful difference. By using a different outcome measure the number of participants required was reduced. Subsequent to this study a larger definitive trial has been funded and is currently recruiting [105]. Surrogate markers are commonly used to make clinical trails feasible, but then make application to clinical practice more difficult. Demonstrating that ursodeoxycholic acid reduces maternal itch is useful, albeit less than the minimum clinically important difference, however it would be more clinically useful to know whether ursodeoxycholic acid, or early term delivery reduce the incidence of intrauterine fetal death.

The minimum clinically important difference in reproductive medicine

In reproductive medicine, measures of success usually focus on the statistical improvement of clinical pregnancy, miscarriage and ultimately live birth rates. However, these measures of success focus on the population level rather than the individual patient. A statistically significant improvement in live birth rate of say 5%, may not necessarily translate into a significant difference for the individual woman when weighed against the potential risks of an intervention. A recent systematic review found that of 32 randomised controlled trials the MCID was explicitly expressed in one. In 17 of 32 studies (53%; 95% CI, 35–71), the MCID was inferred from the sample size calculation, but MCIDs were not supported by citations in any study [106]. This suggests that sample size calculations used a value of Delta without considering patients or healthcare professionals views of clinical significance. In the studies discussion sections, no studies used confidence intervals or the minimum clinically important difference to draw conclusions about the results. Discussing results of clinical trials purely in terms of statistical significance or even worse, trends towards significance fails to understand the importance of translating results into clinical practice.

Summary

Calculating the correct number of participants for a trial is much more complex than a simple sample size calculation. It is important for researchers to consider all variables when determining the sample size of a study. The type of intervention, trial design, the minimum clinical important difference and pragmatic considerations must all be taken into account. Only by doing so, can an adequately sized trial be conducted with the opportunity to provide

definitive and meaningful results for patients.

3.2 Methods

3.2.1 Research questions

1. What is the minimum clinically important difference for septal resection?

2. What sample size is required to statistically demonstrate the minimum clinically important difference?

3.2.2 Setting

The study enrolled participants between February and June 2015 at Nottingham City Hospital, Queens Medical Centre and Nurture Fertility.

3.2.3 Design

A ten item survey was developed consisting of free text, binomial and multiple choice responses. A full version of the survey is included in Appendix 3. Each group of participants were asked to consider three treatment options to reduce the risk of subsequent miscarriage in the hypothetical scenario of having a uterine septum. The three treatment options were vaginal pessaries, hysteroscopic septal resection and open metroplasty. The survey was designed with reference to previously published guidelines on survey based research [107] and was based on a survey previously used to determine the minimum clinically important difference of reducing maternal itch in women with obstetric cholestasis taking ursodeoxycholic acid [104].

Content validity of the survey was assessed by senior academics and clinicians who also assessed the survey for face validity alongside two patient representatives. The feedback received was used to refine the questions and rephrase medical language. To improve validity the reviewers advised conducting the survey as part of a face to face interview which was then piloted on a patient and clinician. This provided the opportunity to further explain, rephrase the questions and ask respondents to justify their response.

I found this method provided more certainty that the respondents understood the concept of the MCID.

Given the range of different constructs measured, internal consistency calculations using principal component analysis were not undertaken. No incentives were offered for participation.

To ensure reliability of the survey intra-observer reliability was calculated using intraclass correlation coefficients (ICC). for the minimum clinically important difference, given by the same respondent on two separate occasions. The intra-class correlation coefficient test assumes normal distribution of the data which was confirmed with Shapiro-Wilk significance value $P<0.05$. A sample size of at least 19 respondents was calculated to observe an expected ICC of 0.90 and an acceptable ICC of at least 0.70 with a power of 80% and a significance threshold of 5%. Therefore, a second identical survey was sent out electronically between 4 to 6 weeks after the initial completion to the first twenty respondents. The intra-observer reliability was assessed calculating two-way mixed ICCs, with absolute agreement and their 95% CIs. The ICC was calculated as 0.96 (CI 0.91–0.99) for pessaries, 0.95 (CI 0.89–0.98) for hysteroscopic septal resection and 0.82 (CI 0.60–0.93) for open metroplasty. An ICC of greater than 0.75 was considered good agreement [94].

3.2.4 Study population

Participants were divided into three groups. Group 1 consisted of women with a history of poor reproductive outcome, defined as one or more of the following; infertility of over 12 months, a history of early, late or recurrent miscarriage or a previous preterm birth. Group 2 consisted of lay participants with no adverse pregnancy history and group 3 were healthcare professionals who care for women with adverse reproductive outcomes.

Women in group one were recruited as part of screening for the pilot septal resection trial, detailed in Section 4.2.3. Women were recruited when they attended for a three-dimensional pelvic ultrasound scan to assess uterine morphology. Group 2 were recruited using contacts from womens' health groups with no previous adverse pregnancy history. Healthcare professionals were recruited from two hospitals Nottingham City Hospital, Queens Medical Centre and Nurture Fertility.

3.2.5 Study conduct

In response to the pilot, the survey was completed by participants in the presence of myself. Participants with previous poor reproductive outcome were given the survey before either before their scan consultation. Lay participants and healthcare professionals were given the survey in their own time. There was no time limit on completion of the questionnaire or interview.

Participants were excluded from the study if the interviewer deemed that the respondent could not communicate their understanding of the concept of MCID or justify their response. For example, a respondent who could not justify stating a smaller MCID for vaginal pessaries than open metroplasty would be excluded. Conversely, if a justification was given then their response was included, for example, a preference for treatment prior to pregnancy rather than risk potential harm to the foetus of vaginal pessaries used during pregnancy.

3.2.6 Data collection

Participants completed the survey on paper and were asked to explain their answers to the interviewer. Results were transferred to an electronic database.

3.2.7 Statistical and data analysis

Microsoft Excel (Microsoft, 2010, Redmond, Washington, USA) was used to calculate descriptive statistics including socio demographic, clinical characteristics and MCID. All non parametric data was summarised by median and interquartile range. The intraclass correlation coefficients were calculated using Statistical Package for the Social Sciences (SPSS version 21.0; SPSS, Chicago, IL, USA)

Using the minimum clinically important difference identified from the study sample size calculations were performed for different values of alpha, beta and delta using standard equations [108].

$$n = f(\alpha/2, \beta) \times [p_1 \times (100 - p_1) + p_2 \times (100 - p_2)]/(p_2 - p_1)^2 \quad (3.2.1)$$

where p_1 and p_2 are the percent 'success' in the control and experimental group respectively, and

$$f(\alpha, \beta) = [\Phi^{-1}(\alpha) + \Phi^{-1}(\beta)]^2 \quad (3.2.2)$$

Φ^{-1} is the cumulative distribution function of a standardised normal deviate.

3.2.8 Ethics and governance

This study was conducted as part of the study titled: "Randomised, controlled study of hysteroscopic metroplasty for women with a septate uterus and a history of miscarriage or preterm labour". This was approved by the National Research Ethics Service in October 2013 (REC reference:13/EM/0362; protocol number: 13GY007).

Participants data were stored anonymously.

3.3 Results

All 109 subjects invited to take part in the study agreed to participate. The baseline demographic data for each group are shown in Table 3.1. All three groups were similar in age. Women with adverse pregnancy history had a tendency to have a higher level of education than lay people. Healthcare professionals were also likely to have a higher educational level than patients.

Participants excluded from the analysis are shown in Table 3.2. Reasons for exclusion were incomplete responses (n=10) or because they did not understand the concept of MCID (n=9).

Group 1 consisted of 48 women with a history of poor reproductive outcome, defined as one or more of the following; infertility of over 12 months, a history of early, late or recurrent miscarriage or a previous preterm birth. The second group contained 39 participants with no adverse pregnancy history. Group three contained 10 healthcare professionals who care for women with adverse pregnancy history.

Table 3.1: Participant demographics, before and after exclusions.

	Previous poor reproductive outcome		*Lay participants*		*Healthcare professionals*	
	all (n=48)	with excl* (n=40)	all (n=39)	with excl* (n=28)	all (n=10)	with excl* (n=10)
Age						
18 – 29	6 (13%)	3 (8%)	3 (8%)	0 (0%)	4 (40%)	4 (40%)
30 - 49	40 (83%)	35 (88%)	33 (85%)	25 (89%)	5 (50%)	5 (50%)
50 - 64	2 (4%)	2 (5%)	3 (8%)	3 (11%)	1 (10%)	1 (10%)
Highest Level of Education						
School to A-level	17 (35%)	11 (28%)	4 (10%)	18 (64%)	0 (0%)	0 (0%)
University or Higher Degree	31 (65%)	29 (73%)	35 (90%)	10 (36%)	10 (100%)	10 (100%)
Male	0 (0%)	0 (0%)	6 (15%)	5 (18%)	3 (30%)	3 (30%)
Poor Reproductive History			-	-	-	-
Subfertility	24 (50%)	19 (48%)	-	-	-	-
Subfertility considering IVF	6 (13%)	5 (13%)	-	-	-	-
Two or more unsuccessful IVF cycles	2 (4%)	1 (3%)	-	-	-	-
One early miscarriage (<12 weeks)	11 (23%)	9 (23%)	-	-	-	-
Two consecutive early miscarriages (<12 weeks)	10 (21%)	8 (20%)	-	-	-	-
Three or more consecutive early miscarriages (<12 weeks)	6 (33%)	14 (35%)	-	-	-	-
Late miscarriage (>12 weeks)	7 (15%)	5 (13%)	-	-	-	-
Preterm birth (<34 weeks)	2 (4%)	2 (5%)	-	-	-	-
None	-		39 (100%)	-	10 (100%)	-

Table 3.2: Reasons for exclusion.

	Women with previous poor reproductive outcome	*Lay participants*
Incomplete	4	6
Did not understand	4	5

3.3.1 Minimum Clinically Important Difference

The minimum clinically important difference is the smallest reduction in the baseline risk of miscarriage (30%) which a person is willing to accept to undergo the intervention. For example, a reduction in miscarriage risk from baseline to 20% would be a MCID of 10%. The MCID is presented as the median value for each group. For women with a history of poor reproductive outcome the difference was a 10% difference (IQR = 5–16) for vaginal pessaries; 20% (IQR = 10–21) for hysteroscopic septal resection; and 25% (IQR = 30-25) for open metroplasty. For lay people with no previous poor pregnancy outcome, median scores were a MCID of 15% (IQR = 5–17) for vaginal pessaries, 20% (IQR = 11–21) for hysteroscopic septal resection and 25% (IQR = 30–25) for open metroplasty. Healthcare professionals median scores were a MCID of 5% (IQR = 5–7) for vaginal pessaries, 15% (IQR = 10–15) for hysteroscopic septal resection and 26.5% (IQR = 29.75–25) for open metroplasty. These results are shown in Figure 3.1.

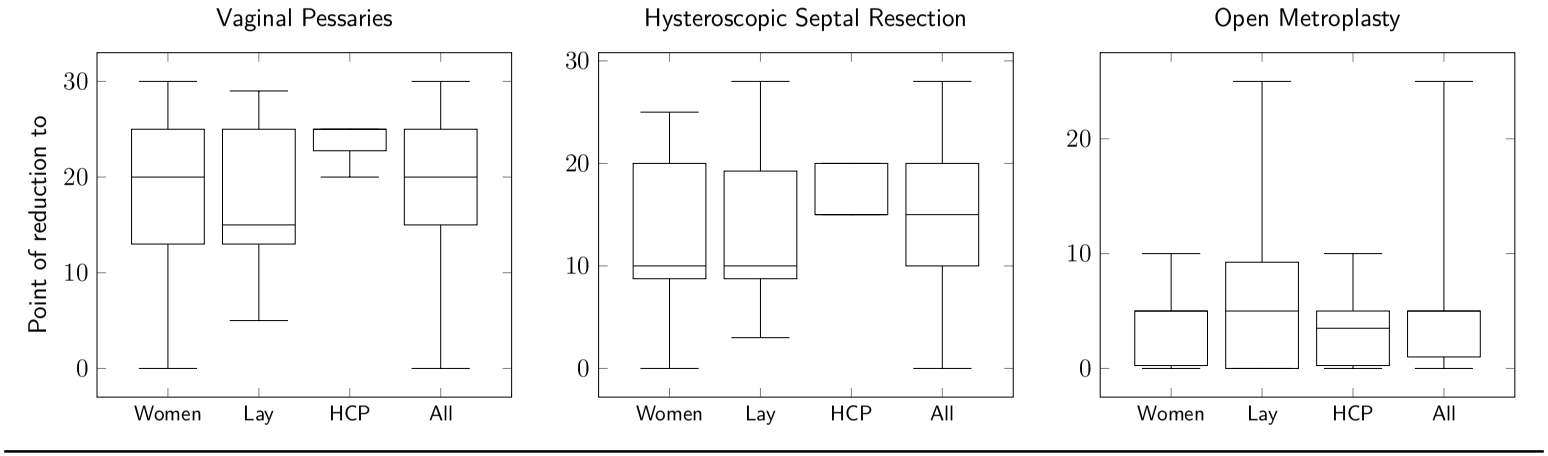


Figure 3.1: Minimum clinically important difference.
The minimum clinically important difference when participants asked how much would their risk of miscarriage need to be reduced from a baseline of 30%, for them to consider each treatment.

3.3.2 Sample size calculations

Sample sizes for a range of alpha, beta and delta are shown in Table 3.3. For septal resection the median MCID values were a reduction in the risk of miscarriage from 30% to 10% for women and lay people and 15% for healthcare professionals respectively. Using the median MCID value for the sample size calculation means the study is not powered to detect a meaningful difference for half the population. It may seem counter-intuitive, but a larger sample size is required to demonstrate a smaller difference. So sample sizes were also calculated powered to detect the smaller MCID identified by healthcare professionals, a reduction in the risk of miscarriage from 30% to 20%, this corresponds with the 75th centile for women, lay people and healthcare professionals, i.e. 75% do not consider this to be a minimum clinically important difference.

When alpha equals 0.05 and beta 0.20, values typically used by the scientific community [108] a sample size of 118 participants, 59 in each arm is required to show a minimum clinically important difference of reduction in the rate of miscarriage from 30% to 10%. Using the same values of alpha and beta, but to identify a smaller minimum clinically important difference of reduction in the rate of miscarriage from 30% to 20% a sample size of 582 participants 291 in each arm is required.

Table 3.3: Sample sizes for a trial of septal resection

Alpha	*Power (1-Beta)*	*Rate of miscarriage in control group (%)*	*Rate of miscarriage in resection group (%)*	*Sample size per group*	*Total sample size*
0.05	0.95	30	10	98	196
0.05	0.95	30	15	195	390
0.05	0.95	30	20	481	962
0.05	0.9	30	10	79	158
0.05	0.9	30	15	158	316
0.05	0.9	30	20	389	778
0.05	0.8	30	10	59	118
0.05	0.8	30	15	118	236
0.05	0.8	30	20	291	582
0.1	0.9	30	10	65	130
0.1	0.9	30	15	129	258
0.1	0.9	30	20	317	634

3.4 Discussion

3.4.1 Main findings

Patients and healthcare professionals have different views of what would be an acceptable minimum clinically important difference for three treatments to prevent miscarriage.

The sample size required to detect a minimum clinically important difference relevant to half of women and lay people, and all healthcare professionals to undergo hysteroscopic septal resection to prevent miscarriage is 118, (59 in each arm) where α=0.05 and (1-β)=0.8.

3.4.2 Strengths and limitations

To eliminate bias that occurs with retrospective MCID calculation this study was performed prospectively asking relevant stakeholders their opinion. Various tools have been used to prospectively calculate the MCID when designing clinical trials. Nonetheless, the majority of trials in reproductive medicine have at best not reported this, or worse, not even considered the MCID when obtaining a value for delta [106]. This study uses a new method for calculating

the MCID by asking patients what improvement they would consider clinically meaningful to make an intervention acceptable.

One limitation of this study is its hypothetical nature. Participants were asked to imagine they had a uterine septum. One group of participants had a history of infertility, miscarriage or preterm birth the cause for this was unknown, but was unlikely to be a uterine septum. Ideally the population studied should be women with a uterine septum, however due to the low prevalence of this combination it was impractical to limit the exercise to this group exclusively. Nonetheless, the results show similar results for women with poor reproductive history and the control group which included six men. This suggests that the MCIDs derived in the study could be applied to women with a uterine septum.

A major flaw in the design of this study is the choice of primary outcome–miscarriage. Despite discussing the importance of primary outcomes in the the introduction to this chapter, I framed the question to women as reduction in miscarriage. In hindsight, it would have been better to frame the question as what is the minimum clinically important difference in rate of live birth people would be willing to accept to the intervention. Miscarriage can be considered the converse to live birth and is not as bad a surrogate endpoint as say, improvement in uterine cavity measurements. Nonetheless, ultimately live birth is what women and health care professionals are interested in and what a trial of septal resection should be powered for.

There is no validated tool or survey to calculate the MCID. Many previous studies used anchor based methods such as a rating scale or validated patient reported outcomes surveys, such as the Medical Outcomes Study 36-Item Short Form Survey Instrument. In reproductive medicine the outcomes of pregnancy, miscarriage and live birth are dichotomous so cannot be placed on a ratings scale. Instead I used an anchor based statement of the expected miscarriage rate to be 30% and asked what improvement would make them consider the treatment to be acceptable.

The survey could only be validated for content and face validity which are subjective criteria that cannot be quantified. The intra-rater reliability was confirmed by statistical measures. Internal consistencies within the survey were addressed by simultaneously interviewing the participant and excluding them is they did not understand the concept of MCID. Nonetheless, this judgement cannot be backed up with statistical testing in the form of principal component analysis. My involvement as an interviewer must also be considered bias. When

setting out to calculate the MCID I was interested in obtaining a value for delta for a sample size calculation. My preconceived ideas could have influenced the responses the participants gave. Survey questions asking about MCID were also leading, by suggesting what would and wouldn't be acceptable. Quoted from the survey in Appendix 3, participants were asked, "For example if the drug reduced the miscarriage risk from 30% to 29% I guess you would probably regard that benefit as not worth the bother and risk of taking a pessary. However, if it reduced the average risk down to say 5%, I presume you would probably want to take the pessary". A better approach would be to provide examples unrelated to septal resection, but I did not think to do this at the start.

The face-to-face nature of this study is beneficial to determine validity of the study. However, choosing to extend this survey on-line would result in a larger sample size for this study and include views from a wider geographical area. Considering the initial research question "what is the minimum clinically important difference for septal resection?", a larger sample size would not confer additional benefit. For example, in this study I used the median MCID, but this could be challenged as incorrect because it ignores half of the participants with a smaller MCID. It could be argued that the correct MCID should be in the lower centiles to include more peoples MCIDs. So unless a larger sample size moved the MCID it is not important.

3.4.3 Interpretation

Feasibility of a trial of septal resection adequately powered to detect a minimum clinically important difference of a reduction in miscarriage from 30% to 10% requires 118 participants, 59 in each arm. Whether this sample size is achievable will depend on the number of eligible women and the recruitment rate. Evidence from prevalence studies, presented in Section 1.2.4 suggests 1% of women with infertility [45], 5% of those with a history of miscarriage [20] will have a uterine septum. Therefore it is likely a multi-centre design will be required to recruit enough participants.

Practical and logistical considerations must also be taken into account. A multi-centre trial will have greater risk of investigators not adhering to the trial protocol, some participants withdrawing from the study, or choosing to undergo resection after being randomised to the control group. Therefore most

Table 3.4: Number needed to treat (NNT) for interventions commonly used in obstetrics and gynaecology.

Population	*Intervention*	*Outcome*	*NNT*	*Reference*
Women with severe pre-eclampsia	Magnesium sulphate	Prevention of seizures	90	Duley 2010 [110]
Women with high risk of pre-eclampsia	Aspirin	Prevention of pre-eclampsia	19	Duley 2007 [111]
Women with mild to moderate endometriosis	Laparoscopic ablation or excision	Additional pregnancy	12	Duffy 2014 [112]

trials aim to recruit slightly more than the sample size calculation.

The result of the MCID for vaginal pessaries is interesting. This study suggests that the MCID for vaginal pessaries would be 5 percentage points. The PROMISE trial was a multi-centre, double-blind, placebo-controlled, randomised trial to investigate whether treatment with vaginal progesterone would increase the rates of live births over 24 weeks gestation women with unexplained recurrent miscarriage [109]. This is similar to the scenario of vaginal pessaries for a uterine septum used in this study, albeit with the converse primary outcome. The PROMISE trial was designed with 80% power to detect a minimally important absolute difference of 10 percentage points. PROMISE did not consider what would be the minimum clinically important difference for women with recurrent miscarriage when calculating the sample size. Therefore PROMISE was not adequately powered to detect the smaller minimum clinically important difference of 5% identified in this study.

In a wider context consideration to what the minimum clinically important difference means to the use of an intervention in clinical practice. An intervention with an absolute risk reduction (ARR) of miscarriage by 20% (the MCID for septal resection), can be rearranged to calculate the number needed to treat (NNT) using the following equation.

$$NNT = 1/ARR \tag{3.4.1}$$

When using an MCID derived by patients, lay people and healthcare professionals the "required" NNT is 5 women. Compared with many interventions currently used in clinical practice, an NNT of 5 is a large effect size, demonstrated in Table 3.4. Therefore it could be questioned whether it is likely that septal resection would show this effect size.

3.4.4 Conclusions

Patients expect a greater reduction in miscarriage risk than healthcare professionals. Most patients and all healthcare professionals would accept reduction in miscarriage risk from 30% to 10% after septal resection. Using conventional values for alpha and beta a sample size of 118 participants is required to detect a minimum clinically important difference. A single centre trial of septal resection is less likely to be feasible as it will take longer recruit enough participants in a reasonable time period.

Chapter 4

Screening

4.1 Introduction

Screening is a process used to identify patients eligible to enter a clinical trial. This is particularly common in phase I or II studies where potential participants undergo pre-screening, where a detailed medical history, examination and are often investigated to ensure they are eligible. Uterine septa are uncommon, even in higher risk populations, and usually an incidental finding, so screening high risk populations would be appropriate to identify women with septa eligible to participate in the study.

Feasibility of a septal resection trial depends heavily on the accrual rate of participants. Identifying participants is easier when a condition is common compared with those of a low prevalence. Existing evidence from the most recent systematic review and meta-analysis, suggests that when using optimal diagnostic tests, the prevalence of uterine septa in women with infertility is 3% (95% CI, 1.3–6.7), miscarriage 5.3% (95% CI, 1.7–16.8) or mixed infertility and recurrent miscarriage 15.4% (95% CI, 12.5–19), [20]. The wide confidence intervals, heterogeneity between studies, and use of different classification systems make estimates of the number needed to screen to identify uterine septa unreliable. For this reason the septal resection pilot trial protocol states:

> 'As this is a pilot study, we estimated the sample size based on the number of patients we could recruit over a period of one year. From our systematic review [20], approximately 5.3% of women with miscarriage have septate uteri. Considering a recruitment rate of 20% and there are approximately 1000 women

> presenting with miscarriage at Nottingham University Hospitals, Queen's Medical Centre over one year period, we will be able to recruit 10 participants in a year. At the end, we hope this study will inform us of the recruitment rate and if a multi-centred randomised trial is feasible. We cannot at this stage specify exactly what levels for each parameter would make the main trial feasible. We have intentionally not drawn hard and fast rules for going ahead or not with the main trial.'

In Chapter 2 reliability in diagnosis of uterine septa using of three-dimensional ultrasound was demonstrated. Therefore three-dimensional ultrasound can be used as a reliable diagnostic screening tool to identify eligible women.

4.2 Methods

4.2.1 Research question

How many women require screening with three-dimensional ultrasound to identify eligible participants?

4.2.2 Setting

This study was carried out in the East Midlands within Nurture Fertility and Nottingham University Hospitals NHS Trust between August 2014 and August 2017.

Nurture fertility sees and treats patients with subfertility. It receives tertiary referrals from the National Health Service and sees patients who self refer privately. During this study the unit moved from Queens Medical Centre in Nottingham to the East Midlands Fertility centre after it joined the Fertility Partnership a group of national and international clinics specialising in IVF [113].

Nottingham University Hospitals NHS Trust provide services to over 2.5 million residents of Nottingham and its surrounding communities [114]. Relevant to this thesis, the trust provides services for secondary fertility, early pregnancy assessment and a specialised recurrent miscarriage clinic at the Queens Medical Centre.

4.2.3 Design

A pilot randomised controlled trial of hysteroscopic septal resection for women with infertility, miscarriage or preterm birth. Participants were randomised to either septal resection or no treatment on a 1 to 1 allocation ratio.

Pilot septal resection trial

Prior to commencing this research, a study titled: "Randomised, controlled study of hysteroscopic metroplasty for women with a septate uterus and a history of miscarriage or preterm labour", was approved by the National Research Ethics Service in October 2013 (REC reference:13/EM/0362; protocol number: 13GY007). The study was designed and protocol developed by a previous PhD candidate who set up the trial sponsored by Nottingham University Hospitals and funded by a small grant from Nottingham University Hospitals Charity. I will describe the trial protocol and methods in brief with particular focus on my involvement in amending the design to fit this thesis.

Amendments

When taking over the lead for this trial, I reviewed the protocol with my supervisors and made some amendments. I applied to the NHS research ethics committee to also include women with primary subfertility, include other centres, and to exclude women with a body mass index over 40. The reasons for this are as follows.

First the extension of the eligibility criteria to include women with infertility came after my literature review and the National Institute for Health and Care Excellence (NICE) Interventional Procedures Programme publishing an overview of hysteroscopic metroplasty of a uterine septum for primary infertility [115]. This reported stated that the role of hysteroscopic metroplasty in patients with otherwise unexplained infertility remains uncertain, yet it remains in use. The clinical efficacy and safety can only be answered through a randomised controlled trial of hysteroscopic septal resection for women with infertility with the same design as that for miscarriage. It therefore seemed sensible to include women with infertility in this pilot study to assess feasibility for an adequately powered trial.

Second, a limit for body mass index was applied for reasons. Firstly, a pragmatic approach. Septal resection was performed as a day case at the Nottingham NHS Treatment Centre. Due to medical restrictions, patients with a body mass index over 40kg/m^2 could not undergo surgery at this centre. Second, all NHS clinical commissioning groups who fund fertility treatment place a restriction on funding for assisted reproduction for women with a BMI over 30 kg/m^2. Reasons behind this limit include using it to ration treatment, but also based on evidence showing assisted reproduction is less effective, and pregnancy complications, including miscarriage, are higher for obese women [116]. Indeed some argue it is unethical to treat obese women. All these points are contentious and a full review of the evidence is outside the scope of this thesis, so for local pragmatic reasons a limit of 40kg/m^2 was applied [117].

Third, several months into this research it became apparent that recruitment rate was lower in practice than the previously estimated target. At an early stage I recognised that feasibility of a definitive trial would be dependent on using a multi-centre approach to recruit enough participants.

The research ethics committee approved all three amendments, the approval letter is included in Appendix 1.

Trial registration

I registered the trial with the ISRCTN registry on 27 January 2015 (trial registration: ISRCTN28960271) before any women were randomised.

Screening

A prospective study design was chosen for three reasons. Firstly, retrospective identification of women from clinical records may identify more women, but would not be true to real life. Some women may have previously undergone septal resection, experienced pregnancy loss, had children or no longer have a desire to conceive. Secondly, prospective design mimics the reality of screening and recruitment for a definitive trial. Finally, alongside feasibility assessment, eligible patients can also be recruited to the pilot septum study, meaning data from these participants could be incorporated into a definitive study.

Screening used a funnel method approach to identify eligible participants. The first step was to identify clinical locations where women are seen with conditions

in the inclusion criteria, namely, infertility, miscarriage, recurrent miscarriage or preterm birth. Step 2 was to identify uterine septa on ultrasound scan, and then confirm using three-dimensional ultrasound. Step 3 involved checking women with a septum against the inclusion criteria, and step 4 the exclusion criteria.

4.2.4 Study population

The pilot septal resection trial protocol inclusion and exclusion criteria were as follows:

Inclusion criteria:

1. A uterine septum diagnosed by three-dimensional ultrasound.
2. Over 18 years old, no upper limit.
3. Wish to conceive.
4. Clinical problem: infertility; or one or more early miscarriages before 12 completed weeks; 1 late miscarriage after 12 completed weeks; or previous preterm delivery.
5. No pregnancy or endometrial surgery between diagnosis and treatment or follow up.

Exclusion criteria:

1. Other uterine anomalies apart from septum.
2. Age $<$18 years old.
3. Not planning to become pregnant.
4. Currently pregnant.
5. Body Mass Index greater than 40kg/m^2.
6. Exclusion criteria does not include women with: antithyroid antibodies, antiphospholipid antibodies, lupus anticoagulant antibodies or anticardiolipin antibodies.

The majority of these criteria were agreed before my involvement in the pilot trial. However an amendment was made to the study protocol, and agreed by research ethics, to include women with primary infertility and exclude women with a body mass index over 40kg/m^2.

Primary infertility was added to the inclusion criteria after a literature review in Chapter 1. Evidence from meta-analysis showed a higher found no difference in the rate of natural conception for women with uterine anomalies (RR 0.96, 95% 0.89 – 1.04). However in women with septate uterus, the probability of pregnancy was significantly decreased compared with normal controls (RR 0.86, 95% CI 0.77–0.96) [54]. Despite the limitations of this meta-analysis of only three studies [24, 46, 50], on balance, it was decided to include women with primary infertility in the pilot trial.

A limit on body mass index was added for logistical reasons. Hysteroscopic septal resection at Nottingham treatment centre was performed as a day case procedure. The centre had an upper limit of 40kg/m^2 for patients to be anaesthetised. Nonetheless, it must be considered that in the United Kingdom most clinical commissioning groups will not fund fertility treatment for women with a body mass index over 30kg/m^2, and many clinics choose to adopt treatment criteria including body mass index. Furthermore, rates of maternal and fetal complications during pregnancy and birth, including miscarriage, increase with body mass index, questioning the ethics of treating obese women.

Fertility

Women referred with subfertility to Nurture, routinely undergo a pre-treatment three-dimensional ultrasound scan. All women who had a pre-treatment scan between August 2014 and August 2017 were included in the study.

Miscarriage

The Early Pregnancy Unit at Queen's Medical Centre in Nottingham sees approximately 1000 new patients per year. This includes women who have viable pregnancies, but also includes women who go on to miscarry or have complications such as an ectopic pregnancy or gestational trophoblastic disease. A two-dimensional ultrasound scan is offered to each woman as part of their early pregnancy investigation. Sonographers in the unit were informed of the study and a poster was placed in each scan room asking for women to be

referred to myself if there was a suspicion of a uterine anomaly. Once referred, I performed a three-dimensional ultrasound scan to classify the anomaly.

Recurrent or late miscarriage

Women with recurrent miscarriage, defined as a history of three or more miscarriages below 12 completed weeks gestation or a second trimester miscarriage are seen in a specialised clinic. These women are not routinely investigated with pelvic imaging. Nonetheless, the recurrent miscarriage population has been identified as having a high prevalence of uterine septa. The most recent systematic review found a prevalence of 15.4% (95% CI 12.5 – 19) [20]. Therefore, each patient referred to the clinic was offered a three-dimensional pelvic ultrasound scan by myself to screen for uterine septa.

Preterm birth

No service existed in Nottingham to follow up women with a preterm birth. Septum prevalence is uncertain in this group, and local clinicians reluctant to resect. Evidence from a clinician survey in Chapter 5 showed less enthusiasm to randomise these women, therefore, I chose not to screen this group unless a specific patient was referred.

4.2.5 Study conduct

Three-dimensional transvaginal ultrasound scans were acquired, manipulated and analysed using the methods described in Chapter 2.

4.2.6 Variables, bias and study size

Nurture pre-treatment scans were performed by one of seven sonographers. Three-dimensional ultrasound scans for women with a previous miscarriage were performed and interpreted by myself.

If a septum was suspected or diagnosed by another sonographer I reviewed the uterine volumes or arranged a further scan to confirm the diagnosis. Where there was uncertainty or disagreement between myself and another sonographer, the images were reviewed by my supervisor, Mr Nick Raine-Fenning.

A three year period was chosen, as this was the duration of my role as a research fellow. Nonetheless, the study was reviewed annually by my supervisors, Mr Nick Raine-Fenning and Professor Jim Thornton to assess the recruitment rate.

4.2.7 Data collection

Patient identifiers were removed and uterine volumes were stored either on the ultrasound machine or on an external hard disk. Age, body mass index and previous pregnancy history were recorded for all women who had a uterine septum. Denominators for the number of patients referred were collected from Nurture Fertility records for fertility patients, number of referrals from the recurrent miscarriage clinic and an estimate of the throughput of patients to the Early Pregnancy Assessment Clinic.

4.2.8 Statistical and data analysis

Results are presented in absolute numbers and percentages.

4.2.9 Ethics and governance

This study was conducted as part of the study titled: "Randomised, controlled study of hysteroscopic metroplasty for women with a septate uterus and a history of miscarriage or preterm labour". This was approved by the National Research Ethics Service in October 2013 (REC reference:13/EM/0362; protocol number: 13GY007).

4.3 Results

Results are summarised in the study flow diagram, shown in Figure 4.1.

Between August 2014 and August 2017, 3040 women were referred for a three-dimensional ultrasound scan with a history of infertility (n=2846), early pregnancy problems (n=5) or recurrent miscarriage (n=189).

A uterine septum was present in 15 women shown in Table 4.1. Seven were identified through Nurture Fertility, two referred by the early pregnancy assessment unit and four from the recurrent miscarriage clinic.

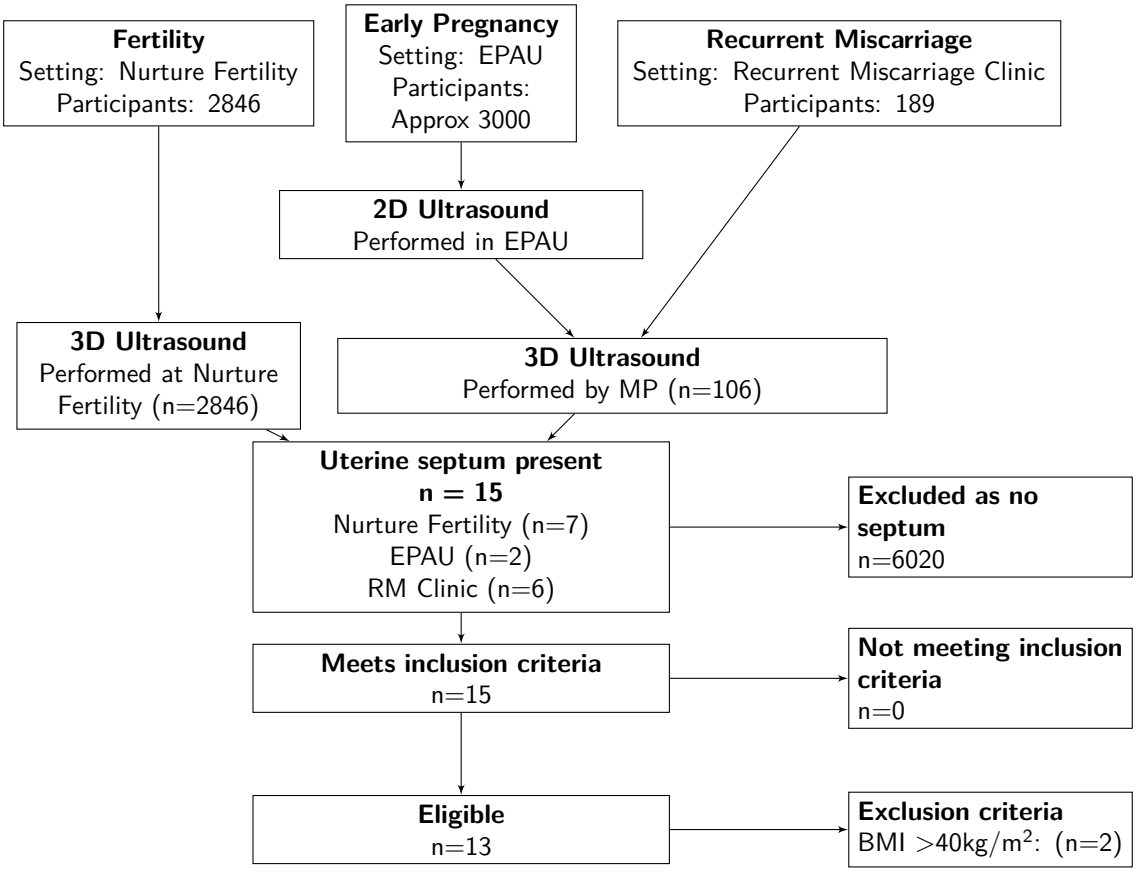


Figure 4.1: Screening Flow Diagram. EPAU: Early Pregnancy Assessment Unit. RM: Recurrent Miscarriage. BMI: Body Mass Index

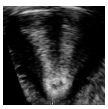
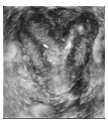
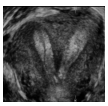
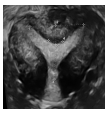
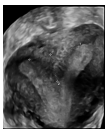
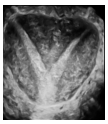
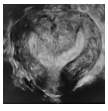
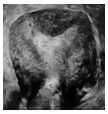
Seven women had a history of subfertility, of whom four was primary subfertility. Ten women had a history of at least one first trimester miscarriage and five had recurrent miscarriage. Two women had previous late miscarriages at 16 or 17 weeks gestation. Four women already had children, of which one was born preterm at 33 weeks gestation and one born by caesarean section for breech presentation.

