Ulipristal acetate versus conventional management of heavy menstrual bleeding (HMB, including uterine fibroids): a randomised controlled trial and exploration of mechanism of action – the UCON Trial

Trial Registration: ISRCTN20426843

Statistical Analysis Plan

SAP Version Number	Protocol Version Number
2.0	7.0

Name of Author:	Versha Cheed	Role:	Trial Statistician	Affiliation:	BCTU University of Birmingham
Signature of Author:	Versha Cheed	Date:	11/8/21		
Name of Chief Investigator:	Hilary Critchley	Role:	Chief Investigator	Affiliation:	University of Edinburgh
Signature of Chief Investigator:	Hilary O.D - Critchley	Date:	11/08/2021		
This Statistical Analysis Plan has been approved by:					
Name of Approver:	Lee Middleton	Role:	Senior Statistician	Affiliation:	BCTU University of Birmingham
Signature of Approver:		Date:	11/8/21		

Statistical Analysis Plan (SAP) Amendments					
SAP version number	**SAP section number**	**Description of and reason for change**	**Timing of change with respect to interim analysis/ final analysis/ database lock**	**Blind Reviewer**	
2.0	-	SAP updated into new version template. This includes the addition of the 'data manipulations' section (section 9.4) which details how the statistician will derive values prior to analysis.	Prior to final analysis.	Name: Signature: Date:	Natalie Rowland  11/8/21
2.0	2	Short description of the impact of the Urgent Safety Measures (USMs) added.	Prior to final analysis.	Name: Signature: Date:	Natalie Rowland  11/8/21
2.0	3	Objectives updated to include analysis of liver function data (added as a result of two USMs).	Prior to final analysis.	Name: Signature: Date:	Natalie Rowland  11/8/21
2.0	4.4	Addition of secondary outcomes added after trial had opened to recruitment: satisfaction with treatment, participant rating of effect of treatment, recommend to a friend and liver function data (UPA group only, including ALT, AST and other measurements according to local protocol). Removal of ICE-CAP capabilities scores (not collected after CRF change).	Prior to final analysis.	Name: Signature: Date:	Natalie Rowland  11/8/21
2.0	4.5	Clarification on how we will handle late returned forms	Prior to final analysis.	Name: Signature: Date:	Natalie Rowland  11/8/21
2.0	4.7	Target sample size updated to reflect new target following the first USM.	Prior to final analysis.	Name: Signature:	Natalie Rowland 

				Date:	11/8/21
2.0	5.1	Clarification that p-values will be produced for the primary outcome (in the primary analysis population only) and Serious Adverse Events (SAEs) only due to issues around multiplicity.	Prior to final analysis.	Name:	Natalie Rowland
				Signature:	
				Date:	11/8/21
2.0	5.3	Changes to analysis populations as a result of the two USMs.	Prior to final analysis.	Name:	Natalie Rowland
				Signature:	
				Date:	11/8/21
2.0	5.5	Definition of valid questionnaire returns clarified (post-randomisation rather than post-treatment).	Prior to final analysis.	Name:	Natalie Rowland
				Signature:	
				Date:	11/8/21
2.0	6.2	Clarification of which populations baseline characteristics will now be presented.	Prior to final analysis.	Name:	Natalie Rowland
				Signature:	
				Date:	11/8/21
2.0	7.2	Clarification of how adherence data will be presented (focusing on time to treatment change rather than tabulations).	Prior to final analysis.	Name:	Natalie Rowland
				Signature:	
				Date:	11/8/21
2.0	9.1	Clarification that all adjustment factors will be treated as fixed effects including centre (the latter considered appropriate due to increased complexity of analytical models as well as number of recruitment centres being relatively small which means a fixed effect is appropriate); clarification on what variables will be included in the generalised linear model if all assessment time are included in the model (repeated measures analysis).	Prior to final analysis.	Name:	Natalie Rowland
				Signature:	
				Date:	11/8/21
2.0	9.2	Clarification of how MMAS will be analysed if regression residuals are skewed; inspect of pooled data during validation processes have indicated that the data may be negatively skewed.	Prior to final analysis.	Name:	Natalie Rowland
				Signature:	
				Date:	11/8/21