Two women with a history of recurrent miscarriage were not eligible due to obesity with a body mass index over 40 kg/m^2 (53 kg/m^2 and 45 kg/m^2 respectively). Both women were advised to lose weight and to re-contact me when their BMI was below 40 kg/m^2. In addition, a letter was sent to the General Practitioner of each woman to advise on weight management. Neither women were re-referred.

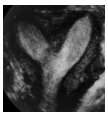
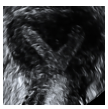
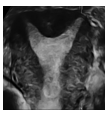
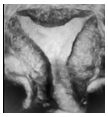
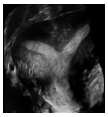
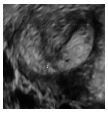
Thirteen women with a uterine septum were eligible for the trial. Demographics and uterine measurements of eligible women are shown in in Table 4.2 and Figure 4.2.

Identifying one eligible participant required approximately 464 women to be screened. Dividing this by patient population, identifying one eligible participant required screening of 407 women in subfertility clinic, 1500 in an early pregnancy assessment unit or 47 women in a recurrent miscarriage clinic.

Table 4.1: Characteristics of patients with a uterine septum.

Patient	*Age*	*BMI*	*History* (G: Gravidity P: Parity)	*Image*	*Eligible*
A	32	53	G3P0 3 first-trimester miscarriages		no*
B	44	24	G2P0 2 terminations of pregnancy		yes
C	32	23	G2P1 39 week caesarean for breech 17 week miscarriage		yes
D	32	24	G2P0 Subfertility 2 unsuccessful IVF cycles		yes
E	37	24	G3P1 39 week forceps 2 first-trimester miscarriages		yes
F	26	21	G2P0 1 first-trimester miscarriage 16 week miscarriage		yes
G	38	35	G4P0 4 first-trimester miscarriages		yes
H	40	24	G7P1 1 preterm birth at 33+6 6 first-trimester miscarriages		yes

Continued...

Patient	Age	BMI	History (G: Gravidity P: Parity)	Image	Eligible
I	31	32	G0P0 Subfertility Polycystic ovarian syndrome		yes
J	26	26	G3P1 Normal birth at 42 weeks 2 first-trimester miscarriages		yes
K	39	25	G0P0 Subfertility Low ovarian reserve		yes
L	33	24	G0P0 Subfertility unexplained		yes
M	36	26	G1P0 Subfertility 1 first-trimester miscarriage		yes
N	34	22	G0P0 Subfertility Low ovarian reserve		yes
O	35	45	G3P0 3 first-trimester miscarriages		no[†]

* † Not eligible as Body Mass Index (BMI) >40 kg/m^2

Table 4.2: Demographics and uterine measurements of eligible women (n=13).

	Mean	*Median*	*Range*
Age (years)	34.5	34	26 to 44
BMI (kg/m^2)	25.4	24	21 to 35
Interostial distance (mm)	34	34	30 to 40
Cavity Indentation (mm)	18	17	11 to 29
Angle of cavity indentation (°)	92	87	63 to 137
Fundal wall thickness (mm)	8	7	4 to 16

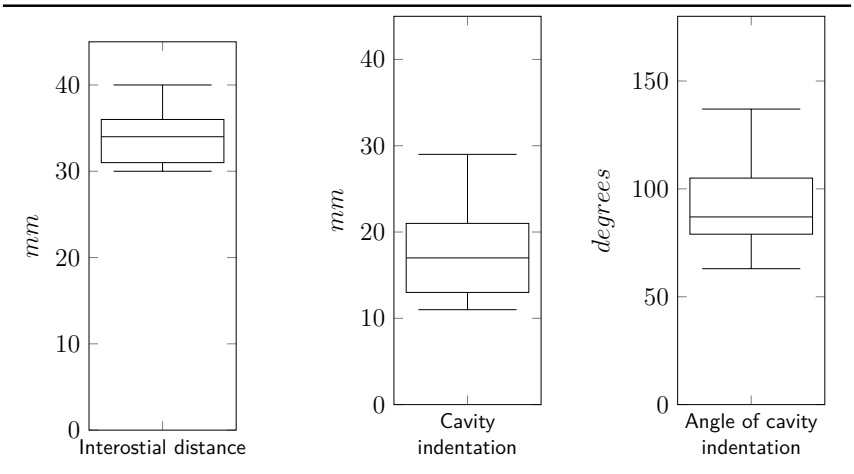


Figure 4.2: Uterine measurements of eligible septa (n=13).

4.4 Discussion

To identify one eligible participant, 464 women needed screening. Divided by patient population, identifying one eligible participant screening required 407 women with infertility, or 47 women with recurrent miscarriage to undergo three-dimensional ultrasound. Approximately one septum was identified per 1500 women who had a two-dimensional scan in the early pregnancy unit.

The main strength of this study is the prospective design. A trial can only recruit prospectively and therefore this study is representative of what may occur going forward. This study directly reflects clinical practice for women with subfertility who are routinely screened for uterine anomalies. However, women with recurrent miscarriage required an additional scan to routine practice and requiring more resources. Eighty-five women seen in the recurrent miscarriage clinic did not go onto have a three-dimensional scan. Reasons for this may include the inconvenience of attending for another appointment to be scanned, as the scan machine was at a different location; not wishing to undergo an intimate scan; not perceiving any benefit to themselves; or not wishing to participate in research. Therefore the number of women who underwent screening to identify eligible participants may be more accurate for women with subfertility compared with recurrent miscarriage. Furthermore women with infertility had their scans checked ensuring quality control of images and septal measurements as at least two sonographers scanned the patients.

Limitations include the selection bias by conducting screening in one centre. Women in Nottingham may be different to those in other locations. For example, Nottingham is known for it's previous research in three-dimensional ultrasound and uterine anomalies and may attract referrals of such women from a wider geography. Nurture fertility receives both tertiary referrals and self referrals. Secondary care is provided at Queens Medical Centre, and so, only women who are due to start assisted reproduction undergo a pre-treatment three-dimensional pelvic ultrasound scan. In March 2015, approximately half way through this study Nurture fertility moved from Queens Medical Centre, a NHS hospital, to the East Midlands Fertility Centre. This saw an increase in the number of women referred for assisted reproduction treatment. Current evidence suggests the pregnancy rate is less for women with uterine septa, so it may be expected that the prevalence would be higher in women with subfertility requiring assisted reproduction treatment. However after this change in location, there was no increase in the number of uterine septa

identified despite the increase in denominator.

The Early Pregnancy Assessment Unit (EPAU) referred two women over the three year study. Given existing prevalence literature it would be expected that more than two women out of about 3000 would have a uterine anomaly. Two reasons for the low referral rate may be sonographers not being aware of the study, or not prioritising it and distortion of uterine anatomy during early pregnancy. The study had been repeatedly promoted in the EPAU by presentation, email and posters in the scan rooms. Nonetheless, I was not based on site to be able to encourage the staff to refer patients on a daily or weekly basis. Furthermore, there was no incentive for sonographers to refer women with a suspected anomaly, and no additional support from research nurses which would come with the trial being granted NIHR CRN portfolio status. Secondly, the majority of patients who come through EPAU are experiencing bleeding or pain in early pregnancy and the main reason for ultrasonography is to risk stratify and make a diagnosis. Common diagnoses are pregnancy of unknown location, miscarriage or viable pregnancy and less frequently ectopic pregnancy, all of which are associated with anxiety for the patient. Uterine anomalies are usually incidental findings and less relevant to the diagnostic question posed in the EPAU, therefore it is likely that sonographers would not make assessment for a uterine anomaly a priority at this scan, resulting in fewer referral to myself. Furthermore, follow up to diagnose a uterine septum on three-dimensional ultrasound must be done when the woman is not pregnant, requiring arranging a further scan after the pregnancy has ended. Therefore the EPAU may not be the best setting to screen patients without additional support from the research team.

Excluding women with a previous preterm birth from screening was pragmatic, as no preterm birth clinic in Nottingham exists, but including them may have resulted in more eligible patients.

Repeating this study at another location, with additional support from the local clinical research network (CRN) is likely to have altered the numbers eligible through screening. For example, a best case scenario would be a location where women with fertility problems, previous miscarriage and preterm birth are all scanned within the same department, with access to three-dimensional ultrasound and support from a research nurse with target numbers for screening and recruitment.

The women identified by screening had a wide spectrum of past pregnancy

history. This included women with primary infertility alone, to women with a history of between one and six first-trimester miscarriages, mid-trimester losses and one woman with a preterm birth. Therefore, it is important to remember a uterine septum in itself is not the pathological condition, instead is is a spectrum of conditions where a septum is present, and resection may be a beneficial intervention for all, some or none of these conditions. This raises the issue that multiple research questions needs addressing: Is septal resection safe and effective for (1) primary infertility, (2) recurrent miscarriage, and (3) preterm birth? Ultimately this requires a trial powered to recruit enough women with each condition to identify a clinically meaningful difference.

This study found that uterine septa were most prevalent in women with recurrent miscarriage, but these women were also more likely to be ineligible due to obesity. This could be expected, because obesity is a recognised risk factor for both subfertility and miscarriage. As fertility treatment is not funded by the NHS for women with a body mass index of more than 30kg/m^2, and Nurture fertility does not treat women with a BMI over 35kg/m^2 it is likely obese women will have been excluded out before entering this study.

These results show a similar trend in prevalence of uterine septa to the existing literature. Uterine septa were more prevalent in women with recurrent miscarriage (3.8%) compared to those with subfertility (0.24%). However these values were much less than those identified by the most recent meta-analysis, of 3% (95% CI, 1.3–6.7) for women with infertility, and 5.3% (95% CI, 1.7–16.8) for those with recurrent miscarriage [20]. Moreover, for women with infertility, the prevalence in this study is below the lower 95% confidence interval of the meta-analysis. Initially, I suspected this was due to more stringent classification criteria, however when reviewing the uterine dimensions this was not the case.

The angle of cavity indentation was over 90° in five women. According to the Salim classification are arcuate uteri, rather than subseptate, and none of these "septa" had a fundal indentation over 15mm, so would considered "normal/arcuate" according to the ASRM classification. Nonetheless, in all cases the fundal indentation was over 50% of the fundal wall thickness, so are classified at partial septate (U2a) according to the ESHRE-ESGE classification.

Another reason why I did not identify many uterine septa could have been clinicians not referring women with a uterine septum to me, due to them not wanting to participate in the pilot trial. A search of the theatre logbook, did not identify any women who had undergone septal resection without being referred

to myself. It was not possible to do the converse; identifying women who had chosen conservative management who were not identified by screening, in essence being shielded from the study. Nonetheless, if this practice existed, of which there is no evidence to suggest it did, it would not affect the feasibility of the study as these patients would not have been recruited anyway.

Women with uterine septa had a median age of 34 years (range 26–44). Data were not collected for the age of women without a septum, but I would estimate that women with a septum are on average older than those without. This is to be expected as evidence shows poor reproductive outcome for women with a uterine septum, so they are more likely to be trying to conceive for longer thus more likely to be older.

Implications of this study on the feasibility of a septal resection trial are that the requirement to screen many women to identify eligible participants will be labour and resource intensive, especially when three-dimensional ultrasound is not a part of routine clinical practice. This will require substantial funding to identify eligible participants. Options could include training early pregnancy unit sonographers in three-dimensional ultrasound to screen women following miscarriage. However as these scans are performed in early pregnancy, uterine morphology will be distorted, possibly resulting in missing uterine septa.

Most women were identified from the subfertility population, as this is where three-dimensional ultrasound was used in routine clinical practice. This may be unique to Nottingham, which is known for its early adoption of three-dimensional ultrasound but may not be the case in other units, and so raises the question of the availability and routine use of three-dimensional ultrasound elsewhere in the country.

4.4.1 Conclusion

Identification of eligible participants required more women to be screened using three-dimensional ultrasound than expected based on existing evidence on the prevalence of uterine septa. The low number of women identified makes a septal resection trial less feasible, unless this can be overcome by screening more women using other trial centres.

Chapter 5

Participation of patients, clinicians and units

A well designed trial is useless if participants cannot be recruited. Many trials, particularly surgical ones, experience difficulty recruiting patients. A review of 63 surgical trials found that the main problem reported in most trials was a very slow recruitment rate, mainly due to the difficulty in finding eligible patients [118]. Chapter 4 identified eligible patients over three years, but these numbers were small. Eligible women will only be randomised if clinicians introduce them to the study, so feasibility of an adequately powered definitive trial be dependent on: (1) clinicians agreeing to randomise their patients; (2) the willingness of women to be randomised; and (3) multiple centres agreeing to participate in the trial. This chapter investigates these three issues.

5.1 Study 1: Clinician opinion of hysteroscopic septal resection

5.1.1 Introduction

Trial feasibility is dependent on support from clinicians. Firstly, it will require clinicians with access to three-dimensional ultrasound to screen potential participants. Second, clinicians will need to perform hysteroscopic septal resection in a manner which is comparable with other surgeons. However thirdly, and most importantly, feasibility will be dependent on clinicians being willing to randomise their patients. Therefore a significant number of clinicians

need to be in equipoise over the benefit of septal resection for this study to be carried out.

5.1.2 Aims

To determine the willingness of clinicians to support a septal resection trial. Specific questions to be addressed are: (1) what is the availability of three-dimensional ultrasound? (2) what frequency is septal resection performed by surgeons? (3) is there agreement in surgical technique and post-operative care? (4) what proportion of clinicians are in equipoise? and (5) what proportion of clinicians are prepared to randomise their patients?

5.1.3 Methods

Clinician Survey

I conducted a survey of clinicians to establish current UK practice of diagnosis, treatment decision and surgical technique.

No formally validated surveys of this nature were identified. Therefore a survey was designed based on previously conducted survey used in the development of randomised controlled trials of progesterone in women with miscarriage [119]. These trials, PROMISE and PRISM, were then funded by the National Institute for Heath Research Health Technology Assessment program [109, 120].

The survey covered: (1) the investigation and classification system of choice used to diagnose uterine anomalies; (2) the availability of three-dimensional ultrasound; (3) the prevalence of patients with a uterine septum; (4) the indication and number of hysteroscopic septal resections performed; (5) post procedure follow up; (6) willingness to randomise women in an randomised controlled trial; and (7) interest in becoming a trial centre.

Face validity was evaluated by two experts, my PhD supervisors, Mr Nick Raine-Fenning and Professor Jim Thornton. Pilot testing was conducted within our department.

The final 19 item survey consisted of 16 closed questions and two open questions asking clinicians to describe their surgical technique and use of any adjuncts. The final survey can be seen in Appendix 5.

An invitation and link to the on-line survey was sent via email in January 2015 to 103 clinicians. This included members of the British Fertility Society, and Association ofEarly Pregnancy Units along with those from the personal network of my supervisors and myself.

5.1.4 Results

Sixty-seven clinicians responded to the survey giving a 65% response rate.

Investigation and diagnosis

The gold standard diagnostic tests used by respondents are shown in Figure 5.1. Three-dimensional transvaginal ultrasound was viewed by over half of clinicians as the gold standard test for diagnosis of a uterine anomaly (N=37; 55%). Magnetic resonance imaging (N=10; 15%), hysteroscopy (N=19; 28%) and hysterosalpingogram (N=1; 1.5%).

The availability of three-dimensional ultrasound and septal resection is shown in Figure 5.2. Three-dimensional ultrasound was routinely used by 37 (56%) clinicians. Most units (N=57; 85%) had access to three-dimensional ultrasound and 45 (68%) clinicians performed septal resection.

Classification systems used are shown in Figure 5.3. There was no consensus in the system used to classify uterine anomalies, the most commonly used classification was the American Fertility Society's (AFS) (N=20; 30%), followed by Salim's modified AFS system (N=18; 27%), ESHRE-ESGE (N=17; 25%). Twelve clinicians did not use a classification system and no clinician used the Vagina Cervix Uterus Adnex-associated Malformation classification system (VCUAM) system.

Septal resection

In total, 46 (69%) clinicians perform hysteroscopic septal resection. Estimations of the incidence of uterine septa seen by clinicians per year and their management decisions are shown in Figure 5.4. The number of uterine septa seen within the last year ranged from 0 to 50, although the median was 9 (IQR 2–20). Conservative management was estimated from 0 to 25 cases, median 2.5 (IQR 0–10). Septal resection was estimated to be between

0 and 40 cases per year. When excluding clinicians who do not perform septal resection the lower whisker moved to 2 per year, with a median of 4 (IQR = 3–9). Confidence in response is shown in Figure 5.5 with 21 (32%) of respondents confident, 24 (36%) believed the estimate was about right and 18 (27%) guessed their answers. Four respondents (6%) did not answer this question.

A variety of surgical equipment and techniques were described and shown in Figure 5.6. Electrodiathermy was used by 21 of 33 (64%), cold scissors by 9 of 33 (27%) and a combination in 3 of 33 (9%). Surgeons described a variety of techniques including using both electrodiathermy knife or hook, using transabdominal ultrasound guidance or bending the loop of a bipolar resectoscope to adapt it for surgery. Practice post procedure is also variable with some surgeons using adjuncts. These include intrauterine device 12 of 33, oral contraceptive or hormone replacement therapy 13 of 33, antibiotics 9 of 33, in-utero balloon catheter 3 of 33 (10%) or adhesion barriers 3 of 33 (10%). One clinician performed repeat hysteroscopy to as a "second look" procedure 6 to 8 weeks after the initial resection.

Clinicians were most likely (73%) to perform resection for recurrent miscarriage, defined as three or more. However resection was also considered for infertility (56%), one previous miscarriage (22%) or previous preterm birth (37%). These likelihoods are presented in Figure 5.7.

Trial involvement

The number of clinicians willing to randomise patients to either septal resection or control for various indications are shown in Figure 5.8. Twelve clinicians (18%) were willing to consider involvement in a randomised controlled trial of hysteroscopic septal resection.

5.1.5 Discussion

Three-dimensional ultrasound is readily available and over half of clinicians routinely use it, but no consensus exists on classification of anomalies. Septal resection was performed by 46 surgeons, but the annual number of resections ranged widely and were for a variety of indications. Approximately half of clinicians would randomise to a septal resection trial, and most likely

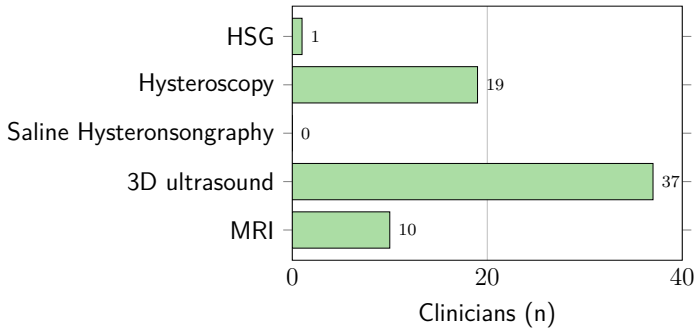


Figure 5.1: What is the gold standard test used in your unit to diagnose and define the presence and type of uterine anomaly?

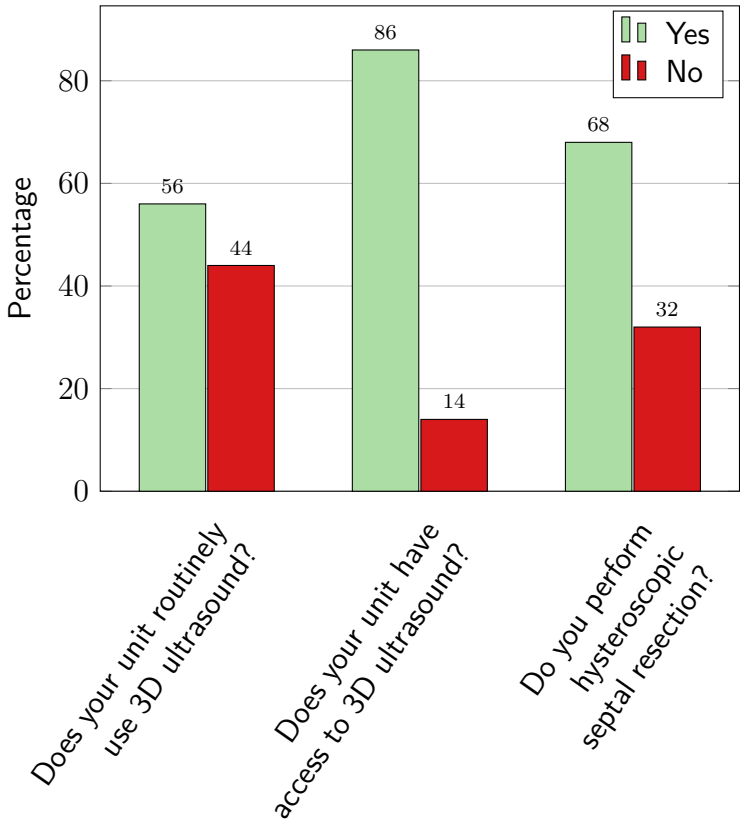


Figure 5.2: Availability of three-dimensional ultrasound and hysteroscopic septal resection (N=67).

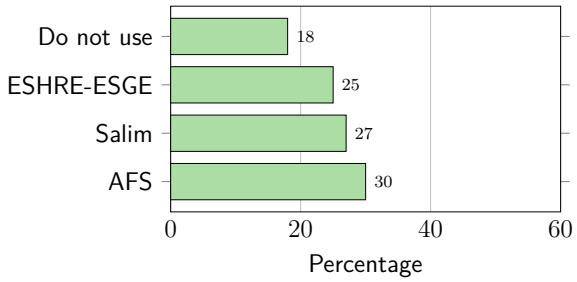


Figure 5.3: Which classification systems do you use to diagnose uterine anomalies? (N=67)

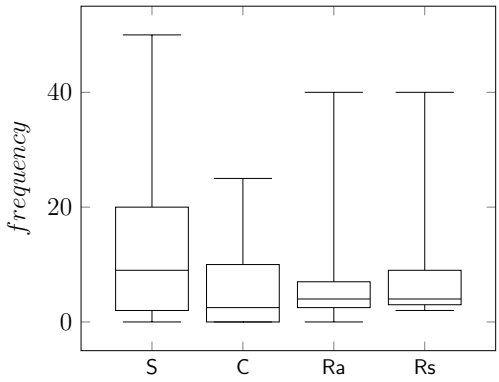


Figure 5.4: Frequency of septal resection: (S) How many uterine septa have you seen in the last year? (C) How many uterine septa have you managed conservatively? (Ra) How many uterine septa have you operated on in the last year? (Rs) How many uterine septa have you operated on in the last year? (excluding clinicians who do not perform septal resection.)

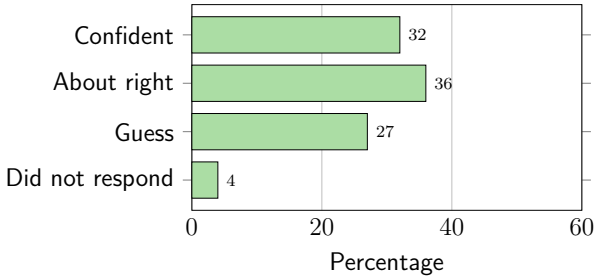


Figure 5.5: How confident are you in your response to the number of septa you have seen and managed in the last year? (N=67)

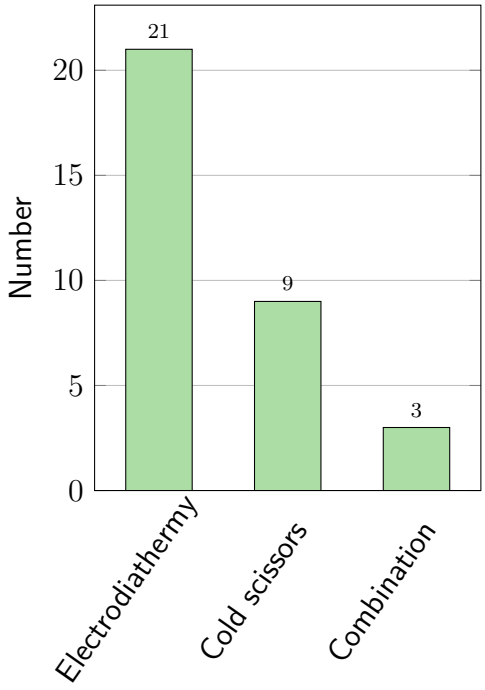


Figure 5.6: Septal resection surgical technique.

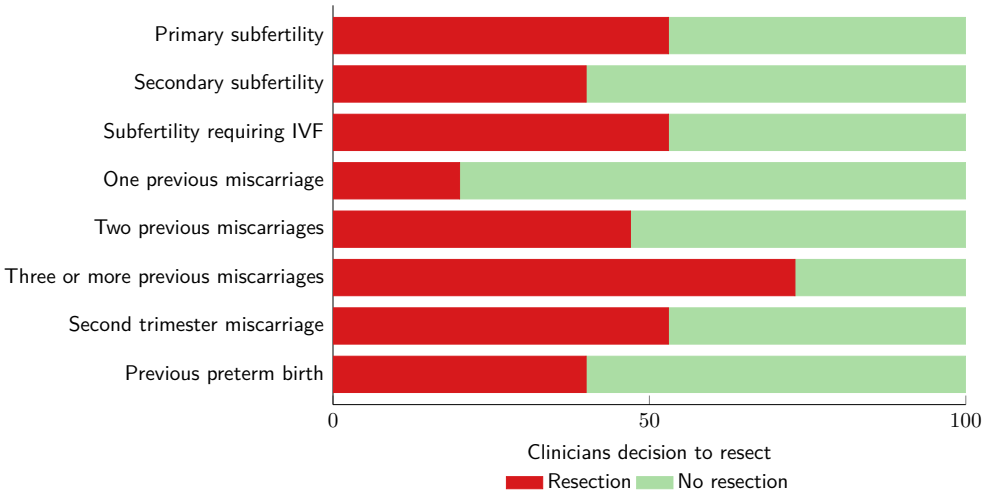


Figure 5.7: For which conditions have you performed septal resection? (N=47)

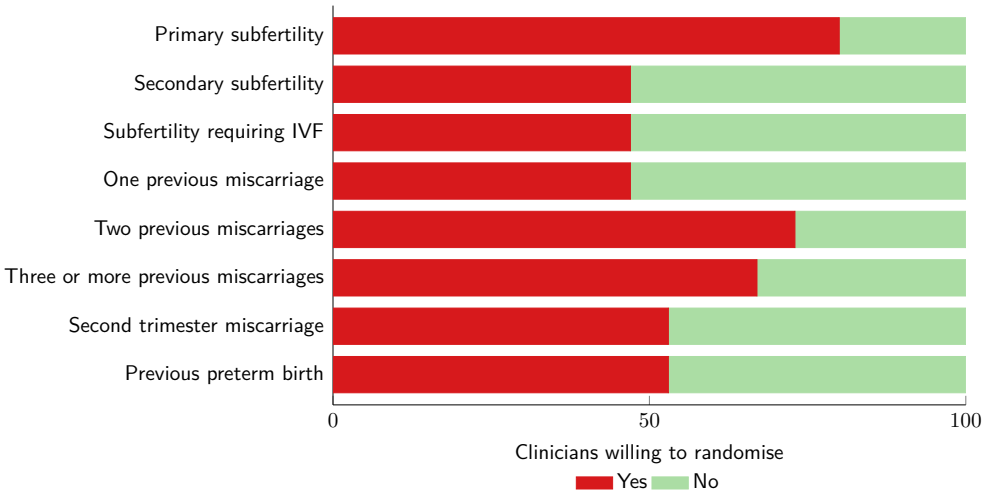


Figure 5.8: For which conditions would you be willing to randomise a patient to either septal resection or control? (N=47)

to randomise women with a history of primary infertility or two previous miscarriages. Only 12 clinicians (18%) showed interest in becoming involved in a trial.

Strengths of this study are that it reflects the wide range of views of clinicians who manage women with uterine septa in the UK and provides data to assess the ongoing feasibility of a randomised controlled trial. It was disappointing to receive a poor response rate to this survey, nonetheless this is consistent with typical response rates to surveys. The majority of respondents worked at tertiary centres, which is where I would expect most septal resections to occur. However respondents are likely to have been self selecting, as those with an interest in uterine anomalies and septal resection more likely to respond. The survey used had not been formally validated, but this is common for surveys assessing viewpoints of clinicians when designing clinical trials. On reflection, the survey failed to ask why many clinicians do not performed septal resection. Reasons for this may be that they were not trained, or did not see enough patients or that they do not believe it works.

In hindsight it would have been useful to ask those not interested in taking part in a septal resection trial why they had come to this decision. Reasons why they were not willing to participate at present can only be derived from speculation from their response to other questions.

Despite over half of units routinely using three-dimensional ultrasound, a significant proportion still rely on other diagnostic techniques. Screening women to identify uterine septa may prove difficult in units with limited access to three-dimensional ultrasound.

Numbers of procedures performed varied widely, which could be due to different populations, size of units and tertiary referrals. The estimates of the number of uterine septa seen and how they had been managed are unlikely to be accurate due to the reliance of recall from individual clinicians. Clinicians performing more resections may be more ready to diagnose a uterus as septate than others who would categorise the indentation as arcuate. For example, one outlier confidently estimated they see over 50 uterine septa per year and resect forty of these. This is inconsistent with median values from other clinicians. It is possible this clinician specialises in hysteroscopic surgery and receives referrals from a wide area. However, looking at other responses from this clinician shows they use hysteroscopy as the gold standard test and the Salim classification system. This is a suboptimal test and Salim classification is intended for use

with three-dimensional ultrasound. Therefore this outlier is likely to categorise uterine morphology as a septum more liberally than others naming uteri as subseptate while others would consider it arcuate.

Surgical technique varied between those who perform resection, but electrodiathermy was the main preference. Reasons for this were not assessed. Some units do not have access to micro-scissors and the method of electrodiathermy was not specified. It could be a standard operative hysteroscope of a small diameter system. These data support the current protocol which does not specify the surgical technique each operator should use as limiting it to a single technique would likely reduce the willingness of clinicians to participate further.

Less than 1 in 5 clinicians were interested in further involvement in a septal resection trial. While this may seem low, it does not mean the trial is not feasible, as a small group of enthusiastic clinicians could recruit sufficient women to conduct the trial.

When clinicians were specifically asked if they will randomise, they are most likely to do so for women with primary infertility, or two previous miscarriages. Only half are prepared to randomise for other indications. These data suggest a significant number of clinicians are in equipoise. So it is then surprising that many were not prepared to take part in a trial. Reasons for this remain unknown.

The high availability of three-dimensional ultrasound is encouraging when considering identifying uterine septa who may be eligible to participate in a trial. Furthermore many units routinely use three-dimensional ultrasound which potentially will provide screening without additional resources.

The median values for number of septa seen was 9 and numbers for conservative management and resection were 2.5 and 4 respectively. These numbers are very small and to achieve an adequate sample size for a definitive study would require the participation of multiple clinicians.

When assessing equipoise it is encouraging clinicians appear split in their decision in which conditions they have resected septa in. When asked for which conditions they would be willing to randomise a patient clinicians are most keen in women with primary subfertility, followed by two then three previous miscarriages. In preterm birth, or second trimester miscarriage just over half of clinicians would be willing to randomise. This supports the direction this thesis

has taken in choosing not to investigate the feasibility of a septal resection trial for previous preterm birth as fewer clinicians are prepared to randomise.

The implication of clinicians performing small numbers of procedures and the not all clinicians are willing to randomise is that trial design should be pragmatic to maximise recruitment. The majority of clinicians use electro-diathermy to perform septal resection, but some use cold scissors or a combination of the two. This supports the pragmatic approach taken in the protocol allowing surgeons to use whichever technique they feel most comfortable. If one technique is mandated it may risk the likelihood of clinicians to participate in the trial or would require further resource to provide additional equipment and training in the alternative technique.

In conclusion, this survey shows there is support for a septal resection trial, but fewer than 1 in 5 clinicians showed interest in taking part. Feasibility will therefore be dependent on active participation of these interested clinicians to become trial sites.

5.2 Study 2: Willingness of patients to be randomised

5.2.1 Introduction

Screening in Nottingham identified 13 women eligible to be entered into the trial, of which two were randomised. I investigated the reasons for non-participation to evaluate if they could be addressed to improve the recruitment rate.

5.2.2 Aims

To assess recruitment rate and reasons why eligible women chose to participate, or not, in the pilot septal resection trial.

5.2.3 Methods

This study used a survey and informal questioning of eligible participants. The recruitment rate was the number of women randomised from eligible patients

over three years.

Survey design

No formally validated surveys were identified, but a literature search found a published survey of why women do not accept randomisation for place of birth [121]. This survey included twelve items on a 4-point Likert scale containing statements respondents could agree or disagree with. Content validity was determined by the researchers by asking the participants to write down their motivation for participation, or not, in the trial.

Using the place of birth survey as a template. I designed a survey which asked respondents to indicate to what extent they agreed or disagreed with eleven statements on 5-point Likert scale. The purpose of this survey was to measure the attitudes of women towards entry into a randomised controlled trial. The addition of a median 'neither agree or disagree', compared to the template survey, allowed the respondent to be indifferent to the statement rather than force them into a preference where it may not exist [122].

Statements respondents were asked to provide their opinion on covered two reasons: not participating in medical research in general, and attitudes towards this particular trial. These covered, autonomy, a preference for a particular arm of the trial, clarity of written and verbal information, and trial withdrawal. The eleven statements and results are presented in Table 5.1.

Evidence from the clinician survey, and my own conversations with doctors caring for women with infertility and recurrent miscarriage, suggested that some clinicians had strong views on the indication of this intervention. Clinicians were at varying ends of the spectrum, some offering resection for small septa with little clinical history, conversely others would not recommend resection of a complete septum after any number of pregnancy losses. To investigate if these views had an effect on participation in Nottingham an additional question addressed the patients interpretation of their clinician's preference.

The final question asked respondents whether they thought hysteroscopic septal resection is still an experimental treatment. The NICE published an interventional procedure overview of hysteroscopic metroplasty of a uterine septum for primary infertility or recurrent miscarriage in 2015. This included the recommendation to ensure that women understand the uncertainty about the procedure's efficacy and its risks [115, 123].

Face validity was evaluated by two experts, my PhD supervisors, Mr Nick Raine-Fenning and Professor Jim Thornton; and by four patients.

Interviews

Alongside the survey I asked eligible women why they did not wish to participate in the trial. Formalised structured interview methodology was not followed. Nonetheless, what women reported provided valuable additional information alongside their survey responses so has been reported in this thesis.

5.2.4 Results

Over three years, two women (15%) were randomised from the thirteen eligible to enter the pilot trial.

Survey

All thirteen eligible women completed the survey when initially approached to participate in the trial. Their responses are summarised in Table 5.1. One woman initially presented after three first trimester miscarriages, and did not wish to enter the study at this point, but after another miscarriage, she entered the trial and was randomised. Her responses from the initial survey are included in the results.

Interviews

Screening identified fifteen women with a uterine septum of whom 13 were eligible. Characteristics of these women are shown in Table 4.1 in Chapter 4. Reasons for participation, or not are, presented for each woman below.

Women A and O both had a body mass index over 40kg/m^2 and neither lost sufficient weight to be considered eligible for the trial.

Woman B was forty four years old and had two previous terminations. The first in her twenties and the second for trisomy 21 four years previous. She now had fertility problems with a poor ovarian reserve. She was also diagnosed with a uterine septum. She could not make up her mind regarding treatment options and so decided to enrol in the trial.

Table 5.1: Reasons for patient non-participation in a septal resection trial.

Proposition	*How much do you agree with this statement?*				
	Strongly disagree	*Mildly disagree*	*Neither agree or disagree*	*Mildly agree*	*Strongly agree*
I had already decided whether to have treatment or not before I was asked to participate in this trial.	0 (0%)	0 (0%)	0 (0%)	10 (77%)	3 (23%)
I want my own choice on what treatment I should have, researchers should not meddle in this.	1 (8%)	5 (38%)	2 (15%)	5 (38%)	0 (0%)
I wanted to be randomly allocated but my partner was against it.	13 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
I do not wish to run the risk of receiving a treatment I do not want.	0 (0%)	0 (0%)	1 (8%)	9 (69%)	3 (23%)
The written information about the trial was not clear.	12 (92%)	1 (8%)	0 (0%)	0 (0%)	0 (0%)
The information the doctor gave me about the random allocation was not clear.	12 (92%)	1 (8%)	0 (0%)	0 (0%)	0 (0%)
Basically I do not participate in medical research.	7 (54%)	6 (46%)	0 (0%)	0 (0%)	0 (0%)
I do not like medical research.	6 (46%)	6 (46%)	1 (8%)	0 (0%)	0 (0%)
I do not understand the purpose of medical research.	12 (92%)	1 (8%)	0 (0%)	0 (0%)	0 (0%)
I never participate in medical research.	12 (92%)	1 (8%)	0 (0%)	0 (0%)	0 (0%)
I am afraid that if I consent to randomisation I cannot get out of it.	0 (0%)	4 (31%)	4 (31%)	4 (31%)	1 (8%)
I think that my doctor had a preference as to which treatment I should receive.	8 (62%)	2 (15%)	1 (8%)	2 (16%)	0 (0%)
Hysteroscopic Septal Resection is still an experimental treatment.	0 (0%)	8 (62%)	0 (0%)	5 (38%)	1 (8%)

Woman C had a complete septate uterus diagnosed when she had a caesarean section at 37 weeks gestation for breech presentation. She subsequently went on to have a 17 week loss. At this point, she recalled her obstetrician informing her that in no circumstances should she let anyone remove the septum. A doctor herself, she weighed up the evidence and was keen to participate in research, but did not want to run the risk of being assigned to septal resection and stated she would withdraw from the study if she was randomised to the intervention. For this reason I did not enrol her in the trial.

Woman D had two failed cycles of IVF and wanted resection before another attempt.

Woman E was 37 years old and had a previous birth at term followed by two first-trimester miscarriages. She did not wish to enter the study as she strongly wanted resection. Her view was that as she was 37 years old her fertility was in decline. She did not consider septal resection to have many risks and did not want to potentially delay resection and miss out on any possible benefit.

Woman F was 26 years old and after a first trimester and 17 week miscarriage wanted to wait and see. She subsequently had a baby at 35 weeks gestation by caesarean section for breech presentation.

Woman G was identified via the recurrent miscarriage clinic. She had four first trimester miscarriages and after discussion with a clinician decided wanted conservative management. She was found to have high levels of uterine natural killer cells and took steroids. A researcher herself, she was eager to participate in medical research, but was not in equipoise regarding septal resection at that point. Several months later, after another miscarriage, her view shifted to equipoise and she was enrolled into the trial.

Woman H was identified by the recurrent miscarriage clinic. She was 40 years old and had a previous preterm birth at 33 weeks gestation who was now fit and well. She wanted a sibling, but had 6 first trimester miscarriages. Initially, she did not want septal resection as the cavity indentation was just over 10mm and broad. Indeed, it could only be classified as a uterine septum by the ESHRE-ESGE classification. After a further miscarriage, her view shifted past equipoise and she underwent septal resection outside of the trial.

Woman I was referred with primary infertility of one years duration. She was not ovulating due to polycystic ovary syndrome and identified to have a partial septate uterus. Despite trial eligibility, after discussion with her clinician she

wanted to address the polycystic ovary syndrome with weight modification and ovulation induction before considering septal resection.

Woman J did not want resection. She conceived naturally and had a normal term live birth.

Woman K wanted resection and did not wish to run the risk of allocation to the control group. She conceived after resection and her pregnancy is ongoing.

Woman L had primary unexplained infertility. After discussion with a consultant she didn't believe it was worth resection. She had a live birth after her first cycle of IVF treatment.

Woman M had subfertility and a first trimester miscarriage. A small septum was diagnosed on scan and she chose to have it resected and was not interested in the trial.

Woman N was 34 years old and had primary infertility and a low ovarian reserve. Her anti Müllerian hormone was 1.5 pmol/l. She was reluctant to consider septal resection and decided not to enrol in the trial.

5.2.5 Discussion

The recruitment rate was 2 per 13 (15%) women with the majority of eligible women unwilling to participate in the pilot septal resection trial.

The survey found that all women had already decided not to participate in the trial before being approached. Their main concern was risking having a treatment they did not want.

Women were not deterred due to their partners, lack of, or poor quality information, or through non-participation in medical research in general. Some women wanted their own choice or were concerned that if randomised to a treatment they did not want, could not withdraw at a later stage. Two women believed their doctor had a preference for which treatment they would receive.

A strength of this study is that willingness to participate data were collected from all women eligible to take part. Additional information obtained by asking women face to face corroborates data from the survey. Each method has strengths, for example, the anonymous survey allowed women to express opinions they would not wish to disclose directly, while interviews enabled me to explore views in more depth.

Limitations from this study are that the methods used lacked rigour. Firstly the survey was not validated to ensure it was measuring what is intended (internal validity) or if the results are applicable to the whole population (external validity). Nonetheless, face validity was assessed to some extent by designing the survey based on a similar one and showing the survey to experts and patients to check language. Formal validity testing was not undertaken because due to the small number of women eligible to participate it would delay the collection of information over the necessary the three year time period.

Formal structured interviews, with transcription and thematic analysis were not conducted. Interviews may have introduced confirmation bias, when inadvertently I may have asked leading questions, confirming beliefs already developed from other sections of this thesis.

Reasons why all the women surveyed had already decided whether to have treatment, or not, before being approached about the trial are unclear. Evidence from the survey and interviews suggest this is multi factorial. Firstly, women felt influenced by clinicians, two women in the survey believed their doctor already had a preference, and when interviewed several women had previously been advised by clinicians to avoid resection or to undergo it. Second, ongoing poor reproductive outcome, as two women changed their minds during the study. One moved from against resection to equipoise and another underwent resection after previously declining surgery. The common factor for both women was they each experienced a further miscarriage before changing their minds. Thirdly, women with other explanations for their infertility or miscarriage were less likely to participate, while those with otherwise unexplained subfertility more like likely to opt for septal resection.

Five women had concerns that once randomised they would not be able to withdraw. Obviously participants can withdraw from research whenever they choose, but these women may have been worried that they would be put under pressure to continue with the trial, or feel as though they had let researchers down if they subsequently withdrew. Placing more emphasis on the rights for women to withdraw may have resulted in more women being randomised, but conversely this may result in more women withdrawing from the study at a later date.

Five women chose septal resection outside of the trial. One of these women previously chose conservative management, but after a further miscarriage chose septal resection. The offer of septal resection outside of the trial is likely

to have limited recruitment. Many unproven interventions are only offered to patients in the context of a research study. However, clinicians in Nottingham already performed septal resection and so were prepared to offer it outside of a trial setting. Herein lies a major factor affecting feasibility. It is unlikely clinicians not already performing septal resection are likely to take it up purely for a randomised controlled trial. Conversely, those performing septal resection are likely to believe it is effective. Allowing women to choose to opt for the intervention outside of the trial is likely to limits the numbers recruited, because 5 out of 13 (38%) women who opted for resection outside the trial may have otherwise entered it.

5.3 Study 3: Exploring the potential for multiple trial sites

5.3.1 Introduction

One year into recruitment it was apparent that insufficient participants could be enrolled from one site. The research ethics committee approved an amendment to run the trial at multiple sites shown in Appendix 1. Feasibility will therefore be dependent on the involvement of other trial sites.

5.3.2 Aim

To explore the potential for multi-centre recruitment to a septal resection trial.

5.3.3 Methods

Clinicians were approached who had expressed an interest from the clinician survey presented in Section 5.1.4. In addition, the British Fertility Society contacted its members by email and individual clinicians from the personal networks of Mr Nick Raine-Fenning and myself were contacted to generate interest.

During this PhD project I was invited to give presentations on uterine anomalies and the background to the septum trial. Presentations to were given at local, regional, national and international meetings. Full details and titles of these

presentations are shown at the beginning of this thesis. The pilot septum trial was promoted at these presentations and members of the audience asked to express their interest in becoming a trial site.

All interested individuals received the trial protocol, and a promotional investigator briefing, shown in Appendix 6. Follow up was by e-mail, telephone or in person. I offered to present the study to the department and assist with local research and development approval.

The stage each unit got to was recorded and reasons for non-participation were sought.

5.3.4 Results

Twelve clinicians, from eleven units expressed interest in becoming trial sites. All sites received a copy of the protocol and investigator briefing. A total of 122 emails were sent to interested sites alongside numerous telephone conversations. Two sites invited me to present the trial to their departments. One site opened for recruitment on 21 September 2016 and has randomised four participants. Reasons for non-participation are shown in Table 5.2.

Five sites did not participate due to the study not being included on the NIHR Clinical Research Network Portfolio. Two sites had limited access to three-dimensional ultrasound. Local service configuration was an issue for two sites, at one site the clinician did not perform septal resection, and the second site required coordination between two organisations to arrange resection which was believed to create an additional barrier. Concerns regarding the diversity of inclusion criteria were given as reasons for non-participation by two centres. One site did not think it could recruit enough participants. Finally, two sites which initially expressed an interest did not respond to repeated attempts to contact them and so were not pursued further.

One centre invited me to present the study to the department, but did not progress to local research and development approval stating several concerns. Firstly, that the population was too broad, and suggested two trials, one for infertility and another for recurrent miscarriage. Secondly, concerns were that the trial was still at pilot stage with no additional funding, and thirdly, that the protocol did not define which classification system should be used to diagnose a uterine septum.

Table 5.2: Reasons for site non-participation in a septal resection trial.

Site	*Extent of communication*	*Reasons for non-participation*
A	Reviewed protocol	Not a portfolio study.
B	Presented at a regional meeting	Restriction of the local health care system. Two Consultants from different sites required to coordinate septal resection. Principal investigator believed this would be a barrier to recruitment.
C	Presented at an international meeting hosted at this site	Not a portfolio study. Clinician keen to be involved but does not perform septal resection.
D	Extensive email and phone correspondence	Unable to arrange a meeting after several attempts to contact. Concluded that site was not interested.
E	Email correspondence	No response to email or telephone calls.
F	Reviewed protocol	No access to 3D ultrasound machine.
G	Email and telephone correspondence	Not a portfolio study. Limited access to 3D ultrasound machines.
H	Departmental presentation	Luke warm reception at departmental presentation. Concerns the trial was only at a pilot stage, classification of uterine septa and trial design involving too diverse population. Not a portfolio study.
I	Reviewed protocol, meeting with site principal investigator.	Not a portfolio study. Local politics.
J	Email and telephone correspondence with principal investigator	Did not think they would recruit many patients. Concerned that the populations were too broad including both infertility and recurrent miscarriage.

5.3.5 Discussion

Initially, twelve clinicians from eleven units expressed interest in becoming a trial site. However only one went on to randomise patients.

The trial was widely publicised through a variety of channels, a clinician survey, regional, national and international presentations, and via our personal networks. Prior to my involvement in this work from August 2014, the pilot study had research ethics approval since October 2013 and was promoted by the previous research fellow Dr Yee Yin Chan and Mr Nick Raine-Fenning, widely known for research using three-dimensional ultrasound and uterine anomalies. It is not possible to know the exact number of clinicians who had heard about this study, but one would estimate that the majority of clinicians who perform septal resection would be aware of this study. Other approaches could have involved requests to contact members of other stakeholder organisations such as the Association of Early Pregnancy Units or the Royal College of Obstetricians and Gynaecologists.

A major limitation was not being included on the NIHR Clinical Research Network Portfolio and this was stated as a reason for non-participation by five units. The NIHR Clinical Research Network Portfolio of studies is a collection of research studies eligible for consideration for support from the Clinical Research Network in all four countries of the United Kingdom. Clinical Research Networks support studies by providing advice, engaging sites and arranging NHS service support for studies [124]. NHS trusts are required to support clinical research and prioritise studies included on the Clinical Research Network Portfolio. I made an application to NIHR for the pilot septal resection trial to be included on the portfolio, but this was refused by NIHR because it was funded by an organisation, Nottingham Hospitals Charity, which does not meet the criteria specified by the Department of Health for studies to receive Clinical Research Network support. It may be possible that more sites would have proceeded to recruitment if consideration was given to the NIHR CRN Portfolio when securing funding for this study.

Access to three-dimensional ultrasound is essential to this study for diagnosis of a uterine septum. Two units did not have sufficient access to three-dimensional ultrasound to become trial sites. Options to overcome this obstacle are either to allow diagnosis using other tests or to supply units with three-dimensional ultrasound equipment and training. Removing the three-dimensional requirement would limit the findings from a trial due to the

problems with accurate, reliable diagnosis, set out in Chapter 2. For example, the TRUST study, an ongoing septal resection trial in the Netherlands, allows diagnosis to be made using hysterosalpingography and diagnostic hysteroscopy, both of which are suboptimal tests and do not allow for uterine measurements to be made. One unit had to "borrow" the machine from another department and the other did not own a scan machine with three-dimensional capabilities. Providing units with a scan machine would significantly increase the costs of this trial, but would overcome this barrier.

Service configuration was a reason for non-participation at one site, but was a contributory factor to recruitment at two other sites. In all three cases, septal resection was only performed by sub-specialists in Reproductive Medicine working in NHS units. Many eligible women are managed by clinicians elsewhere, including those with infertility, miscarriage or recurrent miscarriage managed in secondary care. Furthermore, many units offering assisted conception are outside the NHS and do not have theatre facilities to perform septal resection and rely on referral to another clinician. These complexities mean that eligible patients come from a variety of sources and a common pathway cannot be devised.

Criticisms of the diverse population eligible to participate in the study are justified. Septal resection isn't a treatment for a uterine septum, it is a treatment for the underlying health condition, such as infertility or previous miscarriage. While theories have been proposed to explain the pathogenesis of a septum, described in Section 1.2.6, none have been proven. Indeed, evidence of an association between infertility and a uterine septum is weak, and the pathology for fertility and miscarriage may be different. Nonetheless, this thesis is concerned about demonstrating feasibility of a study and given the small numbers of eligible women identified, further tightening of the inclusion criteria will likely make a trial less feasible.

5.4 Discussion and Conclusions

Only a minority of clinicians (12 of 46) who perform septal resection and women eligible (2 of 13) to participate, were willing to participate in a septal resection trial. This low recruitment rate means feasibility will be dependent on multiple centres recruiting participants. However, only one of eleven interested units went on to become trial sites.

These studies assessed hypothetical willingness to participate by asking clinicians if they would randomise patients, and a significant number said they would. Nonetheless, in reality, virtually none of those interested went on to recruit participants. Reasons for this include not being on the NIHR Portfolio of studies, but also limited availability of three-dimensional ultrasound and local service configuration.

5.4.1 Conclusion

In its present format a definitive septal resection trial is not feasible due to the low recruitment rate and lack of interest from other sites. Achieving the required sample size of 118 participants is not possible in a reasonable time scale, and is likely to be prohibitively expensive.

However, if the barriers identified were overcome, feasibility may be possible. Incentives for sites, such as availability of three-dimensional ultrasound or being included on the NIHR Portfolio of studies, combined with only offering septal resection to women as part of the trial, would shift the balance towards feasibility. Securing funding for a definitive study is more likely if it is considered a research priority.

Chapter 6

Economic evaluation and value of information analysis

6.1 Introduction

As resources are finite, decisions need to be made about which treatments to fund and which not. Economic evaluations can be used to estimate the added costs and added benefits, (or health gains), from one treatment strategy versus another, so can be used to aid decision-makers when choosing where to allocate resources. The same principle can be applied to research, allocating finite funds to where they are most likely to be beneficial. Health economic evaluation can predict the expected cost and benefit from a particular study, such as a trial of septal resection, to see if it is worthwhile.

Decision models are a structured synthesis of all good quality information relevant to a particular decision question. For example, whether a new treatment is cost-effective relative to a current treatment, and so should replace its routine use. Models can be analysed probabilistically yielding a probability distribution for the measure of cost-effectiveness. This decision uncertainty can be used in a value of information analysis to predict return on investment from further research to reduce this uncertainty.

Value of information is centred around the idea of expected loss. Decision uncertainty means there is a probability of making the "wrong" decision, thus incurring a loss of health benefit, such as QALYs forgone. The probability of an error multiplied by the consequence of the error is the expected loss associated with the decision. More information should reduce the probability of the error,

so reduce the expected loss. This expected reduction in expected loss is the expected value of that information, known as the expected value of sample information (EVSI).

Comparing the value of information, specifically the EVSI, with the cost of obtaining it gives a measure of the return on investment. For example, The National Institute for Health and Care Excellence (NICE) are willing to pay £30,000 for one QALY gain in health benefit, this can be compared with the cost per QALY gained of further research, such a randomised controlled trial.

Value of information analysis is seen as a logical next step in economic evaluations. Its use in health technology assessment is beginning to gain traction and a good proportion of health economic evaluations are routinely including a basic value of information analysis [125].

6.1.1 Miscarriage

Miscarriage is defined as pregnancy loss from conception to 24 weeks' gestation. Approximately 15% of all clinically recognised pregnancies end this way [126]. Most miscarriages occur in the first trimester, while only 1% occur after 12 weeks' gestation [127]. Recurrent miscarriage, defined as three or more consecutive pregnancy losses, affects about 1% of couples [128]. About half of miscarriages are chromosomally abnormal [129], although this is lower in women with recurrent miscarriage [130]. It is believed that a proportion of the remainder are of pregnancies carrying lethal recessive genes [131]. Despite evidence that some women with recurrent miscarriage may have an auto-immune, thrombotic or endocrine cause; few immunologic, anti-thrombotic or hormone therapies have yet been shown to be effective. The vast majority remain unexplained and most women receive no explanation for their miscarriage. Despite substantial evidence about the appropriate place for medical, surgical and conservative treatment for ongoing miscarriage, there remain gaps in our knowledge about how to minimise psychological sequelae [132]. There is also uncertainty about the appropriate investigation for women who have had particular numbers of miscarriages [133]. The best way to care for women experiencing miscarriage is uncertain, both when it occurs and in subsequent pregnancies [134].

6.1.2 Progesterone to prevent recurrent miscarriage

To illustrate the concepts of economic evaluation and value of information analysis, I will use the example of progestogen to prevent miscarriage. I chose this example because both progestogen and septal resection are interventions intended to reduce the risk of miscarriage. Progesterone plays an essential role in maintaining pregnancy, and anti-progestogens are used for the termination of pregnancy and induction of labour. Furthermore, recently a large trial was funded to answer this question. In 2009 the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Programme funded a definitive trial of first-trimester progestogen therapy in women with a history of unexplained recurrent miscarriages (PROMISE) at a cost of £1,083,872.64 [135].

Since, no value of information analysis was conducted prior to that trial there is an opportunity to measure the hypothetical value of information before it was done with the real costs and information obtained from the actual trial to test the validity of the method.

I begin with a conventional economic evaluation of the treatment.

In economic evaluation there are always two metrics: cost and benefit, as well as at least two comparators. The cost of an intervention, such as progestogen, can be measured in terms of the monetary cost of providing that intervention, i.e. the cost of a progestogen course. However costs must also include those associated with any potential side effects requiring further treatment, hospital stay for miscarriage treatment, or antenatal and delivery costs. The benefit derived from interventions can be measured in units individual to the intervention such as lives saved or improvement in quality of life, or in this example, QALYs. There are different types of economic evaluation [136]:

Cost-minimisation assumes the benefits produced by two interventions are identical and the only dimension of interest is cost. This is the most basic of health economic evaluations. For example, generic ibuprofen over the branded version Nurofen.

Cost-effectiveness analysis evaluates cost per outcome gained. Outcomes are usually hard outcomes, such as death, or in this example miscarriages prevented. Cost-effectiveness is the added cost to prevent one additional miscarriage. Let's say, progestogen costs £300 per course and five courses prevent 1 miscarriage compared with no progesterone. The cost effectiveness

is £1,500 per miscarriage prevented. This is easy to understand for clinicians, but difficult to compare between different diseases, as they are likely to have different outcomes. For example it is difficult to compare the prevention of miscarriage with another unrelated condition, such as a headache.

Cost-utility analysis considers a common health outcome unit, usually a quality adjusted life year (QALY). This generic outcome unit allows comparison across many, if not all diseases. For example, the cost of progestogen for each QALY gained can be compared with the cost of sumatriptan for treating migraine.

Estimating the utility of a miscarriage prevented is difficult. If a miscarriage is simply the transient pain and suffering of the mother during the process and any ongoing grief or regret she later suffers, the effects are likely to be relatively small. Although acknowledging the considerable uncertainty around the QALY loss from miscarriage, the NICE estimated a 0.1 QALY loss from each miscarriage [137]. However, preventing a miscarriage also leads to the creation of a new individual. Agreeing a monetary value for the creation of a new individual who will probably live a full healthy life is problematic. Health economics does not usually consider "future people" because of philosophical and ethical problems about how they should be counted [138]. However if the future baby is ignored the value of miscarriage is small and the costs to the NHS significantly less than preventing miscarriage. For these reasons, despite the limitations, I will use cost-utility measuring QALYs and lifetime NHS costs of the infant.

Value of information

The results of health economic evaluation recommend the best course of action for decision makers, given current evidence. But there is uncertainty around both the cost and the benefit estimates. There is a possibility that the decision is "wrong", such as deciding to implement progestogen, or no treatment when the converse would be more cost-effective. Further research will reduce this uncertainty, but cost money, which could be used in either other research projects or direct patient care.

Probabilistic sensitivity analysis (PSA) is recommended in guidelines for cost-effectiveness analysis as a means of quantifying decision uncertainty to decision makers [139, 140]. A decision tree model, for progestogen and no

treatment, is described and analysed based on the probability of individual parameters, to estimate mean costs and benefits for each strategy.

The expected value of information (EVPI) is the maximum amount a decision maker should be willing to pay for information prior to making a decision about progestogen or no treatment. It aids decision makers in choosing where future research should be directed, or indeed if any further research would be worthwhile.

I decided to do;

1. A value of information analysis for an area of uncertainty, progestogen to prevent recurrent miscarriage, where a meta-analysis of a number of small, poor quality trials had been followed by a large high quality trial [109, 141].

   My aim was to:

   (a) practice the methodology.

   (b) calculate the point estimate and uncertainty in cost-effectiveness of progestogen before the PROMISE trial via a decision model, the key input being a meta-analysis of four small, poor quality trials [141].

   (c) repeat this process using a prior estimate of effectiveness derived from expert opinion.

   (d) retrospectively predict the expected return on investment from the PROMISE trial based on these models. This would determine whether the PROMISE trial was predicted to be a good investment.

2. Do a prospective value of information study for a trial of septal resection. Compare this with estimated trial cost information derived from my pilot study.

This latter value of information analysis will predict whether a trial of septal resection would be a good use of research funding.

6.2 Study 1: A value of information analysis of progestogen for recurrent miscarriage - before and after PROMISE.

This study is two economic evaluations and value of information analyses of using progestogen to treat women with recurrent miscarriage. Firstly, using prior data from a meta-analysis of four studies available before the PROMISE trial. The second uses the same model, but with a prior elicited from expert opinion. I chose to perform a second analysis using an expert prior to compare results, adjusting for the potential bias from poor quality trials.

6.2.1 Methods

Model structure

A decision tree model was developed to assess the cost-effectiveness of progestogen versus no treatment in terms of additional cost to the UK National Health Service (NHS) per additional birth over 24 weeks gestation. The model is shown schematically in Figure 6.1.

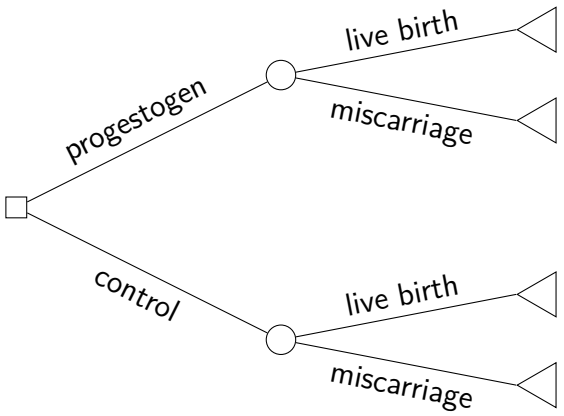


Figure 6.1: Decision tree model structure of progestogen to prevent miscarriage.

The decision problem for the NHS was, which treatment strategy for women with recurrent miscarriage should be adopted: vaginal progestogen (utrogestan 400mg twice-daily), or no drug treatment. The decision tree model for this comparison was developed in R, a statistical computing program [142].

On entering the model, a woman with a history of recurrent miscarriage has a positive pregnancy test. She would either be started on utrogestan, 400mg

twice-daily until 12 weeks gestation, or receive no additional treatment. Women in both arms had previous consultations in a recurrent miscarriage clinic, where progestogen would be prescribed. So no additional costs were incurred for women receiving progestogen. The follow up requirements are also the same for each group.

The model assumes no costs are incurred from side effects of progestogen. Progestogen supplementation has been used for many years in IVF practice for luteal phase support and there is evidence that it is safe to the mother and fetus. A systematic review of observational studies of first trimester sex hormone exposure identified 14 studies, consisting of 65,567 women [143]. In several of these studies progestogens were used alone or with other steroid hormones. They concluded there is no association between first trimester exposure and any external genital malformation. Nonetheless, more recently one case-control study has suggested an association between hypospadias and progestogen use [144]. Intramuscular progestogen has been reported to cause more injection site reactions than placebo [145], however cost and quality of life impact is trivial.

As pregnancy continues, the final outcome for each woman is either a birth over 24 weeks gestation, or a miscarriage. Birth would involve additional antenatal, intrapartum and postnatal care. Miscarriage could resolve spontaneously, or require expectant, medical or surgical treatment. The model assumes women do not have another form of early pregnancy loss, such as an ectopic or molar pregnancy, a stillbirth, or a multiple birth.

Two time horizons were chosen for this analysis. The first in the economic evaluation is the lifetime of the infant and then for the value of information analysis.

The value of information time horizon is supposed to reflect the time period over which the information is useful, i.e. the time until a new technology may be developed making the information redundant. There is no guide for this, but the standard time horizon for these types of analyses is typically 10 years.

Probabilities

Parameter values and associated probability distributions are shown in Table 6.1. The first analysis, used values derived from the best available data. Primarily, from meta-analysis of randomised trials. The second analysis used a

prior derived from expert opinion.

Effectiveness values were derived from data extracted from a meta-analysis of first trimester progestogen prophylaxis in recurrent miscarriage for the outcome of miscarriage [141]. Other meta-analyses and reviews have been published, but contain studies which include women with threatened miscarriage or using progestogen to prevent preterm birth. I decided to use this meta-analysis as it was the most specific to our focused decision question: would progestogen be beneficial for women with unexplained recurrent miscarriage for the outcome of births over 24 weeks gestation?

The meta-analysis included four randomised trials, in which 132 women were treated with progestogen to prevent miscarriage [146–149]. All four trials showed a trend towards benefit, but had wide confidence intervals. One trial showed a statistically significant benefit. Nonetheless, overall there was statistically significant reduction in miscarriage rate with an odds ratio of 0.36 (95% CI 0.19–0.68). The results of this meta-analysis are shown in Figure 6.2.

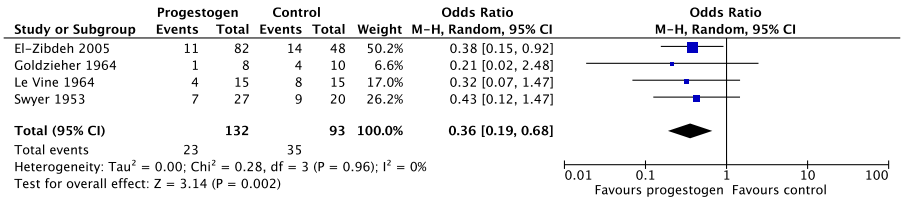


Figure 6.2: Meta-analysis of trials of progestogen in recurrent miscarriage for the outcome of miscarriage.
The figure shows the odds ratio for each study as a square (with size proportional to the amount of information) and the horizontal line depicts the 95% CIs. The open diamond indicates the odds ratio (OR) with 95% CI overall. The vertical line, at the odds ratio of unity, corresponds to the line of no effect.

The meta-analysis combined data from four poor quality studies, the limitations of which are discussed later in this chapter. Taken at face value, the meta-analysis provided clear evidence that progestogens were effective in preventing miscarriage. Nonetheless, even the authors of this meta-analysis, and progestogen enthusiasts, were sceptical that progestogen was this effective, stating that most clinicians have called for a definitive placebo controlled randomised trial targeting live birth as the primary outcome [141]. For this reason, I chose to perform a second analysis using effectiveness data elicited from experts who derived a "more realistic" prior based on their knowledge of the literature and clinical practice. Much has been written on structured expert

Table 6.1: Probability values of progestogen to prevent miscarriage.

Parameter Description	*Mean value*	*95% confidence interval*	*Distribution*	*Source and type of data*
Baseline probability of miscarriage	0.38		Beta (α=35, β58)	Coomarasamy 2011. Random effects meta-analysis of RCTs
Odds Ratio with progestogen	0.36	0.19, 0.68	Log normal	Coomarasamy 2011. Random effects meta-analysis of RCTs
Odds Ratio with progestogen	0.91	0.54, 1.53	Log normal	Distribution fitted to experts' belief.

elicitation processes [150], but the intention this analysis is for illustrative purposes, to see how the results of the value of information analyses vary dependent on the inputs.

The expert elicitation process used a prior based on the opinion of three experts: Professor Jim Thornton, Professor Arri Coomarasamy and myself. A log normal distribution was fitted to the elicited odds ratio and used in the model.

Costs

Costs are shown in Table 6.4 and expressed in 2015 pounds sterling prices. The decision to use the most recent price year was to allow clearer comparison between these analyses and the results of the PROMISE trial.

Progestogens can be administered as either progestogen itself which is usually administered as a vaginal gel or tablet, and 17-alpha medroxyprogesterone acetate which is usually administered by weekly intramuscular injection. Three of four previous trials used intramuscular progestogen [146–148] and one used oral dydrogesterone [149]. Intramuscular progestogen is licensed, as Makena, for the prevention of recurrent preterm birth of singleton gestations in the United States, but not available in the European Union. Vaginal progestogen is not licensed for use in pregnancy but is available in the European Union and widely prescribed for these indications. Dygesterone is an oral progestogen not licensed in the UK or United States but used more commonly in Europe and Asia.

The model assumes the NHS would use vaginal progestogen, due to its EU licensing and reduced cost compared to the intramuscular preparation. The

Table 6.2: Cost of alternative progestogen preparations.

Drug	*Preparation*	*Dose*	*Unit dose*	*Unit cost*	*Cost per patient per day*	*Source*
Utrogestan	Vaginal	400mg bd	200mg	£21 for 21 200mg capsules	£4	BNF 2015
Cyclogest	Vaginal	400mg bd	400mg	£0.86	£1.72	BNF 2015
Makena	Intramuscular	250mg weekly	250mg/ml	£500.33*	£71.48	Drugs.com
Dydrogesterone (Duphaston)	Oral	10mg bd	10mg	£0.26†	£0.52	Medindia.net

* converted from USD using historical exchange rate 1/1/15 1 USD = 0.6419 GBP
† Duphaston 10 tablets 352.5 INR converted from IND using historical exchange rate 1/1/15 1 IND = 0.0102 GBP

PROMISE trial used 400 mg of micronized progestogen (Utrogestan, Besins Healthcare) twice daily. Estimates of the cost of progestogen treatment were taken from the BNF in 2015, as £21.00 for progestogen 200 mg, 21 vaginal capsules. The cost per participant per day, was estimated to be £4.00. Table 6.2 show example costs of other progestogen preparations.

There were no data available for the progestogen course length. Studies within the meta-analysis use progestogen until miscarriage or 36 weeks gestation [148], from diagnosis of pregnancy to 12 weeks gestation [149] or were unclear on course length [146, 147]. The PROMISE trial protocol stated that women were to take progestogen from a positive pregnancy test until 12 weeks gestation [135]. Expert opinion was obtained from a discussion between my supervisor, Jim Thornton, a Professor of Obstetrics and Gynaecology and myself. We reasoned the earliest that the most sensitive of pregnancy tests can detect pregnancy is one week before a period, when a woman is 3 weeks pregnant; and the latest a woman would start progestogen would be at 6 weeks gestation. So the course duration could range between 42 and 56 days. We agreed course lengths between these durations were equally likely.

Miscarriage can resolve with conservative, medical or surgical management. Reference costs and rates of activity are published by the NHS annually. NHS annual reference costs divide the management into miscarriage with, or without, intervention.

Table 6.3 shows costs and activity levels for pathways of care for each outcome branch of the decision tree, namely pregnancy care and miscarriage management. These remain the same throughout the analyses.