2.0	9.3	Missing data approach updated to a Chained Equations approach.	Prior to final analysis.	Name: Signature: Date:	Natalie Rowland  11/8/21
2.0	9.5	Estimates of treatment effects to be taken from repeated measures linear regression (incorporating data at all assessment times) rather than separate linear regressions at each time point). Clarification of which populations data will be analysed.	Prior to final analysis.	Name: Signature: Date:	Natalie Rowland  11/8/21
2.0	9.6	Description of analyses added for secondary outcomes that were added after study start. Where appropriate, single models (e.g. GEE) for repeated assessments are now described, as opposed to separate assessments at each time point. Odds ratios now described rather than relative risks for consistency between outcomes. Clarification of what data will be analysed for each population. Clarification that p-values will not be produced for secondary outcomes.	Prior to final analysis.	Name: Signature: Date:	Natalie Rowland  11/8/21
2.0	9.9	Removal of subgroup analysis due to impact of USMs on sample size.	Prior to final analysis.	Name: Signature: Date:	Natalie Rowland  11/8/21
2.0	9.10	Rationalisation of proposed sensitivity analysis given changes to analysis populations, including addition of examination of recruitment phase (as a result of USMs) by treatment interaction. Clarification of which populations data will be analysed.	Prior to final analysis.	Name: Signature: Date:	Natalie Rowland  11/8/21

Abbreviations & Definitions	
Abbreviation / Acronym	**Meaning**
BCTU	Birmingham Clinical Trials Unit
CONSORT	Consolidated Standards of Reporting Trials
DMC	Data Monitoring Committee
ISRCTN	International Standard Randomised Controlled Trial Number
ITT	Intention to Treat
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SUSAR	Suspected Unexpected Serious Adverse Reaction
TSC	Trial Steering Committee
CRF	Case Report Form
UPA	Ulipristal Acetate
LNG-IUS	Levonorgestrel-releasing intra-uterine system
EMA	European Medicine Agency
PRAC	Pharmacovigilance Risk Assessment Committee
MHRA	Medicines and Healthcare Products Regularity Authority
USM	Urgent Safety Measure
Term	**Definition**
International Standard Randomised Controlled Trial Number	A clinical trial registry
Protocol	Document that details the rationale, objectives, design, methodology and statistical considerations of the study
Randomisation	The process of assigning trial subjects to intervention or control groups using an element of chance to determine the assignments in order to reduce bias.
Statistical Analysis Plan	Pre-specified statistical methodology documented for the trial, either in the protocol or in a separate document.

TABLE OF CONTENTS

1. Introduction

This document is the Statistical Analysis Plan (SAP) for the UCON trial and should be read in conjunction with the current trial protocol. This SAP details the proposed analyses and presentation of the data for the main paper(s) reporting the results for the UCON trial.

The results reported in these papers will follow the strategy set out here. Subsequent analyses of a more exploratory nature will not be bound by this strategy, though they are expected to follow the broad principles laid down here. The principles are not intended to curtail exploratory analysis (e.g. to decide cut-points for categorisation of continuous variables), nor to prohibit accepted practices (e.g. transformation of data prior to analysis), but they are intended to establish rules that will be followed, as closely as possible, when analysing and reporting data.

Any deviations from this SAP will be described and justified in the final report or publication of the trial (using a table as shown in Appendix A). The analysis will be carried out by an appropriately qualified statistician, who should ensure integrity of the data during their data cleaning processes.

2. Background and rationale

The background and rationale for the trial are outlined in detail in the protocol. In brief, UCON is a multicentre, randomised controlled trial to determine if UPA (Ulipristal Acetate) is more effective at reducing the burden of HMB symptoms than LNG-IUS (levonorgestrel-releasing intra-uterine system) after 12 months of treatment.

In February 2018, the European Medicine Agency (EMA) released a drug alert update to say their Pharmacovigilance Risk Assessment Committee (PRAC) had reviewed the benefits and risks following some reports of serious liver injury in Europe. Use of UPA was temporarily halted in the UK and European Union. The UCON trial thus implemented an urgent safety measure (USM) on 12-Feb-2018 which suspended recruitment. Trial participants who were taking UPA were allowed to complete their existing treatment cycle should they wish, but not commence a new cycle. Liver function tests were mandated at appropriate time points.

In May and August 2018, EMA and Medicines and Healthcare Products Regularity Authority (MHRA), respectively, published restrictions and requirements around the use of UPA, including the need for additional blood tests to monitor liver function and rescinded the temporary safety measure that prompted the USM and trial suspension. Recruitment into UCON trial resumed, according to this protocol and under an amended clinical trial authorisation.

In March 2020, EMA and MHRA, respectively, published further restrictions around the use of

UPA and issued a medical recall. The result of this action meant no further recruitment into the trial, and also led to the cessation of treatment for all UPA participants

Following the halt to trial recruitment arising from these USMs, it was necessary to consider the trial analysis populations, considering the restrictions that prevented women taking subsequent course of UPA and any other new potential biases. The general principle was that a revised primary analysis population, free from as much confounding bias as possible should be agreed, and supplemented with a number of sensitivity analysis populations (see section 5.3). These analysis populations were initially proposed by the TMG and then agreed by the independent members of the Trial Steering Committee who were blind to any data accrued to that point. The trial funder (HTA) also approved the changes, following external peer review from an independent statistical expert who was also blind to any data accrued.