Table 6.3: Care pathway costs and activity levels. Source: NHS reference costs 2014-2015 [151]

Ref Cost Code	*Resource item*	*Activity*	*Unit Cost (£)*
Miscarriage costs			
MB08A	Miscarriage with intervention	891	2073.81
MB08B	Miscarriage without intervention	50271	551.56
	Miscarriage costs weighted average		578.07
Live birth costs			
Antenatal costs			
NZ16Z	Ante-Natal Routine Observation	140279	421.40
NZ17A	Ante-Natal False Labour, including Premature Rupture of Membranes, with CC Score 2+	6733	943.07
NZ17B	Ante-Natal False Labour, including Premature Rupture of Membranes, with CC Score 0-1	91260	592.82
NZ18A	Ante-Natal Complex Disorders with CC Score 2+	2944	1648.22
NZ18B	Ante-Natal Complex Disorders with CC Score 0-1	21649	1141.75
NZ19A	Ante-Natal Major Disorders with CC Score 2+	21611	894.11
NZ19B	Ante-Natal Major Disorders with CC Score 0-1	230757	577.74
NZ20A	Ante-Natal Other Disorders with CC Score 2+	2607	728.33
NZ20B	Ante-Natal Other Disorders with CC Score 0-1	37874	529.68
NZ21Z	Ante-Natal Standard Ultrasound Scan	542422	110.77
NZ22Z	Ante-Natal Specialised Ultrasound Scan	37250	131.81
NZ24A	Ante-Natal Therapeutic Procedures, including Induction, with CC Score 2+	1027	1676.63
NZ24B	Ante-Natal Therapeutic Procedures, including Induction, with CC Score 0-1	16421	808.07
	Antenatal cost weighted average		350.20
Delivery costs			
NZ30A	Normal Delivery with CC Score 2+	14806	1761.70
NZ30B	Normal Delivery with CC Score 1	28296	1555.32
NZ30C	Normal Delivery with CC Score 0	159165	1514.13
NZ31A	Normal Delivery, with Epidural or Induction, with CC Score 2+	24853	2319.51
NZ31B	Normal Delivery, with Epidural or Induction, with CC Score 1	24319	2092.52
NZ31C	Normal Delivery, with Epidural or Induction, with CC Score 0	81238	1892.06
NZ32A	Normal Delivery, with Epidural and Induction, or with Post-Partum Surgical Intervention, with CC Score 2+	8970	3091.35
NZ32B	Normal Delivery, with Epidural and Induction, or with Post-Partum Surgical Intervention, with CC Score 1	9402	2705.17
NZ32C	Normal Delivery, with Epidural and Induction, or with Post-Partum Surgical Intervention, with CC Score 0	25221	2245.00
NZ33A	Normal Delivery, with Epidural or Induction, and with Post-Partum Surgical Intervention, with CC Score 2+	1478	3228.91
NZ33B	Normal Delivery, with Epidural or Induction, and with Post-Partum Surgical Intervention, with CC Score 1	6762	1323.58

NZ33C	Normal Delivery, with Epidural or Induction, and with Post-Partum Surgical Intervention, with CC Score 0	7239	1270.43
NZ34A	Normal Delivery, with Epidural, Induction and Post-Partum Surgical Intervention, with CC Score 2+	1174	3627.42
NZ34B	Normal Delivery, with Epidural, Induction and Post-Partum Surgical Intervention, with CC Score 1	1802	3108.82
NZ34C	Normal Delivery, with Epidural, Induction and Post-Partum Surgical Intervention, with CC Score 0	1910	2889.95
NZ40A	Assisted Delivery with CC Score 2+	1486	2588.60
NZ40B	Assisted Delivery with CC Score 1	4149	1711.38
NZ40C	Assisted Delivery with CC Score 0	11911	1670.86
NZ41A	Assisted Delivery, with Epidural or Induction, with CC Score 2+	4429	3266.09
NZ41B	Assisted Delivery, with Epidural or Induction, with CC Score 1	7038	2757.66
NZ41C	Assisted Delivery, with Epidural or Induction, with CC Score 0	16035	2387.51
NZ42A	Assisted Delivery, with Epidural and Induction, or with Post-Partum Surgical Intervention, with CC Score 2+	7622	3683.32
NZ42B	Assisted Delivery, with Epidural and Induction, or with Post-Partum Surgical Intervention, with CC Score 1	10026	3233.05
NZ42C	Assisted Delivery, with Epidural and Induction, or with Post-Partum Surgical Intervention, with CC Score 0	16716	2901.12
NZ43A	Assisted Delivery, with Epidural or Induction, and with Post-Partum Surgical Intervention, with CC Score 2+	511	3878.56
NZ43B	Assisted Delivery, with Epidural or Induction, and with Post-Partum Surgical Intervention, with CC Score 1	1333	3237.64
NZ43C	Assisted Delivery, with Epidural or Induction, and with Post-Partum Surgical Intervention, with CC Score 0	1310	2912.96
NZ44A	Assisted Delivery, with Epidural, Induction and Post-Partum Surgical Intervention, with CC Score 2+	1376	3476.40
NZ44B	Assisted Delivery, with Epidural, Induction and Post-Partum Surgical Intervention, with CC Score 1	1922	3698.36
NZ44C	Assisted Delivery, with Epidural, Induction and Post-Partum Surgical Intervention, with CC Score 0	1590	3401.28
NZ50A	Planned Caesarean Section with CC Score 4+	5490	4491.50
NZ50B	Planned Caesarean Section with CC Score 2-3	21745	3470.70
NZ50C	Planned Caesarean Section with CC Score 0-1	41602	2922.25
NZ51A	Emergency Caesarean Section with CC Score 4+	6555	5847.51
NZ51B	Emergency Caesarean Section with CC Score 2-3	24270	4707.33
NZ51C	Emergency Caesarean Section with CC Score 0-1	65835	3820.37
	Delivery costs weighted average		2442.49

In 2015 the annual reference cost for miscarriage with interventions was £2,073.81 with an activity level of 891; without interventions was £551.56 with an activity level of 50,271 [151]. A weighted average of cost dependent

on activity level of £578.07 was derived as the cost of miscarriage treatment. The additional cost of birth over miscarriage depends on the complexity of antenatal and intrapartum care. Weighted average costs dependent on activity level for antenatal care of £350.20 and intrapartum care of £2442.49 were derived as the cost of birth.

Table 6.4: Cost values.

Parameter Description	*Mean value*	*Distribution*	*Source*	*Type of data*
Progestogen				
Utrogestan per day	£4.00	Constant	BNF 2014	Cost
Course length (days)	49	Uniform (42-56)	Expert opinion	Duration
progestogen Course	£168.00		Imputed	Cost
Pregnancy Care				
Miscarriage	£578.07	Constant	NHS Reference costs 2014-2015	Weighted average
Antenatal Care	£350.20	Constant	NHS Reference costs 2014-2016	Weighted average
Intrapartum Care	£2,442.49	Constant	NHS Reference costs 2014-2017	Weighted average
Live birth	£2,792.69	Constant	NHS Reference costs 2014-2018	Weighted average
Cost of life	£55,005.81	Constant	Nuffield Trust, ONS	Discounted cost

The lifetime cost of a person to the NHS was derived from (1) the average cost per person to the NHS per person per year [152] and (2) the average life expectancy in 2014 [153]. The average cost to NHS per person per year in 2014 was £2039 and the average life expectancy at birth was 81 years. Discounting £2039 by 3.5% per year from birth to 81 resulted in a cost of £55005.81 per additional person to the NHS.

Outcomes

Utility was derived using an estimate for the number of quality adjusted life years (QALYs) gained from an additional birth. The lifetime QALY value was calculated from a paper used to show the effect of deprivation on quality of life [154]. The average of the 1st and 5th quintiles QALYS for males at birth resulted in 71 QALYs for an additional life not discounted.

The utility value of a miscarriage was obtained from the health economic evaluation completed by NICE when evaluating the impact of miscarriage and ectopic pregnancy [137]. An illustrative value of a loss of 0.1 QALYs was used to represent the effects of miscarriage on maternal quality of life.

Point-estimates

The incremental cost-effectiveness ratio (ICER) is used in health economic evaluation to summarise the cost-effectiveness of an intervention. It is defined by the difference in cost between progestogen and no treatment, divided by the difference in their effect. In this case, the ICER is the extra cost to gain one QALY. It was calculated using the following equation:

$$ICER = \frac{\mathsf{E}[\Delta_c]}{\mathsf{E}[\Delta_e]} = \text{additional cost to gain one QALY.} \qquad (6.2.1)$$

Where Δ_c is the difference in cost and Δ_e is the difference in QALYs, between progestogen and no treatment.

Investigation of uncertainty

Effectiveness parameters were extracted from the meta-analysis which used a random effects model. So uncertainty around the mean coefficient and standard error around treatment group is reflected in the probability distribution.

Modelling was performed by describing each parameter by a suitable distribution. A beta distribution was applied to the baseline probability of miscarriage and the log-normal distribution around the odds ratio of miscarriage with progestogen. A beta distribution was chosen to represent a probability as it falls within a boundary of 0 and 1. A uniform distribution was applied to the cost of progestogen course. This is shown in Table 6.1.

Uncertainty in outcomes was investigated using a probabilistic sensitivity analysis of 100,000 replications. Scatter plots of the incremental cost and incremental benefit in QALYs, for progestogen compared with no treatment for each simulation were prepared for each analysis.

This health economic model is a cost-utility analysis, comparing cost to the outcome of an additional QALY. A cost-effectiveness acceptability curve (CEAC) is a graph summarising the impact of uncertainty on the result of an economic evaluation, expressed as the probability that a particular strategy is cost-effective as a function of the willingness to pay for a unit of outcome, in this case, a birth, or miscarriage averted. The graph plots a range of willingness-to-pay thresholds on the horizontal axis against the probability that the intervention will be cost-effective at that threshold on the vertical axis. The

cost-effectiveness acceptability curve helps the decision-maker to understand the uncertainty associated with making a particular decision to approve or reject a new health technology [155].

Cost-effectiveness acceptability curves (CEAC) were generated for each analysis. These reflect the probability that progestogen compared with no treatment is cost effective at various thresholds of willingness-to-pay for a QALY.

Value of information analysis

Cost-effectiveness acceptability curves inform decision makers about the probability that a given strategy is cost-effective. The ICER shows the added cost for every unit of benefit, i.e. cost per additional QALY. However, there is uncertainty surrounding this decision dependent on the probability of cost and benefit parameters.

Value of information is the amount a decision maker should be willing to pay for information before making a decision. It is based on the belief that investment in further research into the effectiveness, costs and utility will reduce uncertainty surrounding these parameters, and reduce overall decision uncertainty. Perfect information, is when there is no uncertainty in all parameters used in the model. The population expected value of perfect information (EVPI) is the difference between the expected value of a decision (the maximum expected net benefit, a rearrangement of the ICER formula), with current information and with perfect information, multiplied by the patient population in the present and future who could benefit from the information. This provides a guide for the upper boundary of investment in future research to answer the decision question.

$$\begin{aligned} \text{EVPI} = (&\text{expected payoff with perfect information} - \\ &\text{expected payoff with existing information}) \\ &\times \text{patient population} \end{aligned} \tag{6.2.2}$$

Rather than calculating all parameter values at once, an alternative strategy is to eliminate uncertainty in one parameter, or a group of parameters at a time. The difference between the maximum expected net benefit with this partial perfect information and with current estimates with some uncertainty gives the

expected value of partial perfect information (EVPPI). This shows the benefit of reducing uncertainty in a parameter or subset of parameters, providing information on where future research should be directed most efficiently. Due to the simplicity of this model, there are only two parameters: the probability of miscarriage with no treatment or progestogen; and length of progestogen course. So the expected value of partial perfect information was not calculated.

Value of information calculations were based on 6000 recurrent miscarriage per year with a time horizon of one year. This was derived from 1% of couples trying to conceive having recurrent miscarriage, with an estimate of incidence of 600,000 pregnancies per year [153, 156].

A willingness to pay threshold was set at £30,000 per QALY. NICE will typically pay up to this amount per QALY when deciding to recommend treatments [157].

Estimated value of sample information (EVSI)

Eliminating all uncertainty from the model is not possible in real life as it is impossible to sample the whole population. However it is possible to assess the cost effectiveness of different research studies by valuing the expected gain of a trial of varying sample sizes. This study is illustrative because a real trial, PROMISE, was performed, so the value of information before the trial can be compared with the cost of the trial itself.

A nested Monte Carlo simulation approach was used to calculate the predicted posterior EVPI for different sample sizes. The inner loop consisted of 100,000 simulations from the prior parameter distribution and the outer loop of 10,000 preposteriors. This was repeated for sample sizes ranging from 1 to 2000. The predicted posterior EVPI was calculated using a simulation based on likelihood function for sampled values from the prior parameter distribution. This process was repeated using both prior distributions: the meta-analysis and expert opinion.

The EVSI is the difference between prior EVPI and the predicted posterior EVPI, which is then multiplied by the patient population, less the number enrolled in the trial, as they cannot benefit from the information.

$$\begin{aligned} \text{EVSI} = [&(\text{expected payoff with sampled information} - \\ &\text{expected payoff with existing information}) \\ &\times \text{patient population}] \text{ - number enrolled in the trial} \end{aligned} \tag{6.2.3}$$

6.2.2 Results

Point-estimate cost-effectiveness

The incremental cost effectiveness ratio (ICER), which is the cost for each additional QALY, using effectiveness data from the meta-analysis and expert opinion, were calculated from 100,000 simulations for each, these were £819.17 and £947.67 respectively.

Sensitivity analysis

Sensitivity analysis using the first prior, from the meta-analysis is shown in the scatter plot in Figure 6.3. This shows the cost-effectiveness plane of the difference in cost versus difference in benefit from 100,000 simulations. Nearly all points are scattered in the north-east quadrant, showing that progestogen is more effective but more expensive. Notice the perfect correlation between difference in cost and difference in QALYs due to the structure of the model.

A second scatter plot of the incremental cost versus incremental benefit for 100,000 simulations using the prior derived from expert opinion is shown in Figure 6.4. No points were scattered in the south-east quadrant suggesting progestogen is unlikely to be effective and cheaper, however points are scattered across the both north and south west quadrants, questioning both the effectiveness and cost of progestogen.

The willingness to pay threshold of £30,000 per QALY is shown in both scatter plot figures. When a simulated point lies within the shaded region, it shows that under these conditions progestogen is more cost-effective. So in the first plot, approximately 99.9% of simulations show progestogen to be more cost effective. Contrast this with the second plot where fewer simulation points are in the shaded area, showing progestogen is less likely to be cost-effective.

Cost-effectiveness acceptability curves are shown in Figures 6.5 and 6.6. These charts show how the probability of cost-effectiveness of progestogen changes

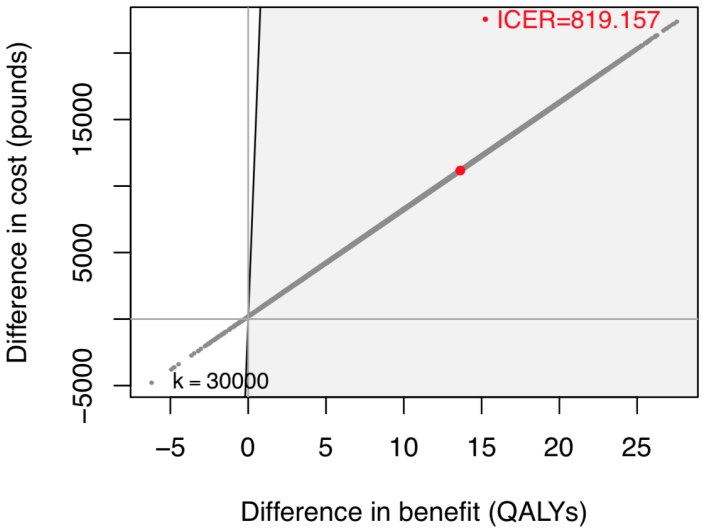


Figure 6.3: Cost-effectiveness plane for progestogen to prevent recurrent miscarriage using meta-analysis prior. Showing 100,000 simulations. ICER: incremental cost effectiveness ratio, k: willingness to pay.

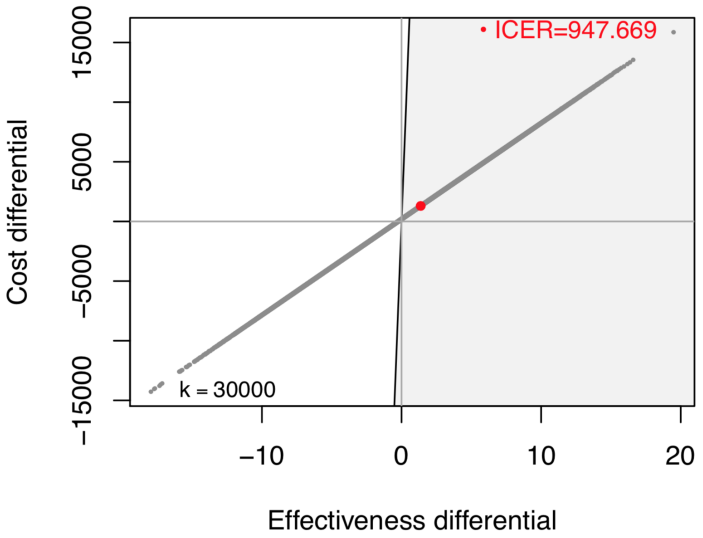


Figure 6.4: Cost-effectiveness plane for progestogen to prevent recurrent miscarriage using a prior derived from expert opinion. Showing 100,000 simulations. ICER: incremental cost effectiveness ratio, k: willingness to pay.

as the willingness to pay increases.

Figure 6.5, using the meta-analysis prior, shows the curve levelling off at willingness to pay thresholds above £1000 per QALY. At a willingness to pay threshold of £30,000 per additional QALY, there is a 99.9% probability that progestogen is a cost-effective alternative to no treatment.

Figure 6.6, using the expert opinion prior looks different. The curve initially falls, then rises, and plateaus at willingness to pay thresholds above £1,000 per additional QALY at around 64%.

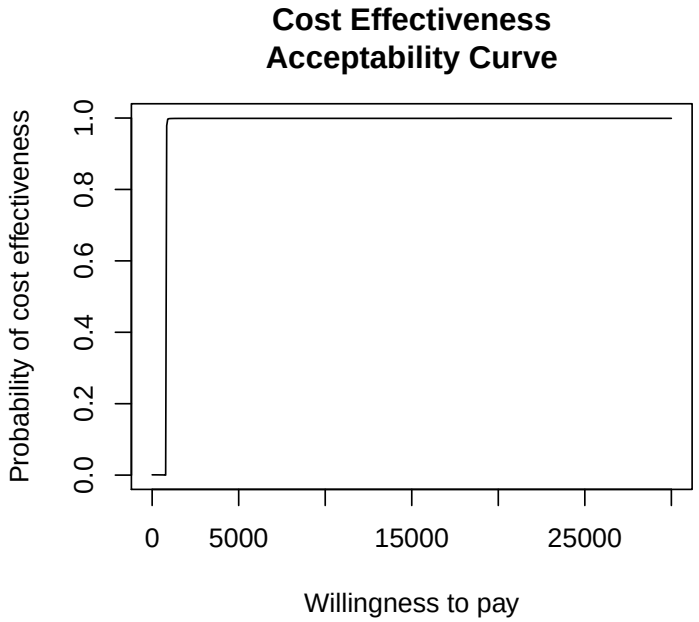


Figure 6.5: Cost-effectiveness acceptability curve for progestogen to prevent recurrent miscarriage using the meta-analysis prior.

Value of information

Uncertainty exists within this model, about the probabilities of the effectiveness of progestogen, and the course length. So if the decision is uncertain, there is a possibility it is wrong. The probability of being wrong multiplied by the loss associated with being wrong, is the expected loss associated with the uncertainty. This is the same as the expected value of eliminating this uncertainty–the expected value of perfect information (EVPI). Per patient EVPI is shown in Figures 6.7 and 6.8.

The population EVPI is the per patient EVPI multiplied by the population who can benefit, in this case it was estimated to be 6000 women per year. At a

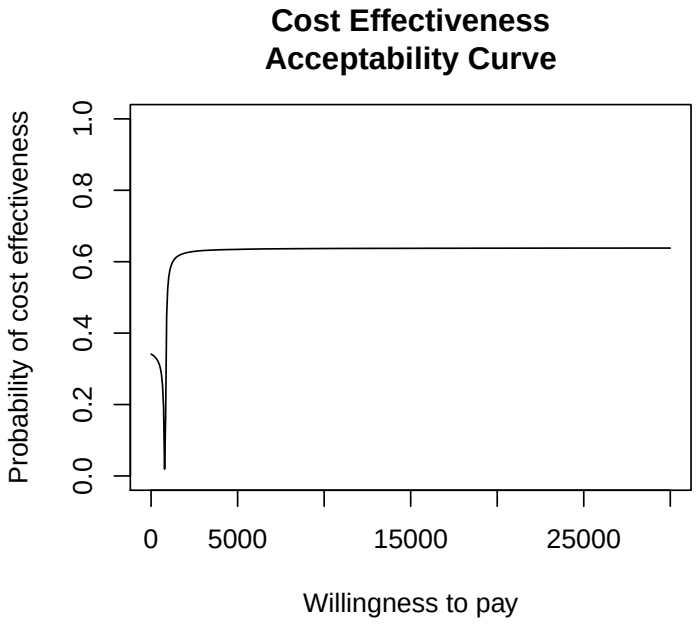


Figure 6.6: Cost-effectiveness acceptability curve for progestogen to prevent recurrent miscarriage using the prior derived from expert opinion.

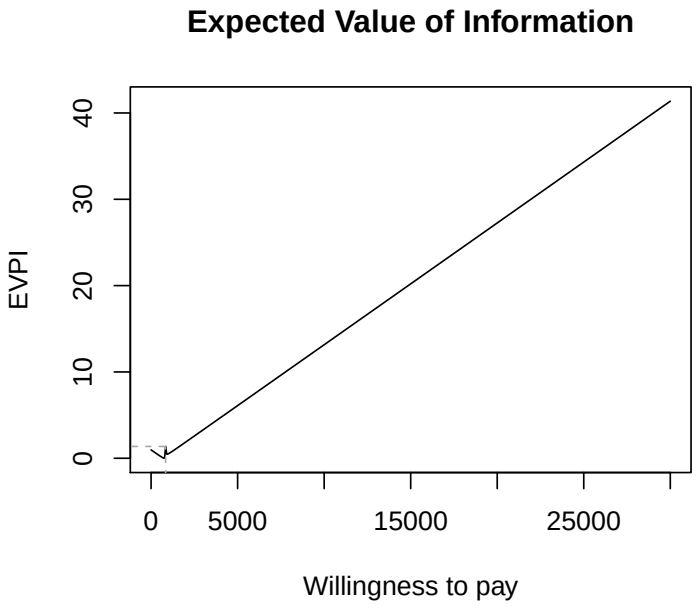


Figure 6.7: Expected value of perfect information (EVPI) for progestogen to prevent recurrent miscarriage using the meta-analysis prior. The dashed line shows the break even point, where the expected difference in benefit between strategies is zero, ie when the optimal decision changes from no treatment to progestogen.

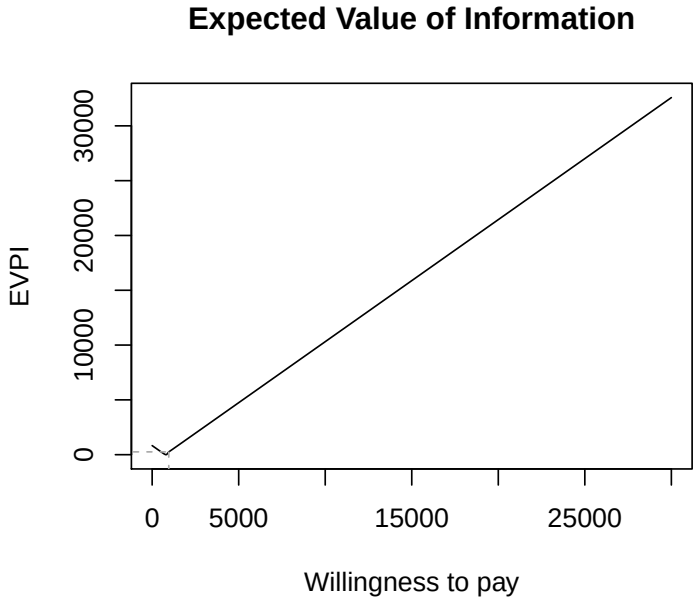


Figure 6.8: Expected value of perfect information (EVPI) for progestogen to prevent recurrent miscarriage using the prior derived from expert opinion. The dashed line shows the break even point, where the expected difference in benefit between strategies is zero, ie when the optimal decision changes from no treatment to progestogen.

willingness to pay threshold of £30,000, using the meta-analysis effectiveness data, population EVPI is £248,190. Using the prior derived from expert opinion the EVPI is £1.68B.

When willingness to pay is zero, the decision maker does not care about outcomes, only which strategy is cheaper. If there is uncertainty in cost there is potentially value in eliminating it. As the willingness to pay approaches the ICER, the EVPI peaks, then falls, before rising again to infinity. The ICER is defined by the difference in cost between progestogen and no treatment, divided by the difference in their effect. In this case, the ICER is the extra cost to gain one QALY. So if one is willing to pay £30,000 for a QALY: an ICER of £29999 is just cost effective, but the decision is extremely sensitive to uncertainty. A small increase in the ICER by £2 and it's not cost-effective. However, if the ICER was £10, or £1,000,000, the decision is less sensitive to uncertainty, and adding or subtracting £2 at these levels will not change the decision. This effect of uncertainty at willingness to pay thresholds close to the ICER is responsible for the peak in EVPI. So as willingness to pay increases above the ICER the effect of uncertainty on the decision decreases. However, more value is placed on the uncertainty in outcome, a QALY. Eventually this becomes so great that the EVPI begins to rise again. So if you value a QALY

infinitely, then it is worth paying anything for reduce uncertainty in it.

In this study, the point estimate ICER remains reasonably stable from £819.17 to £947.67 between the meta-analysis and expert priors. However there is more uncertainty that progestogen is cost-effective, as demonstrated by the £30,000 willingness to pay threshold shaded areas on the both cost-effectiveness planes in Figures 6.3 and 6.4. Basically, the experts' belief is a lower effectiveness of progesterone, resulting in more uncertainty. So uncertainty in the odds ratio is higher, shown by the cluster of points being more spread out, and there is more value in establishing which is the most cost-effective strategy.

Estimated Value of Sample Information

The estimated value of sample information for various sample sizes at a willingness to pay threshold of £30,000 are shown in Figure 6.9.

The EVSI using the meta-analysis prior, shown by the orange line, is practically £0 for all sample sizes. Evidence from the meta-analysis suggests that progestogen is so, and relatively certain to be, cost-effective, i.e. it is effective and cheap, that it is not worth conducting further research to eliminate any uncertainty.

However, the population EVSI calculated from the expert opinion prior, shown by the blue crosses, increases from £0 to approximately £1.6B when the sample size is 25 or more.

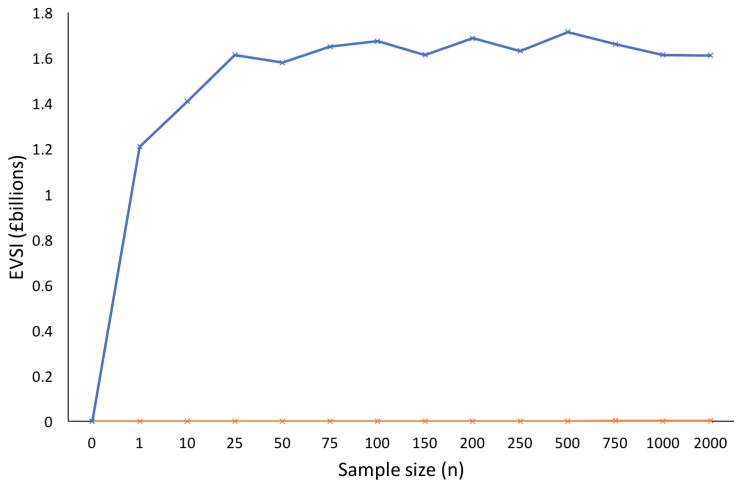


Figure 6.9: Estimated Value of Sample Information (EVSI) for progestogen to prevent recurrent miscarriage. Meta analysis prior (orange) and expert opinion prior (blue). The wobble in the line represent simulation noise.

6.2.3 Discussion

Main findings

This study estimated that when using current evidence the expected value of sample information (EVSI) of conducting a trial of progestogen to prevent miscarriage to the NHS for sample sizes up to 2000 participants over a 1 year time horizon was zero. This is because progestogen is cheap, and the population who may benefit is relatively small. The meta-analysis suggested it was effective with a fairly high degree of certainty, so there was little value for the NHS to being even more certain that it was effective. When using an expert prior, the probability of cost-effectiveness is more uncertain, so the value of perfect information is higher and the expected value of sample information is as much as £1.6B.

In the real world the PROMISE trial cost £1.1m to conduct. So if this value of information analysis had been performed prior to the trial, using existing evidence, it would suggest it was not worthwhile funding.

Nonetheless, the PROMISE trial unexpectedly, according to the meta-analysis, showed that progesterone was ineffective in preventing miscarriage in this scenario (relative risk, 1.04; 95% confidence interval, 0.94–1.15) and so considered by clinicians and patients to be very worthwhile [109].

EVSI recalculation

It is worth noting that in the first iteration of this Chapter, to avoid the complication of assigning a utility value to future life, QALYs were not used as an outcome. Instead outcomes were measured in births over 24 weeks gestation to provide a cost-effectiveness analysis of cost per additional birth over 24 weeks gestation. Therefore a value of 1 was assigned to a birth and 0 to a miscarriage.

However, analyses using these inputs resulted in it never being cost-effective to prevent a miscarriage. This is because the cost of a miscarriage (£578.07) was nearly ten times less than the cost of an additional birth (£5,585.38). So the analyses was revised using QALYs, to consider the cost and utility from a future life, compared with the QALY loss from miscarriage. This is an issue with these analyses, should it focus on the mother, infant or both?

Strengths and limitations

The results of this analysis rest on three crucial assumptions: First, the model structure is plausible; second, that the distributions around the input parameters accurately reflect current evidence and knowledge; and third, the willingness to pay threshold is realistic.

The model structure was designed to reflect likely outcomes from the clinical pathway, the assumption that women would not have an ectopic, molar or multiple pregnancy, or a stillbirth may have some effect on the results. Also assumed, is that progestogen does not affect risk of premature birth with the added resource requirements, but this is a reasonable assumption as many advocate progestogen as a treatment to prevent preterm birth. The second assumption, that the NHS would use utrogestan, rather than cyclogest, a cheaper alternative, or the more expensive intramuscular preparation may mean that the additional cost of progestogen is exaggerated or underestimated respectively. Side-effects of progestogen were also not considered.

This analysis calculated the cost-utility to the NHS rather than society as a whole. Preventing miscarriage will also lead to societal effects. For example, mothers will not require time off work during the miscarriage, but will be entitled to maternity benefits. Furthermore the existence of a new human being, will ultimately lead to a contribution to society, such as their productivity, tax receipts, sense of community and perhaps children of their own. Potential benefits which are impossible to plausibly quantify.

Health economic evaluation, like all statistical analyses is dependent on the quality of the data inputs. The four studies used in the meta-analysis providing effectiveness parameters involved small numbers of women and the trial quality was poor. Different preparations of progestogen were used, none of which were vaginal. Three studies were not randomised, and the fourth did not provide the method of randomisation. Two studies were unblinded and intention to treat analysis was only performed in one study. Researchers in one study excluded 26 (46%) patients, the reasons for this are unclear. The first study was published in 1953, and two others in 1964, the same year the Declaration of Helsinki was adopted and well before the Consolidated Standards of Reporting Trials (CONSORT) statement and clinical trials registries were established. None of the trials were registered.

I have shown that value of information analysis when conducted using seriously

biased information can provide misleading results. For example, in the present example it suggested progestogen was already a cost-effective treatment for recurrent miscarriage, and that further information would be of little value.

On the face of it this suggests that the value of information analysis was flawed because the PROMISE trial unexpectedly showed that progesterone was ineffective in preventing miscarriage in this scenario. The PROMISE results showed the rate of live births was 65.8% (262 of 398 women) in the progesterone group and 63.3% (271 of 428 women) in the placebo group (relative rate, 1.04; 95% confidence interval, 0.94 to 1.15). However, was this really unexpected? In practice many scientists and clinicians believed perfectly well that this evidence was biased for a number of reasons:

1. The poor study quality mentioned above.
2. The huge effect size which is implausible since most miscarriage are due to untreatable causes.
3. The lack of a progestogen deficiency state.
4. Unregistered trials, leading to selective outcome reporting.

Although not presented in this thesis, I conducted a systematic review and meta-analysis of progestogen use to prevent all pregnancy complications [158]. I identified 29 systematic reviews with meta-analyses of progestogen versus placebo for a range of pregnancy indications including: prevention of preterm birth in different populations (singletons, twins, short cervix, previous preterm birth, premature rupture of membranes), prevention of pre-eclampsia, and prevention of threatened or recurrent miscarriage. Nineteen of these meta-analyses concluded progestogens were effective.

The search identified ninety-three randomised controlled trials. In our review, trials were restricted to those of the highest quality, namely, ones that had been prospectively listed in a World Health Organisation recognised publicly-registered trials registry with a predefined primary outcome, and which either achieved their pre-specified sample size, or were both double blind, and failed to achieve their sample size for logistical reasons unrelated to the trial result. The primary outcome had to be included in the trial report. Trials were excluded if they were not registered or registered late. Applying these restrictions left 22 trials included in the meta-analysis. Results are shown in

Figures 6.10 and 6.11. The review concluded that when evaluated in registered double-blind trials with analysis restricted to predefined primary outcomes, progestational agents in pregnancy are ineffective. In the present study, none of the four trials used to derive the prior had been registered, and so could have also been subject to selective outcome reporting. Therefore it is not unexpected that a well designed and properly conducted trial, such as PROMISE, did not find any benefit of progestogen.

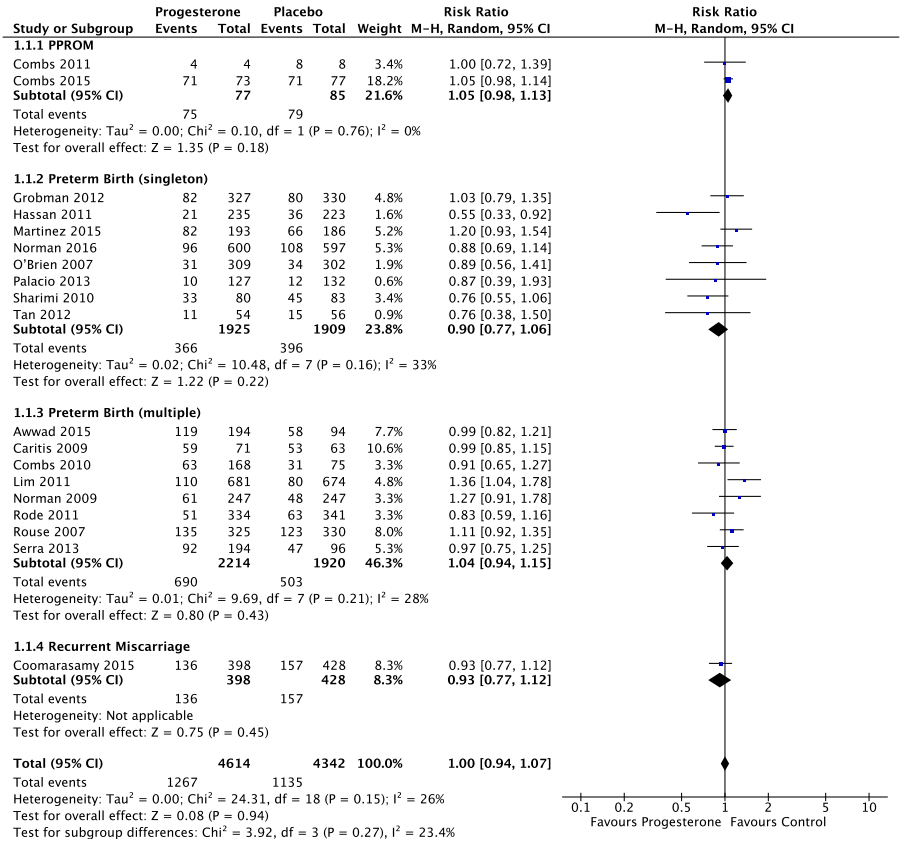


Figure 6.10: Progestogen versus placebo: primary registered dichotomous outcome.
The figure shows the odds ratio for each study as a square (with size proportional to the amount of information) and the horizontal line depicts the 95% CIs. The open diamond indicates the odds ratio (OR) with 95% CI overall. The vertical line, at the odds ratio of unity, corresponds to the line of no effect.

To account for these limitations the second analysis, using a prior derived by expert opinion was performed. Nonetheless, this too was limited by not following a structured expert elicitation process and all three experts knew the results of the PROMISE trial.

These results raise the question of the usefulness of value of information analysis. I have shown in this case, that when existing evidence is poor, the expected value of information is low. Replacing it by using some form of

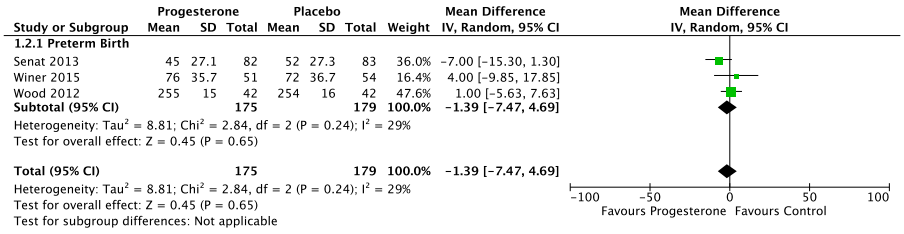


Figure 6.11: Progestogen versus placebo: primary registered continuous outcome.
The figure shows the odds ratio for each study as a square (with size proportional to the amount of information) and the horizontal line depicts the 95% CIs. The open diamond indicates the odds ratio (OR) with 95% CI overall. The vertical line, at the odds ratio of unity, corresponds to the line of no effect.

expert elicitation is helpful, but experts are also bias. Would the NIHR have reconsidered funding for PROMISE had they known the results of this chapter? Or is value of information analysis a waste of resources in itself and trials should be planned anyway? If so, how does that help us choose between competing trials that also have poor quality prior data?

Conclusion

Data before the PROMISE trial showed progestogen is beneficial and cost-effective for women with unexplained recurrent miscarriage at a willingness to pay threshold of £30,000 per QALY. When using existing evidence, the expected value of perfect information (EVPI) and sample information (EVSI) was low compared with the actual cost of the PROMISE trial.

Accounting for misleading bias trials, by eliciting expert opinion to derive new priors, increased uncertainty and the EVSI exceeded the cost of the PROMISE trial making it look like good value for money as the EVSI was approximately 1400 times the cost of the trial.

6.3 Study 2: A value of information analysis of hysteroscopic septal resection for infertility and miscarriage

Comparing the value of information for a trial of septal resection with the results of Study 1 will guide funders whether a such a trial would be good

use of research resources. Furthermore, as the PROMISE trial was funded, it provides a baseline for what may be considered as good allocation of resources.

In this study I perform a economic evaluation and value of information analyses using the same methods as the previous study, but using a model for septal resection.

6.3.1 Methods

I conducted a health economic evaluation and value of information analysis of septal resection for women with a history of infertility or miscarriage to prevent further miscarriage.

Model structure

A probabilistic decision tree model was developed to assess the cost-effectiveness of septal resection versus no treatment in terms of additional cost to the UK National Health Service (NHS) per additional QALY. This model is shown in Figure 6.12.

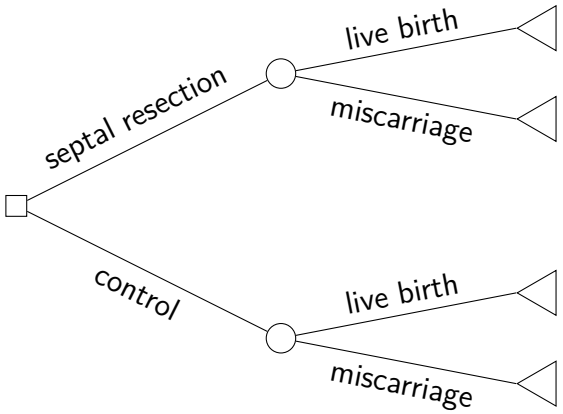


Figure 6.12: Decision tree model structure of septal resection to prevent miscarriage.

The decision tree model for this comparison was developed in R, a statistical computing program [142]. On entering the model, a woman with a history of infertility or miscarriage is diagnosed with a uterine septum. She would either have septal resection or receive no treatment.

The decision problem for the NHS is: Which treatment strategy should be adopted–hysteroscopic septal resection or no treatment?

After treatment, pathways for each arm are identical. The woman either conceives or does not. As pregnancy continues, the outcome for each woman is a birth, or a miscarriage. Birth would involve additional antenatal, intrapartum and postnatal care. Miscarriage could resolve spontaneously, or require expectant, medical or surgical treatment. QALYs are assigned to each outcome as in the previous model.

The model assumes no woman has another form of pregnancy loss, such as an ectopic, molar pregnancy or a stillbirth.

The time horizon chosen was the same as the progestogen study.

Probabilities

Parameter values and associated probability distributions are shown in Table 6.5. These values were derived from the best available data. Primarily, from meta-analysis of observational studies comparing the reproductive outcomes in women with known congenital uterine anomalies who had surgical treatment as the study group and women who had no surgical treatment as the control group. Costs were identified from NHS reference costs.

Table 6.5: Probability values of miscarriage with a uterine septum.

Parameter Description	*Mean value*	*95% confidence interval*	*Distribution*	*Source and type of data*
Baseline probability of miscarriage with a septum	0.33		Beta (α=108, β224)	Chan 2016 Random effects meta-analysis of observational studies
Odds Ratio with septal resection	0.20	0.09, 0.47	Log normal	Chan 2016 Random effects meta-analysis of observational studies
Odds Ratio with septal resection	0.91	0.54, 1.53	Log normal	Distribution fitted to experts' belief.

Effectiveness was derived from data extracted from a meta-analysis of hysteroscopic septal resection to increase the term live birth rate in women with infertility or previous miscarriages [159]. Other meta-analyses and reviews have been published, but include studies without a control group or which compare their results to historical controls or use women as their own controls

before and after surgery. I decided to limit the meta-analysis to these studies as these were the highest quality available.

The meta-analysis included four observational studies, in which 153 women underwent septal resection and 332 did not. All four studies showed a trend towards benefit, but had wide confidence intervals. The benefit was statistically significant three of the four studies. The results of this meta-analysis are shown in Figure 6.13

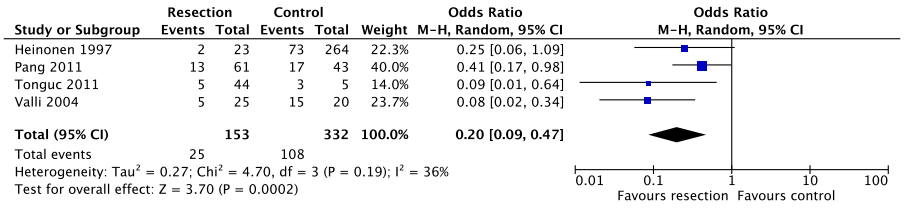


Figure 6.13: Meta-analysis of studies of hysteroscopic septal resection for the outcome of term birth.
The figure shows the odds ratio for each study as a square (with size proportional to the amount of information) and the horizontal line depicts the 95% CIs. The open diamond indicates the odds ratio (OR) with 95% CI overall. The vertical line, at the odds ratio of unity, corresponds to the line of no effect.

Consultation with the same expert panel of Professor Jim Thornton, Professor Arri Coomarasamy and myself, considered that septal resection was as likely as progestogen to be effective. A log normal distribution was applied to the odds ratio and used in the model.

Costs

Costs in this study are the same as Study 1 and shown in Table 6.4 and expressed in 2015 pounds sterling prices. I decided to use this most recent price year to make comparisons between the value of information for further research into septal resection and progestogen as treatments for miscarriage.

Septal resection can be performed in many ways as described in Chapter 1 of this thesis. No data were available specifically for this procedure, nonetheless NHS reference costs provide a cost code for resection or ablation procedure for intra-uterine lesions. This was £1044.08 in 2014-2015 [151]. This covers many types of hysteroscopic surgery including, transcervical resection of endometrium, submucosal fibroids, polyps and septa. The pre-operative and post-operative care for all of these procedures is similar. The model assumes

that no adjuncts are used after septal resection, such as anti-adhesion gels, hormone replacement therapy or intra-cavity balloons. Nonetheless, given the absence of any prospective randomised controlled trials of septal resection it would seem reasonable to accept there is no evidence for these adjuncts.

The model assumes spontaneous conception after septal resection. Reference costs and rates of activity are published by the NHS annually. Tables 6.3 shows costs and activity levels for pathways of care for each outcome branch of the decision tree, namely pregnancy care and miscarriage management.

Outcomes

Outcomes were measured in QALYs as the previous study.

Uncertainty

In the meta-analysis a random effects model was used and the uncertainty around the pooled means is reflected in the probability distribution. Parameters were considered independent a beta distribution was fitted to the baseline probability and a log normal distribution was fitted to the elicited odds ratio and used in the model.

Uncertainty in outcomes was investigated using a probabilistic sensitivity analysis of 100,000 replications. A scatter plot of the incremental cost and outcome for resection compared with no treatment for each simulation was prepared.

The cost-effectiveness acceptability curves (CEAC) was generated to reflect the probability that septal resection is cost effective compared with no treatment at various thresholds of willingness-to-pay for a QALY.

Value of information

Value of information calculations were based on 300 uterine septa per year. This estimate was obtained from the prevalence of uterine septa in women with a history of miscarriage estimated at 5% (95% CI 1.7–16.8) [20]

A willingness to pay threshold of £30,000 per QALY was used.

Estimated value of sample information (EVSI)

A nested Monte Carlo simulation approach was used the same as in Study 1.

6.3.2 Results

Point-estimate cost-effectiveness

The incremental cost effectiveness ratio (ICER), which is the cost for each additional QALY, using effectiveness data from the meta-analysis and expert opinion, were calculated from 100,000 simulations for each, were £868.28 and £1642.70 respectively.

Sensitivity analysis

Scatter plots of the incremental cost versus incremental benefit for 100,000 simulations are shown in Figures 6.14 and 6.15. Points are scattered in three quadrants of the cost-effectiveness plane using both priors. However, when using the meta-analysis, the majority of these points are scattered to the top-right, implying that resection is more likely to be effective, but have a higher cost. When using effectiveness data derived from expert opinion, the points are mainly scattered between the top right and bottom left quadrants, suggesting resection is either more expensive and more effective or cheaper, but less effective.

The cost-effectiveness acceptability curves in Figures 6.16 and 6.17, show how the probability of resection being cost-effective changes as the willingness to pay increases.

At a willingness to pay threshold of £30,000 per additional QALY, using the meta-analysis prior there is a 99.9% probability that resection is a cost-effective alternative to no treatment. Using the expert opinion prior this is 63.5%.

Value of information

There is uncertainty about the probability of the effectiveness of resection. This means that the decision is uncertain and there is a possibility it is wrong. The probability of being wrong multiplied by the loss associated with being wrong is the expected loss associated with the uncertainty. This is the same as the

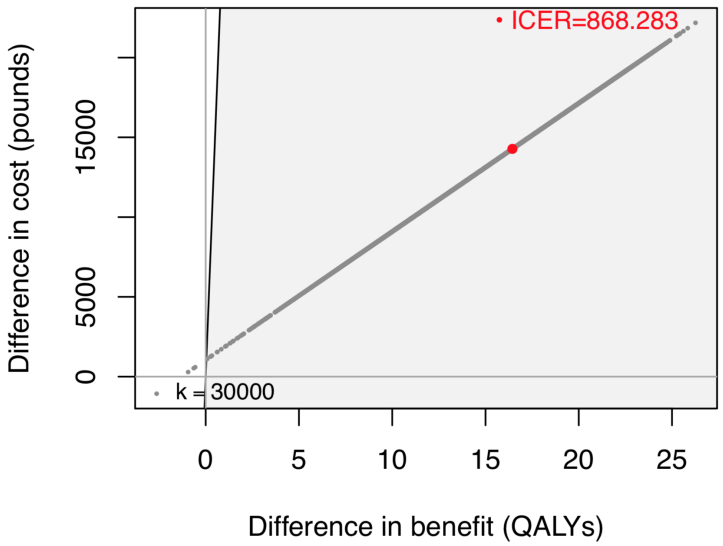


Figure 6.14: Cost-effectiveness plane for septal resection to prevent miscarriage using meta-analysis prior. Showing 100,000 simulations. Δ_c: Incremental costs, Δ_e: Incremental benefits.

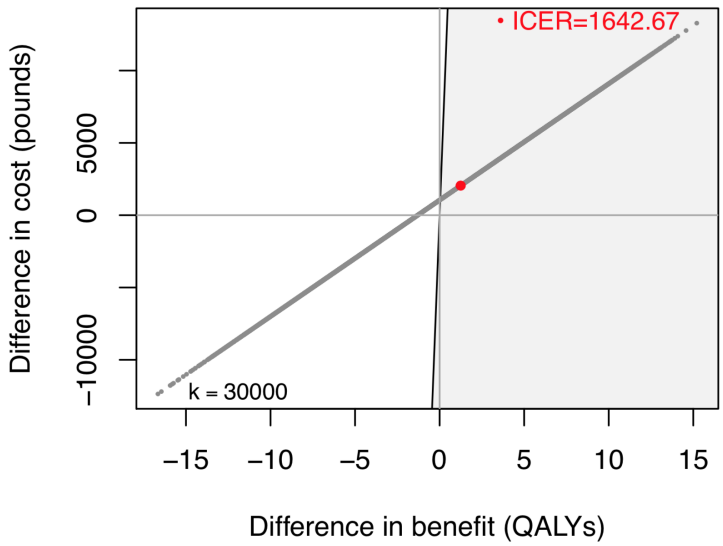


Figure 6.15: Cost-effectiveness plane for septal resection to prevent miscarriage using prior derived from expert opinion. Showing 100,000 simulations. Δ_c: Incremental costs, Δ_e: Incremental benefits.

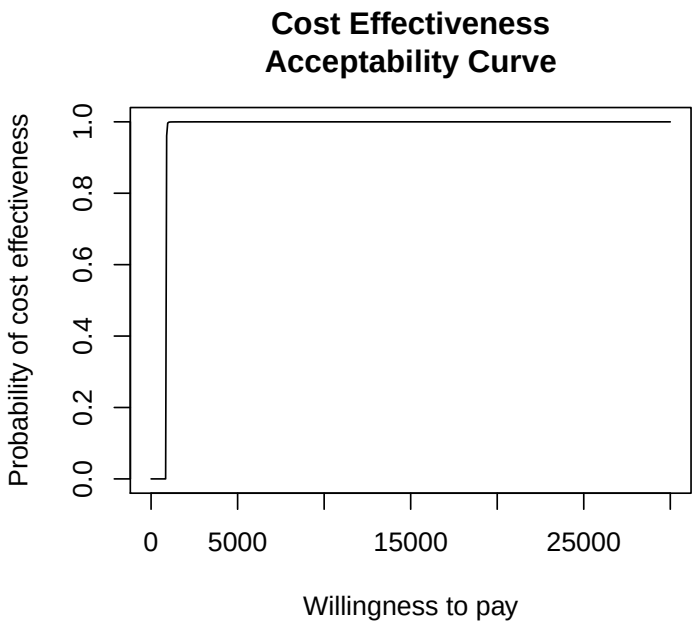


Figure 6.16: Cost-effectiveness acceptability curve and expected value of perfect information for septal resection to prevent miscarriage using meta-analysis prior.

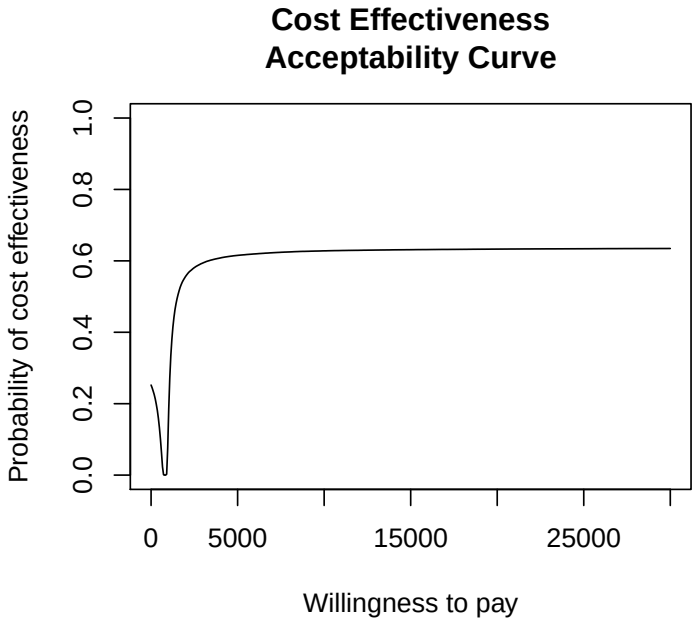


Figure 6.17: Cost-effectiveness acceptability curve and expected value of perfect information for septal resection to prevent miscarriage using expert opinion prior.

expected value of eliminating this uncertainty–the EVPI. Per patient EVPI is shown in Figures 6.18 and 6.19.

The population EVPI for the NHS at a willingness to pay of £30,000 per additional QALY was £3706.20 using the meta-analysis prior and £81m using expert opinion over a 10 year time horizon.

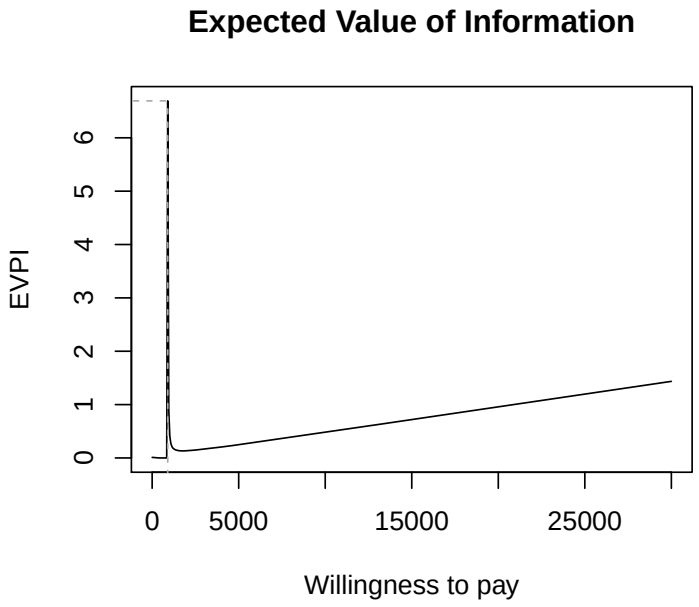


Figure 6.18: Expected value of perfect information (EVPI) for septal resection to prevent miscarriage using the meta-analysis prior. The dashed line shows the break even point, where the expected difference in benefit between strategies is zero, ie when the optimal decision changes from no treatment to septal resection.

The population EVPI is the per patient EVPI multiplied by the population who can benefit, in this case it was estimated to be 300 women per year. At a willingness to pay threshold of £30,000, using the meta-analysis effectiveness data, population EVPI is £3706.20. Using the prior derived from expert opinion the EVPI is £81m.

When willingness to pay is zero, the decision maker does not care about outcomes, only which strategy is cheaper. If there is uncertainty in cost there is potentially value in eliminating it. As the willingness to pay approaches the ICER, the EVPI peaks, then falls, before rising again to infinity. The explanation for this is the same as in Study 1, described in Section 6.2.2.

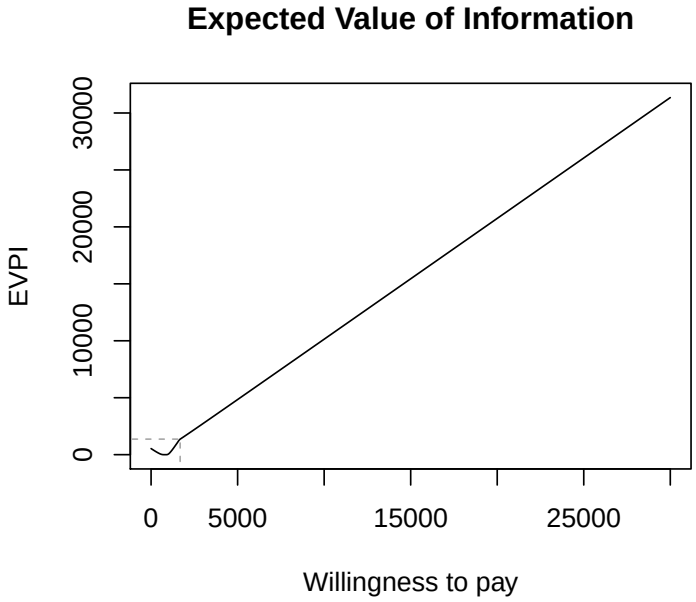


Figure 6.19: Expected value of perfect information (EVPI) for septal resection to prevent miscarriage using the prior derived from expert opinion. The dashed line shows the break even point, where the expected difference in benefit between strategies is zero, ie when the optimal decision changes from no treatment to septal resection.

Estimated Value of Sample Information

The estimated value of sample information for various sample sizes is shown in Figure 6.20. The EVSI using the meta-analysis prior, shown by the orange line, is practically £0 for all sample sizes. Evidence from the meta-analysis suggests that resection is so, and relatively certain to be, cost-effective, i.e. it is effective and cheap, that it is not worth conducting further research to eliminate any uncertainty.

However, the population EVSI calculated from the expert opinion prior, shown by the blue crosses, increases from £0 to approximately £80m for sample sizes over 50.

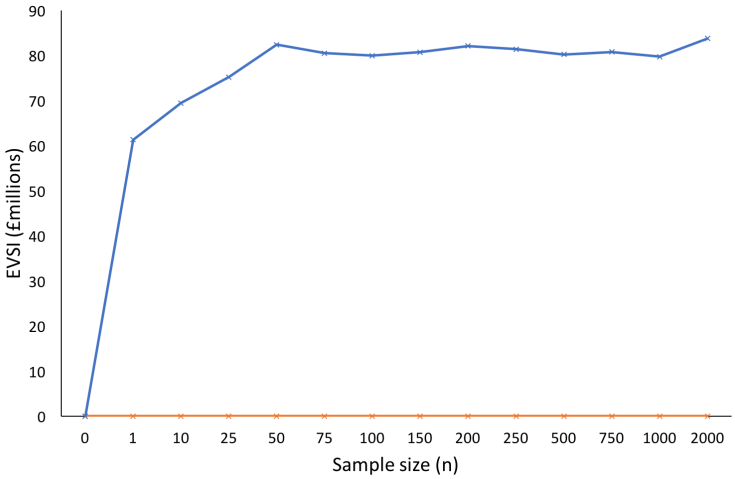


Figure 6.20: Estimated Value of Sample Information (EVSI) of a trial of septal resection to prevent miscarriage. Meta analysis prior (orange) and expert opinion prior (blue). The wobble in the line represent simulation noise.

6.3.3 Discussion

Main findings

This study estimated that there is no value in performing a trial when using effectiveness data from the meta-analysis prior. This is because the cost of resection and the population who can benefit is relatively small. When repeating the analysis using more uncertain effectiveness data the value of sample information is higher at around £80m for the sample size required by a definitive trial.

Strengths and limitations

This is the first economic evaluation and value of information analysis of septal resection to prevent miscarriage.

Efficacy data were limited to the best available evidence, excluding studies which used patients as their own controls. However, despite this the remaining studies remained observational and biased. No participants were randomised and there was no clear definition of what constituted a uterine septum. Moreover the four included studies are likely to be more biased than pre-PROMISE meta-analysis, which at least included controlled studies. Using a prior based on expert opinion may have counteracted the flaws in these studies, but no formal elicitation process was conducted and the experts are likely to be bias too.

The model is simplified, perhaps oversimplified, and does not include risk of surgical complications, changes in rates of preterm birth or the possibility of uterine rupture during pregnancy after septal resection. All of these exclusions would have additional costs to the NHS which were not included. Furthermore women with infertility may require assisted conception with or without resection which is another additional cost.

Septal resection is also performed for other indications than miscarriage. Subfertility and previous second-trimester miscarriage or preterm birth. Therefore the population who may benefit may be larger than the 300 per year estimated in this study.

Interpretation

Given the limitations, interpretation of these analyses must be cautious. The results suggest the costs of a study are less than the value of the information which would be obtained. On first glance, this would be a good use of resources. However it provides less return on investment when compared with other potential research studies, such as the PROMISE trial.

6.4 General conclusions

Studies based on existing evidence, namely meta-analyses of small biased studies, found that the trials of progestogen or septal resection to prevent miscarriage would not be good value for money. However when adjusting for bias, using expert opinion, the value of information is much higher.

This Chapter, questions not only where further research should be directed, but also the methodology of value of information analysis itself. On the one hand, when using flawed prior data, value of information analysis finds little value in further research, supporting the use of both interventions. However, we are in the privileged position to review this in hindsight, knowing that PROMISE showed progestogen doesn't work, demonstrating the flaws of value of information analysis. Only using expert opinion, which is also bias, and was conducted with the knowledge of the PROMISE results, provided inputs for value of information analysis which yield more useful results.

Putting these concerns aside and returning to the focus of this thesis, the feasibility of a septal resection trial, these economic evaluations and value of

information analyses predict that a trial of septal resection will provide less return on investment than other research studies. Therefore is must not be considered a priority for research funders.

Chapter 7

Miscarriage research prioritisation

7.1 Introduction

Traditionally the medical research agenda has been set by researchers or the pharmaceutical industry. However in recent years public and patient involvement in prioritisation of research questions and trial design has become increasingly essential. The National Institute for Health Research (NIHR) and the Medical Research Council (MRC) are the main funders of medical research in the UK. Both bodies mandate that grant applications have involvement from patients in the study design.

INVOLVE, a group established to promote public involvement in research state that involvement should be an "active partnership between the public and researchers in the research process, rather than the use of people as the subjects of research" [160]. INVOLVE are now incorporated into the National Institute for Health Research.

The James Lind Alliance (JLA), established in 2004 is a non-profit initiative with the aim of making healthcare research funders aware of the issues that matter most to patients and clinicians. The JLA whose infrastructure is now funded by NIHR, achieve this through priority setting partnerships which bring together patients, carers and clinicians to identify and prioritise the Top 10 uncertainties, or unanswered questions, about the effects of treatments [161]. The first priority setting partnership was on asthma [162] and the NIHR have since provided £1.2m funding for a study of benefits of breathing exercises as this was a top priority [163]. In women's health, priority setting partnerships have been conducted in preterm birth, stillbirth and most recently, endometrial

cancer [164–166]. No priority setting partnerships have been done for infertility or miscarriage.

Alongside the feasibility studies in this thesis I established the miscarriage priority setting partnership for women, those affected by miscarriage and healthcare professionals to collaboratively identify and prioritise important questions for miscarriage research. The overall aims of the priority setting partnership were much wider than those presented in this thesis which focuses on septal resection. Nonetheless, the James Lind Alliance methodology is accepted as the gold standard for public involvement in research prioritisation [167]. Furthermore, given the history outlined above, and close relationship with NIHR, if research into septal resection is found to be a high priority, there is more chance a definitive trial would be funded.

7.2 Aims

To identify whether a septal resection trial is a research priority for (1) women who have experienced, (2) those affected by miscarriage and (3) healthcare professionals.

7.3 Methods

Before establishing the miscarriage priority setting partnership consideration was given to the subject and scope of the partnership. At the start of my doctoral research, the agreed subject of my work was congenital uterine anomalies, specifically uterine septa. Uterine anomalies are uncommon, many are incidental findings diagnosed in conjunction with other conditions in decreasing strength of association: preterm birth, miscarriage and infertility. Previous priority setting partnership had focussed on individual health conditions, such as asthma, Parkinson's disease or schizophrenia. The first priority setting partnership in women's health prioritised research questions for preterm birth. Potential stakeholders for involvement in a priority setting partnership related my topic were divided by health condition. For example, there is no patient group for women with uterine anomalies, but the Miscarriage Association and Fertility Network UK exist for women with each respective condition. Advice from the James Lind Alliance was to avoid a priority setting

partnership with too broad or narrow scope to make the process achievable, avoid dilution of research findings between conditions and to attract funding for the project. Consideration was given to setting the scope to the following options: (1) (early) pregnancy loss (2) miscarriage (3) fertility or (4) fertility requiring assisted reproduction. Pregnancy loss was discounted, as the Stillbirth Priority Setting Partnership was due to complete, and limiting the scope to early pregnancy loss would exclude women with second-trimester loss. Uterine anomalies have a stronger association with miscarriage than fertility, so I decided to establish the Miscarriage Priority Setting Partnership.

Initiation

The first stakeholders were identified by a process of collective mind mapping between Mr Nick Raine-Fenning, Professor Jim Thornton and myself. We considered representation from the following groups: women who have experienced miscarriage; partners, family members, friends or colleagues of women who have had a miscarriage; and doctors, nurses and professionals allied to medicine with clinical experience of treating miscarriage. Organisations identified at this stage were sent an email and letter outlining the project. These included the Miscarriage Association, Tommys, Fertility Network UK, Ectopic Pregnancy Trust, Birthrights, Mumsnet, Netmums, Babyloss, the Association of Early Pregnancy Units, the Royal College of Obstetricians and Gynaecologists, the Royal College of Nursing, the British Fertility Society and the Royal College of General Practitioners. A follow up email and telephone call with interested organisations asked for their support and suggestions of other organisations or individuals who had not been approached in the initial round to ensure no relevant stakeholder was excluded. This process further identified Scottish Care & Information on Miscarriage, the Hughes Syndrome Foundation, International Society of Ultrasound in Obstetrics and Gynecology, the Society of Radiographers, the RCOG Women's Network and the Muslim Bereavement Support Service. All organisations were asked to become partners.

Partner organisations identified interested individuals to form a steering group. The priority setting partnership was overseen by a steering group whose members are listed in Table 7.1 and independently chaired by Leanne Metcalf, a James Lind Alliance Advisor. The group consisted of women affected by miscarriage, charities representing them, and healthcare professionals. The six stages of the PSP are outlined in Figure 7.1. The protocol was agreed

and published on the James Lind Alliance website before identification of unanswered questions began [168].

Finally, the steering group reviewed partner organisation and stakeholders already approached to ensure a broad spectrum of participants and to verify that all key stakeholders had been invited to be involved.

James Lind Alliance methodology excludes representatives of the pharmaceutical industry and non-clinical researchers as traditionally these group dominated the research agenda. Steering group members completed a declaration of interest form which was reviewed by the Chair to ensure no undue influence on the outcome.

Table 7.1: Miscarriage Priority Setting Partnership steering group members.

Name	*Representative*
Leanne Metcalf (Chair)	James Lind Alliance Facilitator
Matthew Prior (Coordinator)	Ph.D. Student
Professor Jim Thornton	Obstetrician and Researcher, University of Nottingham
Mr Nick Raine-Fenning	Gynaecologist and Researcher, University of Nottingham
Jane Brewin	CEO Tommy's Charity
Barbara Hepworth-Jones	Trustee of the Miscarriage Association
Judy Shakespeare	Retired General Practitioner
Julie Orford	Womens representative
Emily O'Toole	Womens representative
Sophie Willetts	Womens representative
Lucy Easthope	Womens representative
Carmel Bagness	Royal College of Nursing
Maureen Sharkey	Scottish Care & Information on Miscarriage
Arri Coomarasamy	Gynaecologist and Researcher, University of Birmingham
Kim Hinshaw	Clinician
Rachel Small	Chair of Association of Early Pregnancy Units
Lesley Regan	President of RCOG

Consultation

The steering group included women who had experience of miscarriage, charities which represented them, and clinicians. Some members representing charities or healthcare professionals, also had personal experience of pregnancy loss.

When defining the scope of the project at least four conflicting definitions for miscarriage were identified. Most miscarriages occur before 12 weeks gestation but the definition of early miscarriage varies between organisations.

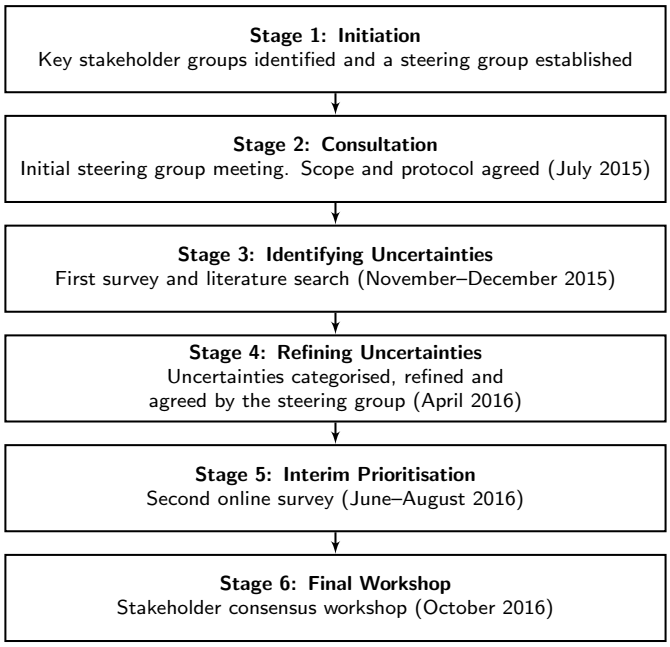


Figure 7.1: Summary of the James Lind Alliance Priority Setting process.

The European Society of Human Reproduction and Embryology (ESHRE) use "spontaneous pregnancy demise before 10 weeks of gestational age" [169]; the RCOG "before 10 completed weeks" [133]; and the National Institute for Healthcare and Clinical Excellence (NICE) "before 13 completed weeks" [137]. In the UK, the legal definition is spontaneous pregnancy loss "before 24 weeks' gestation" [170]. Patient representatives highlighted that many women would not consider a second trimester pregnancy loss to be a miscarriage.

The steering group debated how to define miscarriage for the purposes of the priority setting partnership. Some argued that almost all miscarriages were in the first-trimester and that this should be the main focus. Others argued that it was unlikely that another priority setting partnership would be formed to include late miscarriage. Women who had experienced miscarriage objected to the use of gestational limits, stating there are often discrepancies between the gestation the women thinks she is from her dates and what is measured on ultrasound scan. Furthermore, some women with a history of late pregnancy loss, approaching 24 weeks gestation took offence to it being labelled a miscarriage, as to them it felt like they lost a baby and went through labour. So it was agreed that miscarriage would be defined as anything from a positive pregnancy test to pregnancy loss before fetal survival was possible. Threatened miscarriage was also included. All aspects of miscarriage including conservative, medical and surgical management, psychological sequelae, treatments for

recurrent miscarriage, complementary therapies, nutrition and service design were included within the scope of the priority setting process. Other early pregnancy complications, ectopic pregnancy, gestational trophoblastic disease and termination of pregnancy were excluded.

Identifying questions

A survey was developed based on templates from previous priority setting partnerships [165, 171, 172]. The pilot survey consisted of an explanation of instructions for participants, guidance and example unanswered questions. Content validity was tested by asking a sample of 18 women affected by miscarriage, 3 partners of women affected, 2 family members friends or colleagues of women affected by miscarriage, and 5 health care professionals to complete the survey. Feedback from piloting resulting in changing the language to improve readability, making the explanation more concise to allow respondents to enter their questions more quickly, and limiting the number of responses to three. The full survey was hosted at www.miscarriagepsp.org and is included in Appendix 7. Survey distribution was carried out by partner organisations using their own networks, promotion in newsletters and conferences, printed flyers in clinics, on-line forums and social media.