3. Trial objectives

The primary objective is to determine if UPA is more effective at reducing the burden of HMB symptoms after 12 months of treatment.

Secondary objectives are as follows:

- To ascertain whether UPA use beyond 3 months and up to 12 months duration is associated with histological changes to the endometrium, and if so, whether this compromises safety.
- To ascertain whether UPA is more effective than LNG-IUS in relation to menstrual blood loss, sexual activity, generic quality of life, satisfaction with treatment, patient reported adverse events, and compliance at 3, 6, and 12 months.
- To determine the response to UPA and LNG-IUS treatment difference in the presence of uterine fibroids in terms of i) alleviation of HMB and ii) change in uterine/fibroid volume.
- Collect data on liver function in women taking UPA.

4. Trial methods

4.1. Trial design

UCON is a multicentre, randomised controlled trial. Participants will be recruited from the gynaecologist, out-patient clinics of participating centres, fitting around their current service provision. Recruitment will be supported by dedicated research nurses, who will work with local gynaecology leads. Participants will be randomised in a 1:1 ratio to UPA or LNG-IUS. Participants, investigators, research midwives/nurses and other attending clinicians cannot be blinded to the treatment allocation, as the treatments are so different in route of administration. See Appendix B for trial schema.

4.2. Trial interventions

Ulipristal acetate (UPA) is the comparator. UPA is provided as a 5mg tablet. The trade name for UPA in the European Union is Esmya™ for treatment of uterine fibroids, and is marketed by Gedeon Richter.

Levonorgestrel releasing intra-uterine system (LNG-IUS) is the control group. The LNG-IUS is a contraceptive device that slowly releases a daily dose of 20 µg levonorgestrel into the uterine endometrium. It is a long acting reversible contraceptive preparation that requires removal and reinsertion approximately every three or five years, depending on the product. LNG-IUS is approved for use as a contraceptive and for HMB and in the context of the current trial is manufactured by two companies. Bayer Pharma AG market their LNG-IUS under the name of Mirena™ and Actavis UK Ltd under the name of Levosert.

4.3. Primary outcome measure

The primary outcome will be the condition-specific Menorrhagia Multi-Attribute Scale (MMAS) questionnaire[1] designed and validated to capture the impact of HMB on women's day-today life. HMB is a subjective problem and quality of life is affected by practical difficulties and the impact on social life, psychological well-being, physical health, work routine and family life. The MMAS questionnaire attempts to capture the consequences of HMB on these domains with 6 questions each with 4 levels of response. Summary scores range from 0 (worst affected) to 100 (not affected). The primary time-point for analysis will be at 12 months.

4.4. Secondary outcome measures

Secondary outcomes are as follows:

- MMAS scores measured at the other assessment points (see section 4.5).
- Menstrual bleeding, which will be captured by validated Pictorial Blood Loss Assessment Chart[2] (PBAC). The standard PBAC is a validated and well used assessment of menstrual blood loss in women. Summary scores range from 0 (amenorrhea), with increasing scores indicating worse bleeding (no upper limit). It will be used to generate the incidence of amenorrhoea (=0), light (1-10), normal (10-100) and heavy menstrual bleeding (>100).
- Cycle regularity (ordinal 4 point scale) and duration of period (ordinal 3 option scale).
- Visual analogue scales (0=best outcome, 10=worse outcome) for pelvic pain during periods, intercourse and at other times.
- Uterine Fibroid Symptom and Quality of Life (UFS-QoL) instrument[3], which contains a health related quality of life (HRQoL) domain and a symptom domain. Scores range from 0 at worst to 100 at best. This instrument will only be given to women diagnosed with fibroids.
- Sexual Activity Questionnaire, a measure of sexual functioning, used in other HMB trials[4]. The sexual activity questionnaire is a valid, reliable and acceptable measure for

describing the sexual functioning of women in terms of pleasure, discomfort and habit. Scores for pleasure range from 0 (lowest level) to 18 (highest level), scores for discomfort range from 0 (greatest) to 6 (none), and scores for habit range from 0 (worst outcome) to 3 (best outcome). It is quick and easy to administer and has good face validity delineating between the sexual functioning of pre and post-menopausal women.