A literature search was also performed of Database of Abstracts of Reviews of Effects (DARE), the Cochrane Library, Royal College of Obstetricians and Gynaecologists Guidelines, National Institute for Health and Care Excellence (NICE) Guidelines, and the Health Technology Assessment (HTA) Database to identify published research uncertainties falling within the scope of the priority setting partnership.

Refining questions

The raw questions were entered onto a spreadsheet (Microsoft Excel, 2010, Redmond, Washington, USA). The aim of this stage was to refine all questions gathered from the survey and literature review to a shorter list of summary questions. Firstly, I removed non-questions, statements or comments, and questions not related to miscarriage deemed being out of scope according to our definition. These were reviewed by the steering group to ensure in scope questions were not removed inadvertently.

I then reviewed all remaining questions and assigned them to categories. I did not identify any pre-existing taxonomy for miscarriage, so qualitatively developed categories based on consecutive submissions. Where a respondent included more than one question in the same submission, I duplicated the response and assigned it to each category. In total there were twenty categories, these are shown in Table 7.2 along with example submissions.

Each category was reviewed by at least one healthcare professional and one patient representative from the steering group. Duplicate questions were removed and questions were combined and rephrased to create summary questions. The full list of summary questions was checked and agreed by the steering group.

In addition to the survey, I searched the Cochrane Library from inception until May 2016 for systematic reviews of interventions for miscarriage. All summary questions were checked against the evidence base to determine if they were true questions or if the survey respondent was unaware there was already evidence to answer their question.

Interim prioritisation

Summary questions were ranked by women and those affected by miscarriage and healthcare professionals via an on-line survey. The survey displayed in a randomised sequence all summary questions and asked respondents to choose up to ten which they thought were most important. This second survey was sent to participants identified by partner organisations and advertised using newsletters, on-line and social media. Participants were asked to read the list of questions and identify their ten priorities. Questions were ranked based on the frequency they had been chosen. Analysis of the top priorities for patients and healthcare professionals; and women from different ethnic origins was performed to ensure the overall rankings represented all groups. The top 25 questions were taken to the final workshop.

Final workshop

Participants were invited to the workshop, representing women and partners affected by miscarriage, including representatives from charities, and healthcare professionals. Before the meeting, participants completed a declaration of interest form and individually prioritised the top 25 questions presented in

Table 7.2: Taxonomy for miscarriage questions.

Category	*Total*	*Example submissions*
Association	7	Is there a link with infertility? Are there links between miscarriage and premature birth?, Can miscarriage cause insomnia?
Cause generic	528	What causes miscarriage? Why does it happen?, Why do they occur?, What causes miscarriage early in pregnancy?, What causes recurrent miscarriage?, What causes miscarriage after 12 weeks?
Cause specific	279	Was my age a factor? Does cervical treatment for CIN (excisional or ablative) increase risk of miscarriage? Does stress increase the chance of miscarriage? What proportion of miscarriages are due to lethal genetic anomalies?
Diagnosis	99	How can an early miscarriage be detected? If I'm bleeding does this mean a definite miscarriage? Do regular scans in first 12 weeks really help women who have had recurrent miscarriages?
Embryo Pain	1	At what stage do embryos feel anything similar to pain?
Follow up	8	How soon after a miscarriage can you try again? How should we advise women following a miscarriage in order to optimise the chance of a successful pregnancy in the future?
Incidence	31	What is the incidence of miscarriage in women over 30? How common are they really?
Investigation	168	Why in the UK they don't check the length of a cervix? What investigations are of true clinical value?,
Management	85	Is there anything I can do to help the miscarriage happen quickly? Use of ultrasound guidance to make surgical management of miscarriage safe.
Pathology	31	What exactly is happening when miscarriage is occurring? Why might the amniotic sac break at only 16 weeks pregnant?
Prevention	334	Does aspirin help prevent miscarriage? How can you prevent recurrent miscarriages?, What can patients do to prevent miscarriages? Do early scans reduce the risk of miscarriage?
Prognosis	128	Am I more likely to have another miscarriage after the first one? What is my risk of miscarrying again taking into account my age, past medical history, current medication, weight, diet and lifestyle?
Risk factors	66	Why are some couples affected by miscarriage and others not? Do lifestyle treatments reduce the risk of miscarriage in women with PCOS? What impact father age or lifestyle can affect the chances of miscarriage?
Screening	11	Can we predict which woman is prone to a first miscarriage? How could you be screened prior to pregnancy to prevent chromosome issues?
Social	27	Why is it such a taboo subject? What employer support could be available for parents grieving due to miscarriage?
Support	382	Research into the psychological effects and how this may affect ability to conceive in future? What is the most appropriate way to support women (and partners) after a miscarriage to reduce psychological harm?, How long do women typically grieve following a miscarriage?, Why are people so insensitive particularly staff?
Systems	89	What is the optimal configuration of early pregnancy loss service (in terms of physical outcomes, service user satisfaction, staff satisfaction)? Why am I put on a ward with people having an abortion?, How much evidence is there to suggest hospital is the better environment after a miscarriage?
Terminology	13	Why hospitals call it the product of conception? Why is your baby now referred to as a product?, Why do doctors refer to Miscarriage as "spontaneous abortion?
Treatment	112	How uNKiller cells should be treated? The role of uterine septum resection in women with recurrent miscarriage and septate uterus?
Twins	3	How come when pregnant with twins, I miscarried one, but other twin was fine? Why isn't there better training for hospital staff for a twin/multiple pregnancy with only one surviving baby?

random order labelled A-Y. This allowed them to attend the workshop with an idea of their own top ten. The workshop was chaired by Katherine Cowan, an independent James Lind Alliance Advisor with no previous personal or professional experience of miscarriage. Ground rules were agreed at the start of the meeting to ensure confidentiality of individual's responses and respecting others views and compromise. This process intended to prevent the domination of the discussions by a single member, encourage quieter members of the group to speak and highlight that consensus meant people were unlikely to leave the meeting with all their views being represented in the final top 10. Participants were divided into three groups of mixed representation of women and those affected by miscarriage and healthcare professionals. Each group, chaired by a JLA facilitator was asked to prioritise all 25 questions which were printed on A4 cards. Each card had the interim prioritisation for each group (women and those affected by miscarriage, and professionals) written on the back, to assist with decision making. The three sets of rankings were combined into one list. Participants were then divided into three different groups, again with mixed representation. In the second round, the groups were presented with the combined prioritisation from round one. Participants were asked to focus on whether the correct questions were in the top 10 and had the right prioritisation. In the final round, all three groups' priorities were combined and presented to the entire group. The focus was on the top 10. When disagreement arose, consensus was reached by raised-hands voting. The top ten research priorities in miscarriage were agreed.

7.4 Results

The process undertaken and results are shown in Figure 7.2.

Consultation and identifying questions

In total, 1093 respondents (932 women who have experienced miscarriage, 8 partners, 17 family members, friends or colleagues, 104 healthcare professionals and 8 charitable organisations) completed the survey between 2 November 2015 to 31 December 2015. Overall 3343 questions were collected (3279 questions via the survey and 64 by literature search).

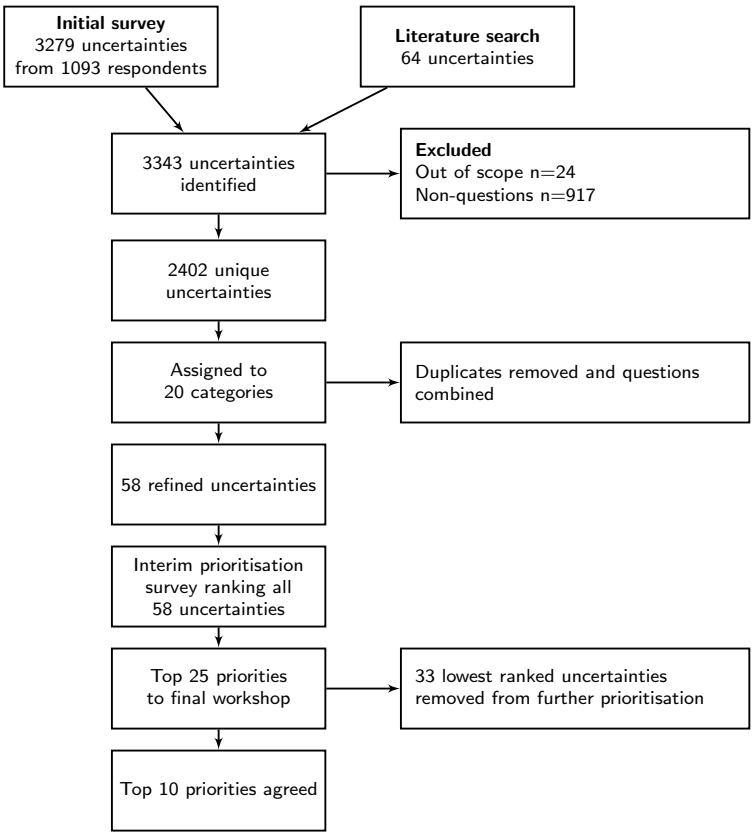


Figure 7.2: Flow diagram of results.

Refining questions

There were 2402 questions remaining after submissions that were either non-questions or not related to miscarriage were excluded. These were assigned to one of twenty categories created shown in Table 7.2. These were refined to 58 summary questions by at least one patient representative and one clinician from steering group. The literature search found no answered questions in miscarriage, so no questions were resolved.

Interim prioritisation

The interim prioritisation survey was completed by 2122 respondents (1797 women who have experienced miscarriage, 22 partners, 68 family members, friends or colleagues, 185 healthcare professionals and 31 people whose experience of miscarriage was unclear) between 6 June 2016 and 1 August 2016.

There was an imbalance in response between those who had experienced miscarriage and healthcare professionals. There was also an ethnic imbalance (1942 white British/Irish (95%) and 79 black, Asian, and minority ethnic (BAME) (3.9%), 18 prefer not to say). Considering this, the overall rankings were compared with those of each subgroup, lay and healthcare professional; white and ethnic minority, to ensure all perspectives were reflected.

The highest 25 overall ranked questions included the following number of questions from each groups top 25: lay 24, health care professional 18, white 25, and ethnic minority 21. No question in the top 8, for health care professionals or the top 18 for ethnic minorities were excluded from the final workshop. This analysis is demonstrated in Table 7.3. Given this analysis the steering group agreed to take the 25 highest overall ranked questions to the final workshop.

Final workshop and top ten priorities.

The one day workshop took place on 14 October 2016 at The Studio in Birmingham. Participants included eleven women who had experienced miscarriage, two male partners of women affected by miscarriage and eight healthcare professionals. The top ten to research priorities for miscarriage were agreed and shown in Table 7.4.

Table 7.3: Analysis by demographic (Q Ref refers to lettered questions in Appendix 8; Priorities in bold indicate questions in the group demographic top 25, but not taken to the final workshop). Lay: includes women who have experienced and those affected by miscarriage. HCP: Healthcare Professionals, BAME: Black, Asian, and minority ethnic.

All Respondents		***Lay***		***HCP***		***White***		***BAME***	
Priority	*Q Ref*	*Q Ref*	*Priority*	*Q Ref*	*Priority*	*Q Ref*	*Priority*	*Q Ref*	*Priority*
1st	AB	AB	1	AB	1	AB	1	AB	1
2nd	V	D	4	V	2	V	2	S	7
3rd	A	A	3	S	7	A	3	A	3
4th	D	V	2	E	11	D	4	V	2
5th	AN	AN	5	W	24	AN	5	AF	6
6th	AF	AH	9	K	12	AF	6	E	11
7th	S	AF	6	B	8	B	8	AG	15
8th	B	AM	10	Y	16	S	7	D	4
9th	AH	S	7	**X**	**33**	AH	9	K	12
10th	AM	B	8	AG	15	AM	10	I	20
11th	E	E	11	**C**	**26**	E	11	Y	16
12th	K	K	12	A	3	K	12	AH	9
13th	AI	AI	13	I	20	AI	13	AL	21
14th	AJ	AJ	14	**AV**	**45**	AJ	14	B	8
15th	AG	F	18	AC	17	F	18	AP	19
16th	Y	AG	15	**AE**	**28**	AG	15	H	23
17th	AC	AC	17	**P**	**48**	Y	16	AC	17
18th	F	Y	16	AF	6	AC	17	AN	5
19th	AP	AP	19	AL	21	AP	19	**AV**	**45**
20th	I	Z	22	AN	5	Z	22	W	24
21st	AL	AL	21	H	23	AL	21	**X**	**33**
22nd	Z	I	20	**AS**	**29**	I	20	AJ	14
23rd	H	H	23	AP	19	H	23	N	25
24th	W	N	25	**BF**	**53**	N	25	**P**	**48**
25th	N	**BC**	**27**	D	4	W	24	**R**	**58**
26th	C	AS	29	M	30	AS	29	BC	27
27th	BC	AE	28	BD	31	AE	28	BD	31
28th	AE	C	26	AJ	14	C	26	C	26
29th	AS	W	24	N	25	BC	27	AD	40
30th	M	M	30	BC	27	M	30	AI	13
31st	BD	BD	31	AM	10	AK	32	AM	10
32nd	AK	AK	32	AQ	34	BD	31	AT	37
33rd	X	BE	35	AK	32	X	33	M	30
34th	AQ	X	33	AT	37	AQ	34	Z	22
35th	BE	AQ	34	L	38	BE	35	AE	28
36th	AY	AY	36	G	41	AY	36	U	39
37th	AT	AT	37	AI	13	AT	37	AK	32
38th	L	U	39	BA	54	U	39	F	18
39th	U	L	38	BB	46	L	38	BB	46
40th	AD	AD	40	AH	9	G	41	L	38
41st	G	G	41	AW	56	AD	40	AR	47
42nd	O	O	42	U	39	O	42	AS	29
43rd	T	T	43	Z	22	T	43	AZ	44
44th	AZ	AZ	44	AR	47	AZ	44	BF	53
45th	AV	BB	46	AY	36	BB	46	G	41
46th	BB	AR	47	AZ	44	AV	45	J	50
47th	AR	AU	49	F	18	AR	47	T	43
48th	P	J	50	R	58	AU	49	AA	55
49th	AU	AO	52	AX	51	J	50	AY	36
50th	J	AV	45	O	42	P	48	O	42
51st	AX	AX	51	AD	40	AO	52	AX	51
52nd	AO	P	48	BE	35	AX	51	BE	35
53rd	BF	BF	53	T	43	BF	53	AQ	34
54th	BA	Q	57	J	50	AW	56	AW	56
55th	AA	AA	55	AA	55	BA	54	Q	57
56th	AW	AW	56	AU	49	Q	57	AO	52
57th	Q	BA	54	Q	57	AA	55	BA	54
58th	R	R	58	AO	52	R	58	AU	49

Table 7.4: Top 10 ranked research questions for miscarriage.

Priority	*Research question*
1	What are the effective interventions to prevent miscarriage, threatened miscarriage and recurrent miscarriage? (e.g. lifestyle, vitamins, aspirin, early scans, HCG, dopamine agonists, progestogen, polytherapy, steroids, oestrogen, metformin, anticoagulants, intravenous immunoglobulin, intralipid and anti-tnf alpha)
2	What are the emotional and mental health impacts of miscarriage in the short and long term for the mother and the partner?
3	What investigations are of true clinical value? (e.g. ultrasound, gene sequencing, natural killer cells, thromboelastography (TEG), microarray testing of the fetus, paternal investigations, plasminogen activator inhibitor polymorphism)
4	To what extent do pre-existing medical conditions cause miscarriage? (e.g. vitamin deficiencies, diabetes, previous infertility, endometriosis, polycystic ovarian syndrome, menstrual irregularities, cervical factors, uterine anomalies, polyps, immunological factors or previous pregnancy complications e.g. caesarean section or preterm birth)
5	What types of emotional support are effective in preventing or treating women or their partners after a miscarriage?
6	Do lifestyle factors (diet, stress, exercise, weight, alcohol, sexual activity, smoking, night shifts or flying) cause miscarriage?
7	To what extent do genetic and chromosomal abnormalities in the fetus cause miscarriage?
8	What preconception tests or interventions prevent miscarriage? (e.g vitamin supplements, folic acid, Dehydroepiandrosterone (DHEA), co-enzyme Q-10 or bariatric surgery)
9	What are the appropriate investigations for women after one, two, or three or more miscarriages?
10	What male factors contribute towards the cause of miscarriage?

Publicity

The results were presented at the Association of Early Pregnancy Units Annual Scientific Meeting in November 2017. Findings were disseminated to funders of miscarriage research and shared by partner organisations through their own networks. The James Lind Alliance disseminated the findings and publicised descriptive reports of the process itself. The results were published in an academic paper from which this chapter is based [173].

Septal resection

The top priority overall, and for women and healthcare professionals was "what are the effective interventions to prevent miscarriage, threatened miscarriage and recurrent miscarriage?". Septal resection being an intervention to prevent miscarriage, could therefore be considered a top priority. Nonetheless, this question was also included as a separate summary question: "Is surgery safe and effective for women with uterine anomalies? (septal resection)". This question was overall ranked 56th out of 58 questions in the interim survey. Women and those affected by miscarriage also ranked it 56th overall, where as healthcare professionals ranked it 41st.

7.5 Discussion

This priority setting partnership has established the top ten research questions for miscarriage. Women and health care professionals collaboratively agreed that research into effective interventions to prevent miscarriage, threatened miscarriage and recurrent miscarriage is the top priority for research. Septal resection is an intervention intended to prevent miscarriage and so can be considered a top research priority. Nonetheless, when asking women and healthcare professionals specifically about septal resection it was considered a lesser priority.

This is the first time a collaborative approach between patients and healthcare professionals has been taken to prioritise unanswered questions in miscarriage for research. The response to both surveys was considerable, although due to the methods of survey distribution I was unable to calculate a response rate. Nonetheless the response was comparable with other priority setting partnerships shown in Table 7.5.

Previous James Lind Alliance priority setting partnerships have produced a top 10 list of uncertainties that focus on treatment [161]. To consider a treatment uncertainty to be answered, requires an up-to-date systematic review. In the Miscarriage Priority Setting Partnership after reviewing the highest level of evidence, Cochrane systematic reviews, no questions were considered answered. Nonetheless, recent priority setting partnerships, including this one, considered questions broader than treatment uncertainties. For example, "what are the emotional and mental health impacts of miscarriage in the short and long term for the mother and her partner?". Clearly this question cannot be answered by a systematic review of treatments, but this is not considered in James Lind Alliance methodology.

While the refining process enabled all original submissions to be mapped to a summary question, it results in dilution of overall meaning and clarity. Many specific questions were merged, creating vague summary questions, which are less helpful for researchers to design focussed studies or for funding bodies to determine which is actually most important. For example, the top priority could lead to trials designed to compare interventions as wide ranging as medications such as aspirin, steroids, intralipid; surgery such or septal resection; or lifestyle interventions; but does not state which is most important. It could be argued it is not the role of patients to prioritise which should be investigated first, because they have less understanding of the underlying pathology. However this goes against the ethos of patient and public involvement which encourages active engagement throughout the research process [160].

Septal resection can be considered the top, or a lesser priority depending on how the findings are interpreted. This highlights a limitation of the refining process where some questions are merged and other left unchanged, or an expression by women that surgical intervention is considered differently to less invasive interventions. Women's perception of surgery as an intervention, when examples included in the top priority list are non-surgical, may be that surgery is an intervention associated with higher risk and therefore are more willing to prioritise research into interventions with fewer perceived risks. Women's perception were explored in the initial consultation finding that some women did not consider a second-trimester loss as a miscarriage, however, there are no documented discussions from the final workshop exploring women's and healthcare professionals perceptions of what constitutes an intervention.

I believe the reason why septal resection was considered separate to other

Table 7.5: Response rate to James Lind Alliance (JLA) priority setting partnerships (PSP).

JLA PSP	*Number of Survey respondents*	*Number of questions submitted*
Bipolar	3,285	14,398
Depression	3,000	10,000+
Sight loss and vision	2,220	4,461
Dementia	1,563	4,116
Autism	1,213	3,331
Miscarriage	**1093**	**3279**
Diabetes (Type 1)	583	1,141
Multiple Sclerosis	507	1,084
Vitiligo	461	1,427
Lyme Disease	253	967
Stroke in Scotland	106	548

interventions is because it is an intervention specific to a defined and limited population, women with a uterine septum. Conversely, lifestyle interventions, vitamins and some proposed medications considered top priority for research can be applied to more women. The priority setting partnership process is bias towards more common conditions, because more patients are likely to be affected, and hence submit it as a question or rank it as a priority. However, this can be at the peril of basic science. For example, it is unlikely blanket use of medications across large numbers of women at risk of miscarriage are likely to be effective, as the cause of most miscarriages is likely aneuploidy or lethal genetic abnormalities [129, 131].

The RCOG previously highlighted septal resection as an area for further research [133]. This priority setting partnership process has confirmed it is also important to women and healthcare professionals when considering miscarriage. However, miscarriage is just part of the spectrum of poor reproductive outcome. Uterine septa are also associated with infertility, miscarriage and preterm birth. Comparing the results of this priority setting partnership with those of the preterm birth unsurprisingly shows some overlap and the top priority is essentially the same for both [164]. This is shown in Table 7.6. Interestingly the preterm birth partnership does not specifically mention uterine anomalies or septal resection in its results, supporting the conclusion that it should be considered a top priority for research. To complete this spectrum of prioritisation a fertility priority setting partnership is required. I would anticipate

there to be considerable agreement with the priorities of the miscarriage and preterm birth priority setting partnerships. So for women with a uterine septum and history of infertility, miscarriage or preterm birth, a trial of septal resection would be the top research priority.

Table 7.6: Themes from the miscarriage and preterm birth priority setting partnerships.

Theme	*Priority*
Intervention	Miscarriage
	1. What are the effective interventions to prevent miscarriage, threatened miscarriage and recurrent miscarriage? (e.g. lifestyle, vitamins, aspirin, early scans, HCG, dopamine agonists, progestogen, polytherapy, steroids, oestrogen, metformin, anticoagulants, intravenous immunoglobulin, intralipid and anti-tnf alpha)
	Preterm Birth
	1. Which interventions are most effective to predict or prevent preterm birth?
Support	Miscarriage
	2. What are the emotional and mental health impacts of miscarriage in the short and long term for the mother and her partner?
	5. What types of emotional support are effective in preventing or treating women or their partners after a miscarriage?
	Preterm Birth
	5. What should be included in packages of care to support parents and families / carers when a premature baby is discharged from hospital?
	9. What emotional and practical support improves attachment and bonding, and does the provision of such support improve outcomes for premature babies and their families?
	10. How can staff support women and their partners in subsequent pregnancies, using a holistic approach, to reduce anxiety, stress and any associated increased visits to healthcare settings?

Conclusion

The results of the Miscarriage Priority Setting Partnership suggests further research into the safety and efficacy of septal resection is a research priority for women who have experienced miscarriage, those affected by miscarriage and healthcare professionals. Therefore, a case could be made for funding a septal resection trial, but consideration must also be given to the operational difficulties and value of information in conducting such a trial.

Chapter 8

Conclusions and further work

The aim of this thesis was to determine the feasibility of a randomised controlled trial of hysteroscopic septal resection for women with uterine septum and a history of infertility or miscarriage. The hypothesis for this work was that such a trial would be feasible. Given the results of the studies presented in this thesis I conclude that a definitive randomised controlled trial is not feasible at present.

A summary of research findings for each chapter are presented in Tables 8.1 to 8.6.

Table 8.1: Summary of research findings: Chapter 2

Aim	• To assess the intra- and inter-rater reliability of three-dimensional ultrasound for diagnosis of congenital uterine anomalies, specifically uterine septa.
Population	• Three-dimensional uterine volumes.
Intervention / Test	• Classification of uterine morphology between raters.
Outcome	• Intra- and inter-rater reliability.
Study design	• Off line analysis of three-dimensional ultrasound volumes. 1. Between experts using their classification of choice. 2. Uterine measurements taken by the same rater six weeks apart. 3. Uterine measurements taken by two raters.
Summary of results	• Diagnosis between expert raters using multiple classification systems showed moderate reliability. Uterine measurements are highly reproducible with cavity indentation showing the highest intra- and inter-rater reliability. Three-dimensional ultrasound is a reliable diagnositic tool for diagnosing a uterine septum providing raters use a consistent classification system based on standardised measurements of uterine morphology.

Table 8.2: Summary of research findings: Chapter 3

Aims	• To establish the minimum clinically important difference for septal resection. • To calculate the sample size required to statistically demonstrate the minimum clinically important difference.
Population	• Women with a history of poor reproductive outcome, defined as one or more of the following; infertility of over 12 months, a history of early, later or recurrent miscarriage or a previous preterm birth. • Lay participants with no adverse pregnancy history. • Clinicians who care for women with adverse reproductive outcomes.
Intervention / Test	• Survey to calculate the minimum clinically important difference for hysteroscopic septal resection. • Sample size calculation using the established minimum clinically important difference.
Outcome	• The minimum clinically important difference for hysteroscopic septal resection. • Sample size for a definitive trial.
Study design	• Survey of views on the minimum clinically important difference required to undergo septal resection.
Summary of results	• The median minimum clinically important difference for women with a history of poor reproductive outcome was a change from 30% risk of future miscarriage to 10% (IQR = 9–20) after hysteroscopic septal resection. • The sample size required to detect a minimum clinically important difference relevant to half of women and lay people, and all healthcare professionals to undergo hysteroscopic septal resection to prevent miscarriage is 118, (59 in each arm) where α=0.05 and (1-β)=0.9.

Table 8.3: Summary of research findings: Chapter 4

Aim	• To determine how many women require screening with three-dimensional ultrasound to identify eligible participants.
Population	• Women with a history of infertility or one or more previous miscarriages.
Intervention / Test	• Screening for uterine septa using three-dimensional ultrasound and eligibility criteria for pilot SEPTUM trial.
Outcome	• Number of women requiring screening to identify eligibile participants.
Study design	• All women referred to Nurture fertility, or the recurent miscarriage clinic underwent three-dimensional ultrasound to identify uterine septa. Women with suspected anomalies on two-dimensional scan were referred for three-dimensional ultrasound. Eligibilty criteria were then applied.
Summary of results	• To identify one eligible participant, 464 women needed screening. Divided by patient population, identifying one eligible participant screening required 407 women with infertility, or 47 women with recurrent miscarriage to undergo three-dimensional ultrasound. Approximately one septum was identified per 1500 women who had a two-dimensional scan in the early pregnancy unit.

Table 8.4: Summary of research findings: Chapter 5

Aims	• To determine the willingness of clinicians to support a septal resection trial. • To assess the reasons why eligible women chose to participate, or not, in the pilot septal resection trial. • To explore the potential for multi-centre recruitment to a septal resection trial.
Population	• Clinicians who care for women with uterine septa. • Eligible women who declined enrolment into the pilot septum trial. • Sites whom originally expressed an interest in participating in the trial.
Intervention / Test	• Assessment of clinicians, women and units views of a septal resection trial.
Outcome	• Views on septal resection, and a such a trial.
Study design	• Mixed methods including surveys and informal qualitative methods.
Summary of results	• Only a minority of clinicians (12 of 46) who perform septal resection and women eligible (2 of 13) to participate, were willing to participate in a septal resection trial. • One of eleven interested units went on to become a trial site.

Table 8.5: Summary of research findings: Chapter 6

Aims	• Calculate the value of information analysis for an area of uncertainty, progestogen to prevent recurrent miscarriage, using data prior to a recently funded large trial. • Calculate the value of information which could be gained by a trial of septal resection. • Compare the value of information of both trials to see if the health economics suggest a septal resection trial is worthwhile.
Population	• Probability of effectiveness data from meta-analyses of previous studies. • Cost data from NHS reference costs.
Intervention / Test	• Decision tree models.
Outcomes	• Incremental cost effectiveness ratio (ICER) • Estimated value of perfect information (EVPI) • Estimated value of sample information (EVSI)
Study design	• Health economic and value of information analysis of decision tree models.
Summary of results	• Using existing evidence, meta-analyses of small biased studies, predicted no value of information of further research in either area. • Using expert opinion with greater undertainty, the expected value of sample information for a trial of progestogen was £1.6B and of resection £80m. • Other trials are likely a higher priority.

Table 8.6: Summary of research findings: Chapter 7

Aim	• To identify whether a septal resection trial is a research priority.
Population	• Miscarriage research uncertainties.
Intervention / Test	• Prioritisation of research questions for miscarriage by women who have experienced miscarriage, those affected by miscarriage and healthcare professionals.
Outcome	• Top 10 ranked research questions for miscarriage.
Study design	• James Lind Alliance Priority Setting Partnership.
Summary of results	• Women and health care professionals collaboratively agreed that research into effective interventions to prevent miscarriage, threatened miscarriage and recurrent miscarriage is the top priority for research.

8.1 Strengths and limitations

The main strength of this research is that it comprehensively covers the majority of the issues which need addressing to demonstrate feasibility. Limitations of this thesis is that it does not address the standard deviation of the outcome measure, live birth after septal resection to address sample size. However, I used methods to estimate the minimum clinically important difference to calculate the sample size.

I also did not address follow up, or rates of compliance with the protocol, in part due to the small number of women randomised and the time scale of this thesis. Unlike studies of fertility treatments where clinical pregnancy and live birth rates are known reasonably quickly after the intervention, septal resection requires a longer follow up period, especially in women who are treated for subfertility. Therefore, in the timescale for this work, and that only six women were randomised it would not be possible to answer this question.

A main limitation to recruiting other sites was exclusion from the NIHR portfolio of studies. One could question why I did not have the foresight to see this obstacle, but the original intention was to recruit enough women in Nottingham to demonstrate feasibility to obtain funding for a definitive trial which would be eligible to be part of the NIHR portfolio of studies. It is difficult to assess if inclusion would change the outcome of this thesis. Five sites stated this was a reason for non-participation, but this may have been an easy way to decline participation and if enough enthusiasm existed within the department it would be possible to participate in the trial.

Value of information analysis is a relatively new concept for research prioritisation in women's health. The models used in this thesis were simple, but demonstrate, that if poor quality biased data is inputted, namely the meta-analyses of small studies, then there is no value in further research. However the PROMISE trial, contrary to previous studies showed progestogen is not effective and was actually very good value. Accounting for biased data using experts opinion may provide better inputs for value of information analyses to prioritise future research projects.

I believe I have demonstrated tenacity to pursue this randomised controlled trial, when facing many challenges. However, there comes a point when a decision needs to be taken regarding continuation of the study. The funding for the pilot study has run out but only six women have been randomised.

While I would like to continue the trial, recruitment will remain difficult and further funding is unlikely.

A randomised controlled trial is by no doubt the definitive method to answer the clinical question of the effectiveness of septal resection for women with a history of infertility, miscarriage or preterm birth and a uterine septum. Nonetheless, as this thesis concludes it is infeasible. Therefore it is worth considering alternative methods to advance our knowledge of uterine anomalies and septal resection.

8.2 Future research

Moving forward from this thesis, further research can be considered in three areas: (1) reproductive prognosis of uterine anomalies and further refinement of classification systems; (2) development research questions based on the priority setting partnership findings; and (3) improving feasibility of a septal resection trial.

Debate continues regarding the classification of uterine anomalies, demonstrated by the publication of two different systems by ESHRE and ASRM, the largest societies of reproductive medicine clinicians in Europe and the United States respectively. Nonetheless, neither system incorporates prognostic data when developing the system, instead attaching labels to subjectively chosen morphometric cut offs. Current evidence on the effect of uterine anomalies on reproductive outcome is limited by the use of different, subjective and unreliable classification systems. An alternative approach would be registering women with uterine anomalies and collecting three-dimensional ultrasound volumes and recording subsequent reproductive outcomes in a database. This methodology would allow independent analysis of uterine measurements which could provide more accurate data on prevalence and effect on reproductive outcome. This could lead to the development of a classification system based with prognostic information. Furthermore, a database would collect prospective data from women with uterine septa who do not wish to be randomised to a clinical trial allowing researchers to analyse if certain septum characteristics, such as dimensions or constitution, have negative prognostic value.

The Miscarriage Priority Setting Partnership presented in this thesis signposts areas for researchers and funding bodies to prioritise. Nonetheless while priority setting partnerships are not intended to devise focussed research questions,

the top ten priorities are wide ranging, so further work is required develop focussed research questions. NIHR consider a good research question to be important to the NHS and its patients, supported by current evidence, of high scientific quality, feasible, timely, clear and well-defined and represent good value for public money [174]. So researchers, in collaboration with funding bodies, patients and healthcare professionals, must break each priority down to develop research questions in a PICO format to design studies which can be funded.

Thirdly, to improve feasibility of a septal resection trial, collaboration with other groups is essential. This could include collaboration with the Dutch TRUST trial, or working with funders to develop incentives for UK centres to participate in a trial of septal resection. Incentives may include being part of the NIHR CRN portfolio of studies, or setting targets for number of women screened, rather than targets for recruitment. This approach was adopted by the TABLET (Thyroid AntiBodies and LEvoThyroxine) trial which went from missing its recruitment targets to easily recruiting participants just by changing the screening pathway. In slightly different circumstances, with external support, a trial of septal resection may still be feasible to complete.

References

[1] Jadad AR and Enkin MW. Bias in randomized controlled trials. In *Randomized Controlled Trials*, pages 29–47. Blackwell Publishing Ltd, 2008. doi:10.1002/9780470691922.ch3.

[2] Heneghan C, Spencer EA, Bobrovitz N et al. Lack of evidence for interventions offered in uk fertility centres. *BMJ*, 2016. 355. doi:10.1136/bmj.i6295.

[3] Sadler T. Urogenital system. In *Langman's Medical Embryology*, pages 232–259. Wolters Kluwer, 12th edition, 2012.

[4] Acién M and Acién P. Normal embryological development of the female genital tract. In *Female genital tract congenital malformations : classification, diagnosis and management*, pages 5–6. Springer, 2015. doi:10.1007/978-1-4471-5146-3.

[5] Acién P, Acién M, and Sánchez-Ferrer M. Müllerian anomalies "without a classification": from the didelphys-unicollis uterus to the bicervical uterus with or without septate vagina. *Fertil Steril*, 2009. 91: 2369–2375. doi:10.1016/j.fertnstert.2008.01.079.

[6] Acién P. Embryological observations on the female genital tract. *Hum Reprod*, 1992. 7: 437–45.

[7] Grimbizis GF and Campo R. Congenital malformations of the female genital tract: the need for a new classification system. *Fertil Steril*, 2010. 94: 401–407. doi:10.1016/j.fertnstert.2010.02.030.

[8] Buttram Jr VC and Gibbons WE. Mullerian anomalies: a proposed classification. (An analysis of 144 cases). *Fertil Steril*, 1979. 32: 40–46.

[9] The American Fertility Society. The American Fertility Society classifications of adnexal adhesions, distal tubal occlusion, tubal

occlusion secondary to tubal ligation, tubal pregnancies, Müllerian anomalies and intrauterine adhesions. *Fertil Steril*, 1988. 49: 944 – 955. doi:http://dx.doi.org/10.1016/S0015-0282(16)59942-7.

[10] Salim R, Woelfer B, Backos M et al. Reproducibility of three-dimensional ultrasound diagnosis of congenital uterine anomalies. *Ultrasound Obstet Gynecol*, 2003. 21: 578–582. doi:10.1002/uog.127.

[11] Troiano RN and McCarthy SM. Mullerian duct anomalies: imaging and clinical issues. *Radiology*, 2004. 233: 19–34. doi:10.1148/radiol.2331020777.

[12] Oppelt P, Renner SP, Brucker S et al. The VCUAM (Vagina Cervix Uterus Adnex-associated Malformation) classification: a new classification for genital malformations. *Fertil Steril*, 2005. 84: 1493–1497. doi:10.1016/j.fertnstert.2005.05.036.

[13] Bermejo C, Martinez Ten P, Cantarero R et al. Three-dimensional ultrasound in the diagnosis of Mullerian duct anomalies and concordance with magnetic resonance imaging. *Ultrasound Obstet Gynecol*, 2010. 35: 593–601. doi:10.1002/uog.7551.

[14] Acién P and Acién MI. The history of female genital tract malformation classifications and proposal of an updated system. *Hum Reprod Update*, 2011. 17: 693–705. doi:10.1093/humupd/dmr021.

[15] Grimbizis GF, Gordts S, Di Spiezio Sardo A et al. The ESHRE/ESGE consensus on the classification of female genital tract congenital anomalies. *Hum Reprod*, 2013. 28: 2032–2044. doi:10.1093/humrep/det098.

[16] Pfeifer S, Butts S, Dumesic D et al. Uterine septum: a guideline. *Fertil Steril*, 2016. 106: 530–540. doi:10.1016/j.fertnstert.2016.05.014.

[17] Grimbizis GF, Di Spiezio Sardo A, Saravelos SH et al. The Thessaloniki ESHRE/ESGE consensus on diagnosis of female genital anomalies. *Hum Reprod*, 2016. 31: 2–7. doi:10.1093/humrep/dev264.

[18] Ludwin A, Ludwin I, Pitynski K et al. Reply: are the ESHRE/ESGE criteria of female genital anomalies for diagnosis of septate uterus appropriate? *Hum Reprod*, 2014. 29: 868–869. doi:10.1093/humrep/deu001.

[19] Ludwin A and Ludwin I. Comparison of the ESHRE-ESGE and ASRM classifications of Müllerian duct anomalies in everyday practice. *Hum Reprod*, 2015. 30: 569–80.

[20] Chan YY, Jayaprakasan K, Zamora J et al. The prevalence of congenital uterine anomalies in unselected and high-risk populations: a systematic review. *Hum Reprod Update*, 2011. 17: 761–771. doi:10.1093/humupd/dmr028.

[21] Charney P. The Gynecologic Examination. In *Practical Gynecology: A Guide for the Primary Care Physician*, pages 4–6. American College of Physicians, 2009.

[22] Padilla LA, Radosevich DM, and Milad MP. Limitations of the pelvic examination for evaluation of the female pelvic organs. *Int J Gynaecol Obstet*, 2005. 88: 84–88. doi:10.1016/j.ijgo.2004.09.015.

[23] Cantuaria GHC, Angioli R, Frost L et al. Comparison of Bimanual Examination With Ultrasound Examination Before Hysterectomy for Uterine Leiomyoma. *Obstetrics & Gynecology*, 1998. 92: 109–112. doi:http://dx.doi.org/10.1016/S0029-7844(98)00131-8.

[24] Woelfer B, Salim R, Banerjee S et al. Reproductive outcomes in women with congenital uterine anomalies detected by three-dimensional ultrasound screening. *Obstet Gynecol*, 2001. 98: 1099–1103. doi:10.1016/S0029-7844(01)01599-X.

[25] National Institute for Health and Clinical Excellence. Fertility problems: assessment and treatment. NICE guideline (CG156), 2013.

[26] Ayida G, Kennedy S, Barlow D et al. A comparison of patient tolerance of hysterosalpingo-contrast sonography (hycosy) with echovist-200 and x-ray hysterosalpingography for outpatient investigation of infertile women. *Ultrasound Obstet Gynecol*, 1996. 7: 201–204. doi:10.1046/j.1469-0705.1996.07030201.x.

[27] Pellerito JS, McCarthy SM, Doyle MB et al. Diagnosis of uterine anomalies: relative accuracy of MR imaging, endovaginal sonography, and hysterosalpingography. *Radiology*, 1992. 183: 795–800. doi:10.1148/radiology.183.3.1584936.

[28] Nicolini U, Bellotti M, Bonazzi B et al. Can ultrasound be used to screen uterine malformations? *Fertil Steril*, 1987. 47: 89–93.

[29] Jurkovic D, Geipel A, Gruboeck K et al. Three-dimensional ultrasound for the assessment of uterine anatomy and detection of congenital anomalies: a comparison with hysterosalpingography and two-dimensional sonography. *Ultrasound Obstet Gynecol*, 1995. 5: 233–237. doi:10.1046/j.1469-0705.1995.05040233.x.

[30] Van Schoubroeck D, Installe A, Raine-Fenning NJ et al. OP11.01: Interobserver variability in the ultrasound diagnosis of congenital uterine anomalies. *Ultrasound Obstet Gynecol*, 2012. 40: 87. doi:10.1002/uog.11490.

[31] Saravelos SH and Li TC. Intra- and inter-observer variability of uterine measurements with three-dimensional ultrasound and implications for clinical practice. *Reprod Biomed Online*, 2015. 31: 557–564. doi:10.1016/j.rbmo.2015.06.024.

[32] McRobbie DW. *MRI from Picture to Proton*, page 138. Cambridge University Press, 2003.

[33] Marten K, Vosshenrich R, Funke M et al. MRI in the evaluation of müllerian duct anomalies. *Clinical Imaging*, 2003. 27: 346–350. doi:10.1016/s0899-7071(02)00587-9.

[34] Altman D. In *Practical Statistics for Medical Research*. London: Chapman and Hall, 1991.

[35] Ludwin A, Pityński K, Ludwin I et al. Two- and Three-Dimensional Ultrasonography and Sonohysterography versus Hysteroscopy With Laparoscopy in the Differential Diagnosis of Septate, Bicornuate, and Arcuate Uteri. *Journal of Minimally Invasive Gynecology*, 2013. 20: 90–99. doi:10.1016/j.jmig.2012.09.011.

[36] Moini A, Mohammadi S, Hosseini R et al. Accuracy of 3-dimensional sonography for diagnosis and classification of congenital uterine anomalies. *Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine*, 2013. 32: 923–7. doi:10.7863/ultra.32.6.923.

[37] Homer HA, Li TC, and Cooke ID. The septate uterus: a review of management and reproductive outcome. *Fertil Steril*, 2000. 73: 1–14.

[38] Simón C, Martinez L, Pardo F et al. Müllerian defects in women with normal reproductive outcome. *Fertil Steril*, 1991. 56: 1192–1193. doi:10.1016/S0015-0282(16)54741-4.

[39] Raga F, Bauset C, Remohi J et al. Reproductive impact of congenital Mullerian anomalies. *Hum Reprod*, 1997. 12: 2277–2281. doi:10.1093/humrep/12.10.2277.

[40] Dreisler E and Stampe Sorensen S. Mullerian duct anomalies diagnosed by saline contrast sonohysterography: prevalence in a general population. *Fertil Steril*, 2014. 102: 525–529. doi:10.1016/j.fertnstert.2014.04.043.

[41] Clifford K, Rai R, Watson H et al. An informative protocol for the investigation of recurrent miscarriage: preliminary experience of 500 consecutive cases. *Hum Reprod*, 1994. 9: 1328–1332. doi:10.1093/oxfordjournals.humrep.a138703.

[42] Acién P. Incidence of Mullerian defects in fertile and infertile women. *Hum Reprod*, 1997. 12: 1372–1376. doi:10.1093/oxfordjournals.humrep.a019588.

[43] Makino T, Hara T, Oka C et al. Survey of 1120 Japanese women with a history of recurrent spontaneous abortions. *Eur J Obstet Gynecol Reprod Biol.*, 1992. 44: 123–30.

[44] Guimarães Filho HA, Mattar R, Pires CR et al. Comparison of hysterosalpingography, hysterosonography and hysteroscopy in evaluation of the uterine cavity in patients with recurrent pregnancy losses. *Archives of Gynecology and Obstetrics*, 2006. 274: 284–288. doi:10.1007/s00404-006-0186-3.

[45] Jayaprakasan K, Chan YY, Sur S et al. Prevalence of uterine anomalies and their impact on early pregnancy in women conceiving after assisted reproduction treatment. *Ultrasound Obstet Gynecol*, 2011. 37: 727–732. doi:10.1002/uog.8968.

[46] Acién P. Reproductive performance of women with uterine malformations. *Hum Reprod*, 1993. 8: 122–126.

[47] Chan YY, Jayaprakasan K, Tan A et al. Reproductive outcomes in women with congenital uterine anomalies: a systematic review. *Ultrasound Obstet Gynecol*, 2011. 38: 371–382. doi:10.1002/uog.10056.

[48] Sorensen SS and Trauelsen AG. Obstetric implications of minor mullerian anomalies in oligomenorrheic women. *Am J Obstet Gynecol*, 1987. 156: 1112–1118.

[49] Zupi E and Solima E. Uterine anomalies prevalence and reproductive outcome in women undergoing diagnostic hysteroscopy. *Gynaecol Endoscopy*, 1996. 5: 147–150.

[50] Shuiqing M, Xuming B, and Jinghe L. Pregnancy and its outcome in women with malformed uterus. *Chin Med Sci J*, 2002. 17: 242–245.

[51] Zlopasa G, Skrablin S, Kalafatic D et al. Uterine anomalies and pregnancy outcome following resectoscope metroplasty. *Int J Gynaecol Obstet*, 2007. 98: 129–133. doi:10.1016/j.ijgo.2007.04.022.

[52] Zhang Y, Zhao YY, and Qiao J. Obstetric outcome of women with uterine anomalies in China. *Chin Med J (Engl)*, 2010. 123: 418–422.

[53] Maneschi F, Zupi E, Marconi D et al. Hysteroscopically detected asymptomatic mullerian anomalies. Prevalence and reproductive implications. *J Reprod Med*, 1995. 40: 684–688.

[54] Venetis CA, Papadopoulos SP, Campo R et al. Clinical implications of congenital uterine anomalies: a meta-analysis of comparative studies. *Reprod Biomed Online*, 2014. 29: 665–683. doi:10.1016/j.rbmo.2014.09.006.

[55] Fayez JA. Comparison between abdominal and hysteroscopic metroplasty. *Obstet Gynecol*, 1986. 68: 399–403.

[56] Dabirashrafi H, Bahadori M, Mohammad K et al. Septate uterus: new idea on the histologic features of the septum in this abnormal uterus. *Am J Obstet Gynecol*, 1995. 172: 105–107.

[57] Sadek SM, Ahmad RA, and Soliman BS. Three dimensional color Doppler transvaginal ultrasound morphologic features of uterine septum and residual cavity in women with reproductive failure. *Middle East Fertility Society Journal*, 2014. doi:10.1016/j.mefs.2014.03.006.

[58] Fedele L, Bianchi S, Gruft L et al. Danazol versus a gonadotropin-releasing hormone agonist as preoperative preparation for hysteroscopic metroplasty. *Fertil Steril*, 1996. 65: 186–188.

[59] Sparac V, Kupesic S, Ilijas M et al. Histologic Architecture and Vascularization of Hysteroscopically Excised Intrauterine Septa. *J Am Assoc Gynecol Laparosc*, 2001. 8: 111–116. doi:10.1016/s1074-3804(05)60559-2.

[60] Raga F, Casan EM, and Bonilla-Musoles F. Expression of vascular endothelial growth factor receptors in the endometrium of septate uterus. *Fertil Steril*, 2009. 92: 1085–1090. doi:10.1016/j.fertnstert.2008.07.1768.

[61] Salim R, Regan L, Woelfer B et al. A comparative study of the morphology of congenital uterine anomalies in women with and without a history of recurrent first trimester miscarriage. *Hum Reprod*, 2003. 18: 162–166.

[62] Jones Jr HW and Jones GE. Double uterus as an etiological factor in repeated abortion: indications for surgical repair. *Am J Obstet Gynecol*, 1953. 65: 325–339.

[63] Tompkins P. Comments on the bicornuate uterus and twinning. *Surgical Clinics of North America*, 1962. 42: 1049 – 1062. doi:https://doi.org/10.1016/S0039-6109(16)36746-9.

[64] Strassmann EO. Fertility and unification of double uterus. *Fertil Steril*, 1966. 17: 165–176.

[65] Ruge P. Einen Fall von Schwangerschaft bei uterus septus. *Z Gebuntshilfe Gynakol*, 1884. 10: 141.

[66] Valle RF and Ekpo GE. Hysteroscopic metroplasty for the septate uterus: review and meta-analysis. *J Minim Invasive Gynecol*, 2013. 20: 22–42. doi:10.1016/j.jmig.2012.09.010.

[67] Valli E, Vaquero E, Lazzarin N et al. Hysteroscopic metroplasty improves gestational outcome in women with recurrent spontaneous abortion. *J Am Assoc Gynecol Laparosc*, 2004. 11: 240–244. doi:10.1016/s1074-3804(05)60206-x.

[68] Colacurci N, De Placido G, Perino A et al. Hysteroscopic Metroplasty. *J Am Assoc Gynecol Laparosc*, 1998. 5: 171–174.

[69] Roy KK, Singla S, Baruah J et al. Reproductive outcome following hysteroscopic septal resection in patients with infertility and recurrent abortions. *Arch Gynecol Obstet*, 2011. 283: 273–279. doi:10.1007/s00404-009-1336-1.

[70] Paradisi R, Barzanti R, Natali F et al. Metroplasty in a large population of women with septate uterus. *J Minim Invasive Gynecol*, 2011. 18: 449–454. doi:10.1016/j.jmig.2011.03.017.

[71] Tehraninejad ES, Ghaffari F, Jahangiri N et al. Reproductive outcome following hysteroscopic monopolar metroplasty: an analysis of 203 cases. *Int J Fertil Steril*, 2013. 7: 175–180.

[72] Ayas S, Gurbuz A, and Tuna G. Hysteroscopic resection of uterine septum improves reproductive performance in women with unexplained infertility. *Journal of Medical Sciences*, 2011. 41: 595–601. doi:10.3906/sag-1006-897.

[73] Patel A, Goudar S, Geller S et al. Drape estimation vs. visual assessment for estimating postpartum hemorrhage. *International Journal of Gynecology & Obstetrics*, 2006. 93: 220–224. doi:10.1016/j.ijgo.2006.02.014.

[74] Royal College of Obstetricians and Gynaecologists (RCOG). Diagnostic Hysteroscopy Under General Anaesthesia. Consent Advice No.1, 2008.

[75] Litta P, Spiller E, Saccardi C et al. Resectoscope or Versapoint for hysteroscopic metroplasty. *International Journal of Gynecology & Obstetrics*, 2008. 101: 39–42. doi:10.1016/j.ijgo.2007.10.013.

[76] Pai HD, Kundnani MT, Palshetkar NP et al. Reproductive performance after hysteroscopic metroplasty in women with primary infertility and septate uterus. *J Gynecol Endosc Surg*, 2009. 1: 17–20. doi:10.4103/0974-1216.51904.

[77] Abbas AM, Abdallah M, and Michael A. *Middle East Fertility Society Journal*, 2017 in press. doi:10.1016/j.mefs.2017.02.002.

[78] Kasapoglu T, Kasapoglu D, and Deren O. Successful management of the recurrent uterine rupture after the uterine septum resection. *Case Reports in Women's Health*, 2015. 8: 13–16. doi:10.1016/j.crwh.2015.10.001.

[79] Angell NF, Tan Domingo J, and Siddiqi N. Uterine rupture at term after uncomplicated hysteroscopic metroplasty. *Obstetrics and gynecology*, 2002. 100: 1098–9.

[80] Kowalik CR, Goddijn M, Emanuel MH et al. Metroplasty versus expectant management for women with recurrent miscarriage and a septate uterus. *Cochrane Database Syst Rev*, 2011. page CD008576. doi:10.1002/14651858.CD008576.pub3.

[81] Lin K, Zhu X, Xu H et al. Reproductive outcome following resectoscope metroplasty in women having a complete uterine septum with double cervix and vagina. *International Journal of Gynecology & Obstetrics*, 2009. 105: 25–28. doi:http://dx.doi.org/10.1016/j.ijgo.2008.11.018.

[82] Heinonen PK. Reproductive performance of women with uterine anomalies after abdominal or hysteroscopic metroplasty or no surgical treatment. *J Am Assoc Gynecol Laparosc*, 1997. 4: 311–317.

[83] Pang LH, Li MJ, Li M et al. Not every subseptate uterus requires surgical correction to reduce poor reproductive outcome. *Int J Gynaecol Obstet*, 2011. 115: 260–263. doi:10.1016/j.ijgo.2011.07.030.

[84] Tonguc EA, Var T, and Batioglu S. Hysteroscopic metroplasty in patients with a uterine septum and otherwise unexplained infertility. *Int J Gynaecol Obstet*, 2011. 113: 128–130. doi:10.1016/j.ijgo.2010.11.023.

[85] Cararach M, Penella J, Ubeda A et al. Hysteroscopic incision of the septate uterus: scissors versus resectoscope. *Hum Reprod*, 1994. 9: 87. doi:10.1093/oxfordjournals.humrep.a138326.

[86] Colacurci N, De Franciscis P, Mollo A et al. Small-diameter hysteroscopy with Versapoint versus resectoscopy with a unipolar knife for the treatment of septate uterus: A prospective randomized study. *Journal of Minimally Invasive Gynecology*, 2007. 14: 622–627. doi:10.1016/j.jmig.2007.04.010.

[87] Candiani GB, Vercellini P, Fedele L et al. Repair of the uterine cavity after hysteroscopic septal incision. *Fertil Steril*, 1990. 54: 991–994. doi:10.1016/S0015-0282(16)53992-2.

[88] Yang JH, Chen MJ, Chen CD et al. Optimal waiting period for subsequent fertility treatment after various hysteroscopic surgeries. *Fertil Steril*, 2013. 99: 2092–2096.e3. doi:10.1016/j.fertnstert.2013.01.137.

[89] Ludwin A, Ludwin I, Kudla M et al. Diagnostic accuracy of three-dimensional sonohysterography compared with office hysteroscopy and its interrater/intrarater agreement in uterine cavity assessment after hysteroscopic metroplasty. *Fertil Steril*, 2014. 101: 1392–1399. doi:10.1016/j.fertnstert.2014.01.039.

[90] Ludwin A, Ludwin I, Pitynski K et al. Role of morphologic characteristics of the uterine septum in the prediction and prevention of abnormal healing outcomes after hysteroscopic metroplasty. *Hum Reprod*, 2014. doi:10.1093/humrep/deu110.

[91] Kha KS and Chien PF. Evaluation of a clinical test. I: Assessment of reliability. *BJOG: An International Journal of Obstetrics and Gynaecology*, 2001. 108: 562–567. doi:10.1111/j.1471-0528.2001.00150.x.

[92] Nederlands Trial Register [Internet]. Academic Medical Center (AMC), Amsterdam, 2009 Feb 18 -. Identifier NTR1676, The Randomised Uterine Septum Transsection Trial; [1 page]. Availiable from: `http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=1676` [Accessed: 27 August 2017].

[93] Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics*, 1977. 33: 159–174.

[94] Kramer MS and Feinstein AR. Clinical biostatistics. LIV. The biostatistics of concordance. *Clinical pharmacology and therapeutics*, 1981. 29: 111–23.

[95] Bland JM and Altman D. Statistical Methods For Assessing Agreement Between Two Methods Of Clinical Measurement. *The Lancet*, 1986. 327: 307–310. doi:http://dx.doi.org/10.1016/S0140-6736(86)90837-8.

[96] Walter SD, Eliasziw M, and Donner A. Sample size and optimal designs for reliability studies. *Statistics in Medicine*, 1998. 17: 101–110. doi:10.1002/(SICI)1097-0258(19980115)17:1<101::AID-SIM727>3.0.CO;2-E.

[97] Martins WP, Raine-Fenning NJ, Leite SP et al. A standardized measurement technique may improve the reliability of measurements of endometrial thickness and volume. *Ultrasound Obstet Gynecol*. 38: 107–115. doi:10.1002/uog.9016.

[98] Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Controlled clinical trials*, 1981. 2: 93–113.

[99] Dupont WD and Plummer Jr WD. Power and sample size calculations: A review and computer program. *Controlled clinical trials*, 1990. 11: 116–128. doi:http://dx.doi.org/10.1016/0197-2456(90)90005-M.

[100] Jaeschke R, Singer J, and Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled clinical trials*, 1989. 10: 407–15.

[101] Copay AG, Subach BR, Glassman SD et al. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*, 2007. 7: 541–546. doi:10.1016/j.spinee.2007.01.008.

[102] Carragee EJ and Cheng I. Minimum acceptable outcomes after lumbar spinal fusion. *Spine J*, 2010. 10: 313–320. doi:10.1016/j.spinee.2010.02.001.

[103] Current Controlled Trials [Internet]. Nottingham University Hospitals NHS Trust, 2008 Jun 27 - 2011 Mar 24. Identifier ISRCTN37730443 , Pilot study for obstetric cholestasis trial; [1 page]. Availiable from: `http://www.isrctn.com/ISRCTN37730443` [Accessed: 21 March 2018].

[104] Chappell LC, Gurung V, Seed PT et al. Ursodeoxycholic acid versus placebo, and early term delivery versus expectant management, in women with intrahepatic cholestasis of pregnancy: semifactorial randomised clinical trial. *BMJ*, 2012. 344: e3799. doi:10.1136/bmj.e3799.

[105] Current Controlled Trials [Internet]. University of Oxford, 2015 Aug 26 -. Identifier ISRCTN91918806 , A trial investigating pregnancy and infant outcomes in women with intrahepatic cholestasis of pregnancy treated with ursodeoxycholic acid or placebo.; [1 page]. Availiable from: http://www.isrctn.com/ISRCTN91918806 [Accessed: 21 March 2018].

[106] Glujovsky D, Sueldo CE, Borghi C et al. Misleading reporting and interpretation of results in major infertility journals. *Fertil Steril*, 2016. doi:10.1016/j.fertnstert.2015.12.134.

[107] Jones TL, Baxter MA, and Khanduja V. A quick guide to survey research. *Ann R Coll Surg Engl*, 2013. 95: 5–7. doi:10.1308/003588413X13511609956372.

[108] Pocock S. *Clinical Trials: A Practical Approach*, pages 124–125. Wiley, 1983.

[109] Coomarasamy A, Williams H, Truchanowicz E et al. A Randomized Trial of Progesterone in Women with Recurrent Miscarriages. *N Engl J Med*, 2015. 373: 2141–2148. doi:10.1056/NEJMoa1504927.

[110] Duley L, Gülmezoglu AM, Henderson-Smart DJ et al. Magnesium sulphate and other anticonvulsants for women with pre-eclampsia. *Cochrane Database of Systematic Reviews*, 2010. doi:10.1002/14651858.CD000025.pub2.

[111] Duley L, Henderson-Smart DJ, Meher S et al. Antiplatelet agents for preventing pre-eclampsia and its complications. *Cochrane Database of Systematic Reviews*, 2007. doi:10.1002/14651858.CD004659.pub2.

[112] Duffy JM, Arambage K, Correa FJ et al. Laparoscopic surgery for endometriosis. *Cochrane Database of Systematic Reviews*, 2014. doi:10.1002/14651858.CD011031.pub2.

[113] Nurture Fertility. About us, 2017. http://www.nurturefertility.co.uk/about-us/ [Accessed: 29 January 2017].

[114] Nottingham University Hospitals NHS Trust. About the Trust, 2017. https://www.nuh.nhs.uk/about-us/ [Accessed: 29 January 2017].

[115] National Institute for Health and Care Excellence. Interventional procedure overview of hysteroscopic metroplasty of a uterine septum for primary infertility. Interventional procedures guidance [IPG509], 2015.

[116] Rittenberg V, Seshadri S, Sunkara SK et al. Effect of body mass index on IVF treatment outcome: an updated systematic review and meta-analysis. *Reproductive biomedicine online*, 2011. 23: 421–39. doi:10.1016/j.rbmo.2011.06.018.

[117] Balen PAH and Anderson RA. Impact of Obesity on female reproductive health: British Fertility Society, Policy and Practice Guidelines. *Human Fertility*, 2007. 10: 195–206. doi:10.1080/14647270701731290.

[118] Wartolowska K, Collins GS, Hopewell S et al. Feasibility of surgical randomised controlled trials with a placebo arm: a systematic review. *BMJ Open*, 2016. 6. doi:10.1136/bmjopen-2015-010194.

[119] Coomarasamy A. Surveys [one of women with recurrent miscarriage (n = 88) and the other of gynaecologists (n = 102) treating women with recurrent miscarriage], 2008. Unpublished data.

[120] Current Controlled Trials [Internet]. Birmingham Clinical Trials Unit (BCTU), 2015 Feb 11 -. Identifier ISRCTN14163439 , PRISM: Progesterone in spontaneous miscarriage; [1 page]. Availiable from: `http://www.isrctn.com/ISRCTN14163439` [Accessed: 27 August 2017].

[121] Hendrix M, Van Horck M, Moreta D et al. Why women do not accept randomisation for place of birth: Feasibility of a RCT in the Netherlands. *BJOG: An International Journal of Obstetrics and Gynaecology*, 2009. 116: 537–542. doi:10.1111/j.1471-0528.2008.02103.x.

[122] Likert R. A technique for the measurement of attitudes. *Archives of Psychology*, 1932. 22 140: 55. doi:2731047.

[123] National Institute for Health and Care Excellence. Interventional procedure overview of hysteroscopic metroplasty of a uterine septum for recurrent miscarriage. Interventional procedures guidance [IPG510], 2015.

[124] National Institute for Health Research. NIHR Clinical Research Network Portfolio, 2017. https://www.nihr.ac.uk/research-and-impact/nihr-clinical-research-network-portfolio/ [Accessed: 18 October 2017].

[125] Wilson E. A practical guide to value of information analysis. *PharmacoEconomics*, 2015. 33: 105–121. doi:10.1007/s40273-014-0219-x.

[126] Kline J, Stein Z, and Susser M. *Conception to birth: epidemiology of prenatal development*, volume 14. New York New York Oxford University Press 1989., 1989.

[127] Wilcox AJ, Weinberg CR, O'Connor JF et al. Incidence of Early Loss of Pregnancy. *New England Journal of Medicine*, 1988. 319: 189–194. doi:10.1056/NEJM198807283190401.

[128] Rai R and Regan L. Recurrent miscarriage. *The Lancet*, 2006. 368: 601–611. doi:10.1016/S0140-6736(06)69204-0.

[129] Stephenson MD, Awartani KA, and Robinson WP. Cytogenetic analysis of miscarriages from couples with recurrent miscarriage: a case-control study. *Hum Reprod*, 2002. 17: 446–451.

[130] Sullivan AE, Silver RM, LaCoursiere DY et al. Recurrent fetal aneuploidy and recurrent miscarriage. *Obstet Gynecol*, 2004. 104: 784–788. doi:10.1097/01.AOG.0000137832.86727.e2.

[131] Daher S, Mattar R, Gueuvoghlanian-Silva BY et al. Genetic Polymorphisms and Recurrent Spontaneous Abortions: An Overview of Current Knowledge. *American Journal of Reproductive Immunology*, 2012. 67: 341–347. doi:10.1111/j.1600-0897.2012.01123.x.

[132] Lee C and Slade P. Miscarriage as a traumatic event: A review of the literature and new implications for intervention. *Journal of Psychosomatic Research*, 1996. 40: 235–244. doi:10.1016/0022-3999(95)00579-X.

[133] Royal College of Obstetricians and Gynaecologists. Green Top Guideline No 17: The investigation and treatment of couples with recurrent first-trimester and second-trimester miscarriage, 2011.

[134] Practice Committee of the American Society for Reproductive Medicine. Evaluation and treatment of recurrent pregnancy loss: A committee opinion. *Fertil Steril*, 2012. 98: 1103–1111. doi:10.1016/j.fertnstert.2012.06.048.

[135] National Institute for Health Research HTA - 08/38/01 [Internet]. Imperial College London, 2009 Oct - 2016 May. Identifier ISRCTN192644181 , First trimester progesterone therapy in women with a history of unexplained recurrent miscarriages: A randomised, double-blind, placebo-controlled, multi-centre trial [The PROMISE (PROgesterone in recurrent MIScarriage) Trial]. Availiable from: `https://www.journalslibrary.nihr.ac.uk/programmes/hta/083801#/` [Accessed: 27 August 2017].

[136] Drummond M. *Methods for the Economic Evaluation of Health Care Programmes*, pages 7–25. Oxford medical publications. Oxford University Press, 2005.

[137] National Institute for Health and Clinical Excellence. Ectopic pregnancy and miscarriage: diagnosis and initial management. NICE guideline (CG154), 2012.

[138] Parfit D. *Reasons and Persons*, pages 356–357. Oxford University Press, 1984.

[139] Baio G and Dawid AP. Probabilistic sensitivity analysis in health economics. *Statistical Methods in Medical Research*, 2015. 24: 615–634. doi:10.1177/0962280211419832.

[140] Claxton K, Sculpher M, McCabe C et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics*, 2008. 14: 339–347. doi:10.1002/hec.985.

[141] Coomarasamy A, Truchanowicz EG, and Rai R. Does first trimester progesterone prophylaxis increase the live birth rate in women with unexplained recurrent miscarriages? *BMJ*, 2011. 342: d1914. doi:10.1136/bmj.d1914.

[142] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[143] Raman-Wilms L, Tseng AL, Wighardt S et al. Fetal genital effects of first-trimester sex hormone exposure: A meta-analysis. *Obstetrics & Gynecology*, 1995. 85: 141–149. doi:10.1016/0029-7844(94)00341-A.

[144] Carmichael SL, Shaw GM, Laurent C et al. Maternal progestin intake and risk of hypospadias. *Archives of pediatrics & adolescent medicine*, 2005. 159: 957–62. doi:10.1001/archpedi.159.10.957.

[145] Meis PJ, Klebanoff M, Thom E et al. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. *New England Journal of Medicine*, 2003. 348: 2379–85. doi:10.1056/NEJMoa035140.

[146] Swyer GIM and Daley D. Progesterone Implantation in Habitual Abortion. *BMJ*, 1953. 1: 1073–1077. doi:10.1136/bmj.1.4819.1073.

[147] Goldzieher JW. Double-blind trial of a progestin in habitual abortion. *JAMA*, 1964. 188: 651–654. doi:10.1001/jama.1964.03060330031008.

[148] Levine L. Habitual Abortion. A Controlled Study of Progestational Therapy. *West J Surg Obstet Gynecol*, 1964. 72: 30–36.

[149] El-Zibdeh M. Dydrogesterone in the reduction of recurrent spontaneous abortion. *The Journal of Steroid Biochemistry and Molecular Biology*, 2005. 97: 431–434. doi:10.1016/j.jsbmb.2005.08.007.

[150] O'Hagan A, Buck C, Daneshkhah A et al. *Uncertain Judgements: Eliciting Experts' Probabilities*. Statistics in Practice. Wiley, 2006.

[151] Department of Health. NHS Reference Costs 2014–15, 2015. `https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015` [Accessed: 12 January 2017].

[152] Nuffield Trust. Health spending per head by country of the UK, 2014. `https://www.nuffieldtrust.org.uk/chart/health-spending-per-head-by-country-of-the-uk` [Accessed: 16 September 2017].

[153] Office for National Statistics. Birth Summary Tables, England and Wales: 2014, 2014. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsummarytablesenglandandwales/2015-07-15 [Accessed: 10 April 2017].

[154] Love-Koh J, Asaria M, Cookson R et al. The social distribution of health: Estimating quality-adjusted life expectancy in england. *Value in Health*, 2015. 18: 655 – 662. doi:https://doi.org/10.1016/j.jval.2015.03.1784.

[155] Barton GR, Briggs AH, and Fenwick EA. Optimal Cost-Effectiveness Decisions: The Role of the Cost-Effectiveness Acceptability Curve (CEAC), the Cost-Effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfection Information (EVPI). *Value in Health*, 2008. 11: 886 – 897. doi:http://dx.doi.org/10.1111/j.1524-4733.2008.00358.x.

[156] Stirrat GM. Recurrent miscarriage. *The Lancet*, 1990. 336: 673–675.

[157] Mason H, Jones-Lee M, and Donaldson C. Modelling the monetary value of a QALY: a new approach based on UK data. *Health Economics*, 2009. 18: 933–950. doi:10.1002/hec.1416.

[158] Prior M, Hibberd R, Asemota N et al. Inadvertent p-hacking among trials and systematic reviews of the effect of progestogens in pregnancy? A systematic review and meta-analysis. *BJOG*, 2017. 124: 1008—-1015. doi:10.1111/1471-0528.14506.

[159] Chan YY. *Study to Evaluate the Prevalence, Importance, and Treatment of Women with Congenital Uterine Anomalies*. Ph.D. thesis, University of Nottingham, 2016.

[160] INVOLVE. What is public involvement in research?, 2017. http://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/ [Accessed: 23 October 2017].

[161] JLA Team at NETSCC and Cowan K. *The James Lind Alliance Guidebook. Version 6*, pages 42–50. National Insitute for Health Research, 2016.

[162] Elwyn G, Crowe S, Fenton M et al. Identifying and prioritizing uncertainties: patient and clinician engagement in the identification of research questions. *Journal of Evaluation in Clinical Practice*, 2010. 16. doi:10.1111/j.1365-2753.2009.01262.x.

[163] Bruton A, Kirby S, Arden-Close E et al. The BREATHE study: Breathing REtraining for Asthma — Trial of Home Exercises. A protocol summary of a randomised controlled trial. *Primary Care Respiratory Journal*, 2013. 22: PS1–PS7. doi:10.4104/pcrj.2013.00047.

[164] Duley L, Uhm S, and Oliver S. Top 15 UK research priorities for preterm birth. *The Lancet*, 2014. 383: 2041–2042. doi:10.1016/S0140-6736(14)60989-2.

[165] Heazell AEP, Whitworth MK, Whitcombe J et al. Research Priorities for Stillbirth: Process Overview and Results from the UK Stillbirth Priority Setting Partnership. *Ultrasound Obstet Gynecol*, 2015. 46: 641–647. doi:10.1002/uog.15738.

[166] Wan YL, Beverley-Stevenson R, Carlisle D et al. Working together to shape the endometrial cancer research agenda: The top ten unanswered research questions. *Gynecologic Oncology*, 2016. 143: 287–293. doi:10.1016/j.ygyno.2016.08.333.

[167] Smith J and Morley R. The emergency medicine research priority setting partnership. *Emerg Med J*, 2015. 32: 830. doi:10.1136/emermed-2015-205353.

[168] James Lind Alliance. Miscarriage Priority Setting Partnership, 2014. `http://www.jla.nihr.ac.uk/priority-setting-partnerships/miscarriage/` [Accessed: 23 October 2017].

[169] Kolte AM, Bernardi LA, Christiansen OB et al. Terminology for pregnancy loss prior to viability: a consensus statement from the ESHRE special interest group, early pregnancy. *Hum Reprod*, 2014. doi:10.1093/humrep/deu299.

[170] Births and Deaths Registration Act 1953, s. 7. `http://www.legislation.gov.uk/ukpga/Eliz2/1-2/20/section/41` [Accessed: 21 March 2017].

[171] Knight SR, Metcalfe L, O'Donoghue K et al. Defining Priorities for Future Research: Results of the UK Kidney Transplant Priority Setting Partnership. *PLOS ONE*, 2016. 11: 1–14. doi:10.1371/journal.pone.0162136.

[172] Boney O, Bell M, Bell N et al. Identifying research priorities in anaesthesia and perioperative care: final report of the joint National Institute of Academic Anaesthesia/James Lind Alliance Research Priority Setting Partnership. *BMJ Open*, 2015. 5. doi:10.1136/bmjopen-2015-010006.

[173] Prior M, Bagness C, Brewin J et al. Priorities for research in miscarriage: a priority setting partnership between people affected by miscarriage and professionals following the James Lind Alliance methodology. *BMJ Open*, 2017. 7: e016571. doi:10.1136/bmjopen-2017-016571.

[174] JLA Team at NETSCC and Cowan K. *The James Lind Alliance Guidebook. Version 6*, pages 54–55. National Insitute for Health Research, 2016.

Appendices

1 Ethical approval

NRES Committee East Midlands - Nottingham 1
Royal Standard Place
Nottingham
NG1 6FS

Tel: 0115 8839697

14 July 2015

Mr Nicholas Raine-Fenning
Consultant Gynaecologist & Reader of Reproductive Medicine and Surgery
The University of Nottingham
Nottingham University Research and Treatment Unit in Reproduction (NURTURE)
B Floor, East Block
Queen's Medical Centre, Nottingham
NG7 2UH

Dear Mr Raine-Fenning,

Study title:	**Randomised, controlled study of hysteroscopic metroplasty for women with a septate uterus and a history of miscarriage or preterm labour.**
REC reference:	**13/EM/0362**
Protocol number:	**13GY007**
Amendment number:	**1**
Amendment date:	**22 June 2015**
IRAS project ID:	**131663**

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Notice of Substantial Amendment (non-CTIMP)		22 June 2015
Participant information sheet (PIS)	1.2	19 May 2015
Research protocol or project proposal	1.2	19 May 2015

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

2 Off-line Analysis of ultraSonographic Images Study (OASIS): Uterine Anomalies

Name:

Unit:

Do you have experience using 3D ultrasound in gynaecology?:

How many years have you been using 3D ultrasound to assess uterine anomalies?:

What classification system do you use for diagnosing uterine anomalies:

Case Number	*Is the uterus normal? (Yes/No)*	*What is your diagnosis?*	*Is the endometrial cavity shape normal? (Yes/No)*	*Describe the 'internal' contour of the fundal aspect of the endometrial cavity (Concave/Convex)*	*Can you see two tubal ostea? (Yes/No)*	*Is the external 'outer' contour of the uterus normal? (Yes/No)*	*Is there a uterine septum? (Yes/No)*	*What is the septal length? (cm)*	*What is the septal width (cm)?*
1									
2									
3									
4									
...									
20									

3 Sample Size Survey

In some women the uterus (womb) is divided into two halves by a thin membrane, called a septum. This may affect fertility and increase the risk of miscarriage. We are considering conducting a study to assess the role of treatments for women with a womb septum. Before we decide whether to proceed we need to know what patients think about it.

Please complete this questionnaire to help us design the study.

Please tell us a bit about you

1. **What is your age?**

   ☐ 18-29 years old

   ☐ 30-49 years old

   ☐ 50-64 years old

   ☐ 65 years and over

2. **What is the highest level of education that you have completed?**

   ☐ Some secondary school

   ☐ Secondary school GCSE level

   ☐ A - Level Trade/technical/vocational training

   ☐ University graduate

   ☐ Postgraduate degree

3. **What sex are you?**

   ☐ Male

   ☐ Female

4. **How many children have you had?**

   ☐ 0

   ☐ 1

   ☐ 2

   ☐ 3

☐ 4

☐ 5 or more

5. **Have you suffered from or experienced any of the following (you can select more than one option):**

   ☐ Subfertility or infertility (Trying for a baby for more than 12 months)

   ☐ Subfertility or infertility and are considering IVF

   ☐ 2 or more unsuccessful IVF cycles and are planning to go for further IVF treatment

   ☐ One early miscarriage (<12 weeks pregnant)

   ☐ Two consecutive (one after another) early miscarriages (<12 weeks pregnant)

   ☐ Three or more consecutive (one after another) early miscarriages (<12 weeks pregnant)

   ☐ Late miscarriage (> 12 weeks pregnant)

   ☐ Premature delivery (<34 weeks pregnant)

   ☐ None of the above

6. **Treatment One - Vaginal Pessaries**
   Imagine that a medicine had been developed to reduce the risk of miscarriage but that we are not sure if it works. It is widely used and considered safe in pregnancy. The medicine is taken as a pessary inserted into the vagina twice a day starting when you first have a positive pregnancy test until you are 12 weeks pregnant. Our question is;

   How much would the risk of miscarriage need to be reduced to make you bother using the pessary?

   To help you answer the risk of miscarriage varies from 0% (no risk of miscarriage) to 100% (every woman will miscarry). Please imagine that the average risk of miscarriage for women with a womb septum is 30%. What reduction in average score from 30% to what would make you wish to use the pessary?

For example if the drug reduced the miscarriage risk from 30% to 29% I guess you would probably regard that benefit as not worth the bother and risk of taking a pessary. However if it reduced the average risk down to say 5%, I presume you would probably want to take the pessary.

☐ What would be the cut off for you? .....

7. **Treatment Two - Keyhole Surgery**
   In hysteroscopic septal resection a thin flexible tube with a camera on the end (a hysteroscope) is inserted into the vagina, through the cervix and into the womb. Instruments are passed through the hysteroscope into the womb and the septum is removed. It is performed regularly but we do not know for sure if it is beneficial. Hysteroscopy is a common procedure in gynaecology and can be performed as a day case. Complications include bleeding (1 in 1000), infection (1 in 250), damage to the neck of the womb (1 in 135) and making a hole in the womb (1 in 500 women). Patients may develop fluid overload if an excessive amount of fluid is used during hysteroscopic resection. Overall complication rates for this type of surgery are just under 1%. The operation would be performed before you try to conceive. Our question is;

   How much would the risk of miscarriage need to be reduced to make you agree to have the operation?

   Fill in the risk of miscarriage after surgery that would make you wish to have the operation. We assume that a risk of 29% would not be worthwhile. But that a risk of 5% would be well worth it.

   ☐ What would be the cut off for you? .....

8. **Treatment Three - Major Surgery**
   In open septal resection a cut is made is made along the bikini line of the abdomen. The womb is cut open and the septum removed before the womb is stitched back together. It is rarely performed, requires at least three days in hospital and has a high risk of complications. We are unsure how beneficial this operation is. Complications include bleeding, blood transfusion, hysterectomy, infection and making a hole in the

womb. The overall complication rates for this type of surgery are unknown but much higher than keyhole surgery. There is a risk of rupture of the uterus in any future pregnancy which may be dangerous for you and the baby. In any future pregnancy the baby would require delivery by Caesarean Section. The operation would be performed before you try to conceive and you should not try to conceive until 12 weeks after the procedure. Our question is;

> *How much would the risk of miscarriage need to be reduced to make you want to have the operation?*

Fill in the risk of miscarriage after surgery that would make you wish to have the operation. We assume that a risk of 29% would not be worthwhile. But that a risk of say 5% would be well worth it. Where is the cut-off?"

☐ What would be the cut off for you? .....

9. **How easy did you find this survey?**

   ☐ I understood it completely and I am confident in my answers

   ☐ It was difficult, but I think I understood it.

   ☐ I'm fairly confident that my answers are valid. I couldn't understand it at all.

10. **Please share any comments you may have with us.**

Thank you for taking the time to complete this survey.

4 Patient Survey

In some women the uterus (womb) is divided into two halves by a thin membrane, called a septum. This may affect fertility and increase the risk of miscarriage. We are seeking your views about a possible research study into the role of surgery for women with a womb septum.

The aim of the trial is to compare surgery with no treatment for women with a womb septum.

Current national guidance about surgical removal of a womb septum says it is acceptable to perform for women with multiple miscarriages, but that the evidence to support it is weak. There is even less evidence for women who have infertility.

We want to find out if surgical removal of a uterine septum is safe and effective for women with previous miscarriage or infertility. The best way to do this is with a randomised trial. Women in the trial will have either surgery or not.

Women in the trial won't be able to choose whether they have the operation or not, and neither will their doctors. However all women will be followed up to find out about the outcome of any future pregnancies.

We would really like to know what you think about this trial and how you might feel if you had a womb septum, and previous miscarriage or infertility and were asked to take part. Most of the questions are tick-boxes but there are also spaces for you to comment if you'd like to. All your answers will remain anonymous.

Thank you,

Dr Matthew Prior
Mr Nick Raine-Fenning,
Professor Jim Thornton,

University of Nottingham

Please tell us a bit about you

1. **What is your age?**

   ☐ 18-29 years old

   ☐ 30-49 years old

   ☐ 50-64 years old

   ☐ 65 years and over

2. **What is the highest level of education that you have completed?**

   ☐ Some secondary school

   ☐ Secondary school GCSE level

   ☐ A - Level Trade/technical/vocational training

   ☐ University graduate

   ☐ Postgraduate degree

3. **What sex are you?**

   ☐ Male

   ☐ Female

4. **How many children have you had?**

   ☐ 0

   ☐ 1

   ☐ 2

   ☐ 3

   ☐ 4

   ☐ 5 or more

5. **Have you suffered from or experienced any of the following (you can select more than one option):**

   ☐ Subfertility or infertility (Trying for a baby for more than 12 months)

   ☐ Subfertility or infertility and are considering IVF

- ☐ 2 or more unsuccessful IVF cycles and are planning to go for further IVF treatment
- ☐ One early miscarriage (<12 weeks pregnant)
- ☐ Two consecutive (one after another) early miscarriages (<12 weeks pregnant)
- ☐ Three or more consecutive (one after another) early miscarriages (<12 weeks pregnant)
- ☐ Late miscarriage (> 12 weeks pregnant)
- ☐ Premature delivery (<34 weeks pregnant)
- ☐ None of the above

6. **Do you think a trial to assess the benefits and disadvantages of surgery to remove a womb septum is a good idea?**

   - ☐ Yes
   - ☐ No

7. **You can add any comment here, if you'd like to:**

8. **If you had subfertility or infertility (trying for a baby for more than 12 months), do you think you'd be willing to take part in the trial we've described?**

   - ☐ Yes
   - ☐ No

9. **If you had subfertility or infertility and were considering IVF, do you think you'd be willing to take part in the trial we've described?**

   - ☐ Yes
   - ☐ No

10. **If you had 2 or more unsuccessful IVF cycles and were planning to go for further IVF treatment, do you think you'd be willing to take part in the trial we've described?**

☐ Yes

☐ No

11. **If you had one early miscarriage (<12 weeks pregnant), do you think you'd be willing to take part in the trial we've described?**

    ☐ Yes

    ☐ No

12. **If you had two consecutive (one after another) early miscarriages (<12 weeks pregnant), do you think you'd be willing to take part in the trial we've described?**

    ☐ Yes

    ☐ No

13. **If you had three or more consecutive (one after another) early miscarriages (less than 12 weeks pregnant), do you think you'd be willing to take part in the trial we've described?**

    ☐ Yes

    ☐ No

14. **If you had a late miscarriage (over 12 weeks pregnant), do you think you'd be willing to take part in the trial we've described?**

    ☐ Yes

    ☐ No

15. **If you had a premature delivery (less than 34 weeks pregnant), do you think you'd be willing to take part in the trial we've described?**

    ☐ Yes

    ☐ No

16. **In this trial, a computer would decide whether you had the surgery or not. This may mean you do not know for a few days which you were having. Do you think this would be:**

    ☐ Very acceptable

- ☐ Acceptable
- ☐ No opinion
- ☐ Unacceptable
- ☐ Very unacceptable

17. **If you had a womb septum and chose to have it removed, what do you think would be the most – and least - important outcomes for you? (Please rank 1-5, with 1 being the most important and 5 being the least important)**

    - ☐ Another pregnancy
    - ☐ Another miscarriage
    - ☐ A premature birth less than 28 weeks
    - ☐ A premature birth less than 34 weeks
    - ☐ A live birth at term

18. **Please share any comments you may have with us.**

Thank you for taking the time to complete this survey.

5 Clinician Survey

We are developing a randomised controlled trial for hysteroscopic septal resection in women with a history of miscarriage or infertility who have a septum.

Please complete this short survey to help us identify the diagnosis, prevalence and management of women with uterine anomalies.

Many thanks

Matthew, Jim and Nick

Matthew Prior - Clinical Research Fellow
Jim Thornton - Professor of Obstetrics and Gynaecology
Nick Raine-Fenning – Reader of Reproductive Medicine and Surgery
The University of Nottingham

1. **What is your name?**

2. **Where do you work?**

3. **Which classification systems do you use to diagnose uterine anomalies?**

   ☐ American Fertility Society 1988

   ☐ Salim 3D 2003

   ☐ VCUAM 2005

   ☐ ESHRE / ESGE 2013

   ☐ I do not use a classification system

4. **What is the Gold Standard test in your Unit used to diagnose and define the presence and type of uterine anomaly?**

- ☐ MRI
- ☐ 3D Transvaginal Ultrasound
- ☐ Saline Hysterosonography
- ☐ Hysteroscopy
- ☐ HSG

5. **Does your unit routinely use 3D ultrasound?**

   - ☐ Yes
   - ☐ No

6. **Does your unit have access to 3D ultrasound?**

   - ☐ Yes
   - ☐ No

7. **How many uterine septa have you seen in the last year?**

8. **How many uterine septa have you managed conservatively in the last year?**

9. **How confident are you in the response to Q7 and Q8?**

   - ☐ Confident
   - ☐ About right
   - ☐ Guess

10. **Do you perform hysteroscopic septal resection?**

    - ☐ Yes
    - ☐ No

11. **How many uterine septa have you operated on in the last year?**

12. **In which conditions did you choose to operate on uterine septa?**

    ☐ Primary subfertility

    ☐ Secondary subfertility

    ☐ Subfertility requiring IVF

    ☐ One previous miscarriage

    ☐ Two previous miscarriages

    ☐ Three or more previous miscarriages

    ☐ Second trimester miscarriage

    ☐ Previous preterm birth

13. **Please describe the surgical technique used to resect a uterine septum. (E.g. sharp / blunt dissection, scissors, electrodiathermy?)**

14. **Do you offer follow up imaging post procedure?**

    ☐ Yes

    ☐ No

15. **Do you use any adjuncts? E.g. IUD, barriers, antibiotics, oral contraceptive pill, HRT.**

16. **What follow up imaging do you offer post procedure?**

    ☐ 3D ultrasound

    ☐ MRI

- ☐ HSG
- ☐ Other (please specify)

17. **Would you be willing to randomise a patient with a uterine septum to septal resection with one or more of the following conditions?**

    - ☐ Primary subfertility
    - ☐ Secondary subfertility
    - ☐ Subfertility requiring IVF
    - ☐ One previous miscarriage
    - ☐ Two previous miscarriages
    - ☐ Three or more previous miscarriages
    - ☐ Second trimester miscarriage
    - ☐ Previous preterm birth

18. **Would you interested in becoming a centre as part of a randomised controlled trial of hysteroscopic septal resection for women with a uterine septum and a history of sub fertility, miscarriage or preterm birth?**

    - ☐ Yes
    - ☐ No

19. **Thank you for your interest please leave a contact email address so we can follow this up.**

Thank you for taking the time to complete this survey.

6 Investigator Briefing

A pilot randomised controlled trial of hysteroscopic septal resection for women with a history of infertility, miscarriage or preterm birth

INVESTIGATOR BRIEFING

SEPTUM

Systematic **E**valuation of the **P**revalence and **T**reatment **U**terine **M**alformations studies

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Nottingham Hospitals Charity

Nottingham University Hospitals NHS
NHS Trust

SEPTUM

Nottingham University Hospital NHS Trust

CLINICAL STUDY PROTOCOL

Dr Matthew Prior, Research Fellow

Agenda

Nottingham University Hospital NHS Trust

- Study overview
- Study population
- Study intervention
- Study follow up

SEPTUM

Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Study Title: SEPTUM

Nottingham University Hospital NHS Trust

A pilot randomised controlled trial of hysteroscopic septal resection for women with a history of infertility, miscarriage or preterm birth

As part of

Systematic **E**valuation of the **P**revalence and **T**reatment **U**terine **M**alformations studies

SEPTUM

Systematic **E**valuation of the **P**revalence and **T**reatment **U**terine **M**alformations studies

Where we are up to…

- Sponsor: Nottingham University Hospitals
- Funder: Nottingham University Hospitals Charity
- Ethical Approval: Granted October 2013
  - Extended to allow multiple sites July 2015
- Trial Registration: ISRCTN 28960271

Rationale for the study

Nottingham University Hospital NHS Trust

Metroplasty versus expectant management for women with recurrent miscarriage and a septate uterus (Review)

Kowalik CR, Goddijn M, Emanuel MH, Bongers MY, Spinder T, de Kruif JH, Mol BWJ, Heineman MJ

THE COCHRANE COLLABORATION®

SEPTUM

Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Rationale for the study

Nottingham University Hospital NHS Trust

Metroplasty versus expectant management for women with recurrent miscarriage and a septate uterus (Review)

Main results

No randomised controlled trials were identified for inclusion.

Authors' conclusions

Hysteroscopic metroplasty in women with recurrent miscarriage and a septate uterus is being performed in many countries to improve reproductive outcomes in women.This treatment has been assessed in non-controlled studies, which suggested a positive effect on pregnancy outcomes. However, these studies are biased due to the fact that the participants with recurrent miscarriage treated by hysteroscopic metroplasty served as their own controls. Until now, the effectiveness and possible complications of hysteroscopic metroplasty have never been considered in a randomised controlled trial.Taking this into account there is insufficient evidence to support this treatment in these women. A randomised controlled trial is urgently needed and currently underway (www.studies-obsgyn.nl/trust NTR 1676).

THE COCHRANE COLLABORATION®

SEPTUM

Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Study Synopsis

Nottingham University Hospital NHS Trust

- **Randomised Controlled Trial**
- **Pilot / Feasibility Study**
- **Intervention:** Hysteroscopic Septal Resection
- **Comparator:** No intervention
- **Duration:** Three years
- **Study Centres:** Across the UK

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Study Aim

Nottingham University Hospital NHS Trust

To determine the feasibility of performing a randomised controlled study of hysteroscopic septal resection versus no intervention for women with a uterine septum and a history of infertility, miscarriage or preterm birth.

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Study Objectives

Nottingham University Hospital NHS Trust

- Determine the prevalence of uterine septae in the three populations of infertility, miscarriage and preterm birth.
- Examine the numbers of patients needed to screen in relation to the numbers randomised into the study.
- Gather the reasons for women not wishing to take part in the SEPTUM study.
- Identify the barriers and facilitators to the diagnosis of uterine septae using three-dimensional ultrasound within real clinical practice.
- Identify the barriers and facilitators to the use of hysteroscopic septal resection within real clinical practice.
- Gather the reasons for clinicians choosing not to enrol their patients in the SEPTUM study.
- Monitor adherence to the trial protocol across multiple sites.
- Determine follow up rate for women in the control and intervention arms of the study.
- Test the methods of data collection and determine the time required collecting study data.

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Agenda

Nottingham University Hospital NHS Trust

- Study overview
- Study population
- Study intervention
- Study follow up

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Inclusion Criteria

Nottingham University Hospital NHS Trust

- Women trying to conceive
- History of infertility, one or more early miscarriages (<12 completed weeks), one late miscarriage (>12 completed weeks until 23 completed weeks), or preterm delivery, regardless of previous viable or live birth
- Age: > 18 years at randomisation
- Confirmed septate uteri (pre-op or intra-op) on three-dimensional ultrasound scan
- Willing and able to give informed consent
- No pregnancy or endometrial surgery between diagnosis and treatment or follow up

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Exclusion Criteria (1)

Nottingham University Hospital NHS Trust

- Currently pregnant
- Current or past treatment for septate uteri
- Use of any contraceptions (Females and males)
- Contraindication to hysteroscopic septal resection
- Age < 18 years old
- Body Mass Index above 40kg/m2

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Exclusion Criteria (2)

Nottingham University Hospital NHS Trust

- Exclusion criteria does not include women with:
  - Antithyroid antibodies
  - Antiphospholipid antibodies
  - Lupus anticoagulant antibodies
  - Anticardiolipin antibodies

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Withdrawal Criteria

Nottingham University Hospital NHS Trust

- Participants may voluntarily withdraw participation in this study at any time.

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Agenda

Nottingham University Hospital NHS Trust

- Study overview
- Study population
- Study intervention
- Study follow up

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Study Intervention

Nottingham University Hospital NHS Trust

- Hysteroscopic septal resection
- General anaesthesia, day case.

- Using an operating hysteroscope with
  - Scissors
  - Versapoint
  - Other bipolar electrode system.

- The choice of distension media and instrument depends on the surgeon and instruments .

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Agenda

Nottingham University Hospital NHS Trust

- Study overview
- Study population
- Study intervention
- Study follow up

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Study Follow Up

Nottingham University Hospital NHS Trust

- Repeat ultrasound 6 – 8 weeks post op
- Study participants will be participating in the study for approximately 24 months from treatment if not pregnant, or 34 months from treatment until the end of the first clinical pregnancy.

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Nottingham University Hospital NHS Trust

Study Procedures – 3D Ultrasound

- This is what we are using for diagnosis of a uterine septum
- Any stage of the menstrual cycle
- Image standardisation

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Nottingham University Hospital NHS Trust

Image standardisation

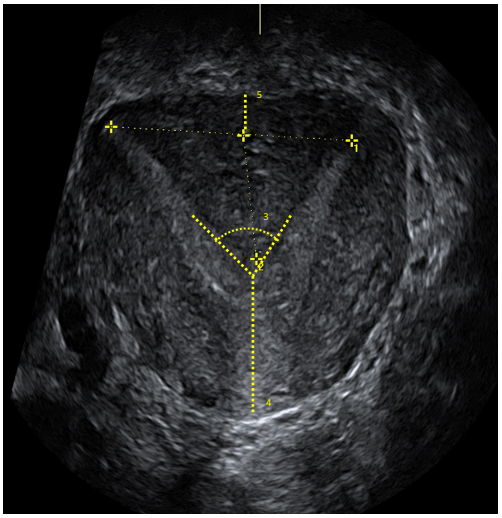
- Coronal Plane
- Both ostea visible
- Following measurements needed
- Save the volume and the image used to make the diagnosis and submit to CRF.

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Nottingham University Hospital NHS Trust

Diagnosis

1: Septum width
2: Septum length
3: Septal angle
4: Residual Cavity Length
5.: Fundal uterine wall thickness



SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Nottingham University Hospital NHS Trust

Safety

- Abnormal 3D ultrasound scans suggesting malignancy
- Hysterectomy
- Uterine rupture
- Death

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Nottingham University Hospital NHS Trust

Endpoints

- First trimester miscarriage (12 completed weeks), second trimester miscarriage (>13 weeks until 23 completed weeks)

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Nottingham University Hospital NHS Trust

Endpoints

- Ectopic pregnancy
- Live birth beyond 28 weeks, 32 weeks and 37 weeks gestation
- Mode of delivery
- Placenta praevia
- Morbidly adherent placenta
- Post-partum haemorrhage of > 1000ml
- Birth weight
- Surgical complications: Uterine perforation, fluid overload, endometritis, bleeding, incomplete resection, synechia/adhesions

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Nottingham University Hospital NHS Trust

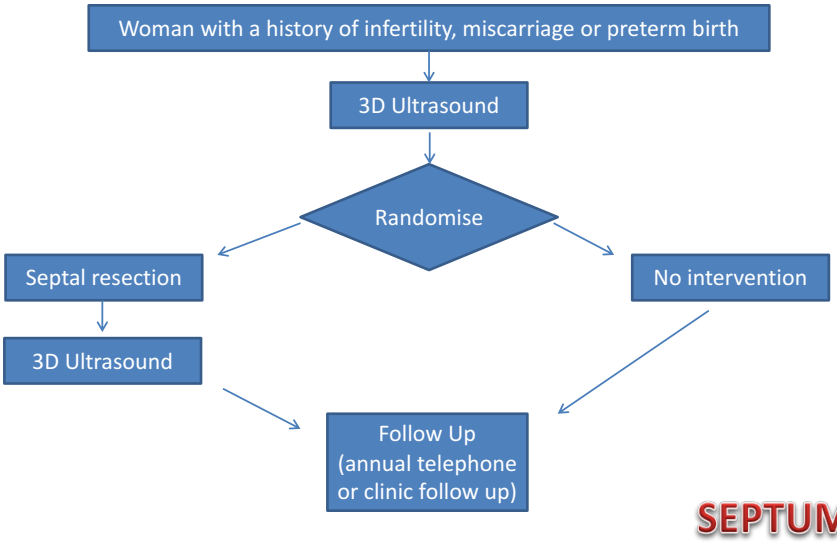
Case Report Form

- We have a bespoke software system to manage each patients case record.
- This deals with data entry, randomisation and follow up.
- You will also need a site log to record patient identifiable information required for follow up.

SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Nottingham University Hospital NHS Trust

Patient Pathway



SEPTUM
Systematic Evaluation of the Prevalence and Treatment Uterine Malformations studies

Nottingham University Hospital NHS Trust



Contact:
Dr Matthew Prior
matthew.prior@nottingham.ac.uk
T: 07817 627712

SEPTUM

Systematic **E**valuation of the **P**revalence and **T**reatment **U**terine **M**alformations studies

7 Identifying the top research priorities for miscarriage

Why we need your help

We are asking for your help because we want to improve care and treatment for women who experience miscarriage and those affected by it. We know that many people have important unanswered questions about miscarriage and we want to find out what they are. We want to use these questions to set priorities for research that will improve care.

Who you are

We would like the following groups of people to submit questions to our survey:

Woman who have had a miscarriage. People affected by miscarriage e.g partners, family members, friends, colleagues. Professionals involved in caring for women who have experienced miscarriage – doctors, nurses, sonographers and other professionals with experience of miscarriage. Professional bodies, patient groups, charities and other organisations involved with miscarriage. We are not just asking women who miscarry but also seeking the views of those affected by miscarriage and professionals involved in caring for women who have had a miscarriage. This is because all groups have ideas about what research questions need to be asked, and which criteria they would use to judge success or failure.

What we are asking you to do

Please think about your own experience and let us know if there are any unanswered questions or if there is anything that could have been done better.

These might include how best to make the diagnosis, how best to treat the miscarriage, emotional support and how to prevent miscarriages happening in future.

You may want to think about the following:

- What would you have liked to know when you first thought you might be having a miscarriage?
- What questions did you have when you were first given the diagnosis?
- What questions about the different treatment options would you have liked answers to?
- How would you have liked to be supported?
- What would you have liked to happen next?
- What questions did you want to know about why you had miscarried?
- How would you improve care in the early pregnancy clinic?
- What is it normal to feel during and after a miscarriage and is there anything that can be done to help?
- Is there anything that would have helped you to know about your risk in another pregnancy?
- Did you have any questions about treatment, drugs or surgery that may help reduce the chance of miscarriage?

About you

We also ask for a little information about you. This section of the survey is optional.

What will happen to your question(s)

All the question(s) we get from everyone who takes part in this survey will be gathered together and responses will kept anonymous. The questions that we haven't yet got the answers to will be published in the UK Database for Uncertainties of Effects of Treatment (UK DUETS).

We will bring together groups of women, partners, family members and healthcare professionals to look at these questions and identify the most important ones. You can be involved in this, if you like.

This will result in a 'top 10' list of research topics which we will use to influence future research. For example it could help decide which research projects should get funding.

What unanswered questions about miscarriage would you like to see answered by research?

Question 1

Question 2

Question 3

Guidance

You can submit up to three questions on this form. Previous 'Priority Setting Partnerships' have found that the most useful format for such a survey is for respondents to write short, simple questions that they would most like to be answered.

Here are some examples from similar surveys in other conditions:

- Are breathing exercises helpful in controlling asthma?
- What is the evidence for gargling with aspirin to relieve a sore throat?
- What psychological support should be provided for patients in Intensive Care?

More information about this survey can be found at
www.miscarriagePSP.org/survey

About You

We want to make sure we are getting a broad representation of people contributing ideas. Please tell us a little about your experience, background, and how you came across the survey. All responses are optional.

1. What is your gender?

    ☐ Male

    ☐ Female

    ☐ Prefer not to say

    ☐ How old are you?

    ☐ Less than 18 years

    ☐ 18-24 years

    ☐ 25-29 years

    ☐ 30-34 years

    ☐ 35-39 years

    ☐ 40-44 years

    ☐ 45-49 years

    ☐ 50 years or over

    ☐ Prefer not to say

2. Where do you live?

    ☐ East Midlands

    ☐ East of England

    ☐ London

    ☐ North East

    ☐ North West

    ☐ Northern Ireland

    ☐ Scotland

    ☐ South East

    ☐ South West

    ☐ Wales

- ☐ West Midlands
- ☐ Yorks and Humber
- ☐ Outside the UK
- ☐ Prefer not to say
- ☐ Other (please specify)

3. What is your ethnic group?

   - ☐ White English/Welsh/Scottish/Northern Irish/British White Irish
   - ☐ White Gypsy or Irish Traveller
   - ☐ Mixed White and Black Caribbean
   - ☐ Mixed White and Black African
   - ☐ Mixed White and Asian
   - ☐ Asian / Asian British Indian
   - ☐ Asian / Asian British Pakistani
   - ☐ Asian / Asian British Bangladeshi
   - ☐ Asian / Asian British Chinese
   - ☐ African
   - ☐ Caribbean
   - ☐ Prefer not to say
   - ☐ Other (please specify)

4. Which best describes you?

   - ☐ I am a woman who has experienced miscarriage
   - ☐ I am the partner of a woman who has experienced miscarriage
   - ☐ I am a family member, friend or colleague of women who has experienced miscarriage?
   - ☐ I am a healthcare professional?
   - ☐ I am a healthcare professional body
   - ☐ I am a charity
   - ☐ Other (please specify)

5. If you are a woman who has had a miscarriage, how many miscarriages have you had?

. . . . . .

6. At what stages were your miscarriages?

   ☐ Before 6 weeks

   ☐ Between 6 and 12 weeks

   ☐ After 12 weeks

7. How long is it since your last miscarriage?

   . . . . . .

8. If you are a healthcare professional, which best describes your role?

   ☐ Obstetrician / Gynaecologist

   ☐ General Practitioner

   ☐ Nurse

   ☐ Midwife

   ☐ Sonographer

   ☐ Healthcare assistant

   ☐ Counsellor

   ☐ Other (please specify)

9. Where did you hear about the Miscarriage Priority Setting Partnership?

   ☐ Social Media

   ☐ Email

   ☐ Flyer or poster

   ☐ At a conference

   ☐ In an Early Pregnancy Unit

   ☐ From a friend

   ☐ Other (please specify)

Thank you for completing this survey. Please provide your email address if you would like to be kept up to date with the project.

We're looking for a broad range of people to take the survey, so please also share it with others friends or colleagues you know.

If you have been affected by miscarriage and would like to access support, the following organisations may be able to provide help and advice (click to link)

The Miscarriage Association Tel: 01924 200 799

Scottish Care and Information on Miscarriage Tel:0141 552 5070

Tommys Tel: 0800 0147 800

8 All refined questions

Reference	*Refined Uncertainty*
A	What causes miscarriage?
B	What causes recurrent (more than three to the same woman) miscarriages?
C	What causes late (after 14 weeks) miscarriage?
D	What causes missed miscarriage, where the baby dies before the miscarriage occurs?
E	Do lifestyle factors (diet, stress, exercise, weight, alcohol, sexual activity, smoking, night shifts or flying) cause miscarriage?
F	Does fetal gender cause miscarriage in some women?
G	Do immunological abnormalities cause miscarriage?
H	To what extent do genetic and chromosomal abnormalities in the fetus cause miscarriage?
I	What male factors contribute towards the cause of miscarriage?
J	Do medical tests cause miscarriage? (such as internal examination, ultrasound scanning, amniocentesis or chorionic villus biopsy)
K	To what extent do pre-existing medical conditions cause miscarriage? (e.g. vitamin deficiencies, diabetes, previous infertility, endometriosis, polycystic ovarian syndrome, menstrual irregularities, cervical factors, uterine anomalies, polyps, immunological factors or previous pregnancy complications e.g. caesarean section or preterm birth)
L	Does infection cause miscarriage?
M	Do medicines cause miscarriage? (e.g. contraceptive pill, progesterone, vaccines, antidepressants, antibiotics, antihistamines, or vitamin supplements)
N	To what extent do parental factors cause miscarriage? (e.g. genetic, chromosomal, blood group)
O	Does super fertility (more frequent implantation and ease of pregnancy) cause miscarriage?
P	What causes preterm spontaneous rupture of membranes as a type of late miscarriage?
Q	Does breastfeeding when pregnant cause miscarriage?
R	Does ethnicity affect miscarriage?
S	Can miscarriage of a current pregnancy be predicted? (e.g. by ultrasound, blood test or vaginal tests).

T	How common is miscarriage in the United Kingdom?
U	Are miscarriage rates affected by gestational age, geography, maternal age?
V	What are the appropriate investigations for women after one, two, or three or more miscarriages?
W	What investigations are of true clinical value? (e.g. ultrasound, gene sequencing, natural killer cells, thromboelastography (TEG), microarray testing of the fetus, paternal investigations, plasminogen activator inhibitor polymorphism)
X	What is the optimal process for conservative, medical and surgical management of miscarriage? (efficacy, location, safety, pain relief, symptoms, information)
Y	Does the method of management of miscarriage affect future fertility or miscarriage?
Z	What are the risks after a miscarriage? (Such as change in menstrual pattern or mental health problems)
AA	What is the best method of contraception after miscarriage?
AB	What are the effective interventions to prevent miscarriage, threatened miscarriage and recurrent miscarriage? (e.g. lifestyle, vitamins, aspirin, early scans, HCG, dopamine agonists, progestogen, polytherapy, steroids, oestrogen, metformin, anticoagulants, intravenous immunoglobulin, intralipid and anti-tnf alpha)
AC	What preconception tests or interventions prevent miscarriage? (e.g vitamin supplements, folic acid, Dehydroepiandrosterone (DHEA), co-enzyme Q-10 or bariatric surgery)
AD	Does alternative or complimentary medicine prevent miscarriage? (e.g. traditional Chinese medicine, acupuncture)
AE	Does a dedicated recurrent miscarriage clinic with reassurance scans reduce the rates of miscarriage?
AF	What is the probability of a further miscarriage after one, two or three or more previous miscarriage(s)?
AG	What is the best time to conceive after a miscarriage?
AH	Why does society not talk about miscarriage?
AI	What is the best societal support for women with miscarriage (leave from work e.g. same as stillbirth)?
AJ	What types of emotional support are effective in preventing or treating women or their partners after a miscarriage?

AK	How can health care professionals be prepared to offer patient's guidance or referral?
AL	What types of care are effective in preventing mental health problems for mothers and fathers during and immediately after a miscarriage?
AM	What are the emotional and mental health impacts of miscarriage in the short and long term for the mother and the father?
AN	What type of care reduces anxiety during a subsequent pregnancy after miscarriage?
AO	What is the best way to support the extended family?
AP	Could women be better prepared for the possibility of miscarriage?
AQ	Does health professional follow up improve the likelihood of a subsequent successful pregnancy?
AR	Does counselling improve the likelihood of a subsequent successful pregnancy?
AS	Do dedicated early pregnancy units offer benefit to women having a miscarriage?
AT	What can be done to improve the coordination between different miscarriage services?
AU	When should women book with their midwife in pregnancy?
AV	Which treatments are effective for incompetent cervix? (e.g. progesterone, cervical cerclage or arabin pessary)
AW	Is surgery safe and effective for women with uterine anomalies? (septal resection)
AX	Does levothyroxine reduce the risk of miscarriage in women with thyroid antibodies?
AY	Does taking folic acid reduce the risk of miscarriage?
AZ	Does previous surgery on the uterus cause miscarriage?
BA	Do uterine muscle relaxants treat threatened miscarriage?
BB	Does preimplantation genetic screening improve outcome versus natural conception in women with unexplained recurrent miscarriage?
BC	Is bed rest or any lifestyle or personal interventions effective in preventing miscarriage?
BD	What factors increase the chance of embryo implantation?
BE	Do embryos feel pain?
BF	What is the best way to manage the loss of one twin? (including both medical and emotional support)