- Generic Quality of Life[5] (EQ-5D-5L)
  - EQ-5D index score (-0.59=worst outcome, 1.0=best outcome).
  - EQ-5D health thermometer (0=worst outcome, 100=best outcome).
- Satisfaction with treatment outcome measured on a 5-point Likert scale.
- Participant rating of effect of treatment on HMB over 12 months measured on a 4-point Likert scale.
- Whether participant is willing to recommend the treatment to a friend (yes/no).
- Surgical intervention (hysterectomy, endometrial ablation and other gynaecological surgery).
- Adherence to trial treatments and reasons for changing treatment, as reported by the participant.
- Serious Adverse Events (SAEs) and reactions reported by participants, principally those that are serious and detailed in the respective Summary of Product Characteristics (SmPC) and those that are unexpected.
- Clinical measurements via pelvic ultrasound: uterine volume, evidence of adenomyosis, presence of fibroids, largest fibroid volume, endometrial thickness, endometrial appearance (regular/ irregular), evidence of ovarian cysts.
- Clinical measurement via endometrial biopsy: primary diagnosis (normal/ benign/hyperplasia/malignant) and further sub-diagnosis if non-normal including presence or absence of PAEC (PRM-associated endometrial change) or other non-physiological changes
- Clinical measurement via blood samples: liver function (including alanine transaminase (ALT) and asparate aminotransferase (AST) and other tests according to local protocols) once mandated by the USM
- Other blood sample measurements: serum haemoglobin and oestradiol levels.

4.5. Timing of outcome assessments

The schedule of trial procedures and outcome assessments are given in Appendix C1. Forms that have been returned late outside of the specified window (see section 5.5 for definition), will be excluded from the primary analysis but included in a sensitivity analysis (section 9.10).

4.6. Randomisation

Participants will be randomised in a 1:1 ratio to either UPA or LNG-IUS.

The Birmingham Clinical Trials Unit will provide third party web-based randomisation with

telephone back-up. A minimisation procedure using a computer-based algorithm will be used to avoid chance imbalances in the following potentially important variables:

- Age (<35, ≥35 years)
- BMI (<25 kg/m^2, ≥25 kg/m^2)
- Presence of any fibroid >2cm, as determined by the ultrasound scans
- Duration of symptoms: <1 year or ≥1 year
- Site: Individual Site
- Agreement to enter sub-study: Both/ MRI only/ Biopsy only/ Neither or N/A

To avoid any possibility of the treatment allocation becoming too predictable, a random factor will be included within the algorithm whereby allocation to the minimised treatment group will occur with probability less than one.

4.7. Sample size

The trial has been designed to be able to detect a clinically useful difference in MMAS score between the two groups at twelve months with high power. The ECLIPSE Trial[6], which evaluated the effectiveness of LNG-IUS against Standard treatment for HMB using MMAS as the primary outcome, demonstrated a difference of 13 points between the groups with a standard deviation of 24 points. This difference is considered to be clinically meaningful[6] (22) and is equivalent to approximately 0.5 standard deviations. To detect a difference of this size with 90% power (p=0.05) would require 86 women in each group (172 in total). To allow for a 20% loss to follow-up or pregnancy, the sample size has been inflated to 220 women in total.

Subsequent to the first USM notification we will aim to recruit enough women to ensure our primary analysis population (see section 5.3) is unaffected by enforced non-compliance or knowledge of the USM during the follow-up period. This means we will need to recruit 302 women in total into the study, with a target of 172 participants used in the primary analysis as per the original sample size target. 302 was calculated from the number of participants who had completed 12 month assessment prior to the first USM (89), taking into account the total number who had been randomised up to this point (198). An additional 104 participants, gaining data on 83 would be required to reach 172.

Following the second USM notification the trial was halted to new recruits on 236 participants.

4.8. Framework

The objective of the trial is to test the superiority of one intervention to another.

The null hypothesis is that there is no difference in MMAS score between the intervention groups. The alternative hypothesis is that there is a difference between the groups.

4.9. Interim analyses and stopping guidance

If UPA is overwhelmingly better or worse than LNG-IUS with respect to the primary outcome, then this may become apparent before the target recruitment has been reached. Alternatively, new evidence might emerge from other sources that UPA is definitely more, or less, effective than LNG-IUS. To protect against this, during the main period of recruitment to the study, interim analyses of the primary outcome and adverse events will be supplied, in strict confidence, to the independent Data Monitoring Committee (DMC), along with updates on results of other related studies, and any other analyses that the DMC may request.

The DMC will advise the chair of the TSC if, in their view, any of the comparisons in the trial have provided both (a) "proof beyond reasonable doubt" that for all, or for some, types of patient one particular treatment is definitely indicated or definitely contraindicated in terms of a net difference in the major endpoints, and (b) evidence that might reasonably be expected to influence the patient management of many clinicians who are already aware of the other main trial results. The TSC can then decide whether to close or modify any part of the trial. Unless this happens, however, the TMG, TSC, the investigators and all of the central administrative staff (except the statisticians who supply the confidential analyses) will remain unaware of the interim results.

Appropriate criteria of proof beyond reasonable doubt cannot be specified precisely, but a difference of at least $p<0.001$ (similar to a Haybittle-Peto[7] stopping boundary) in an interim analysis of a major endpoint may be required to justify halting, or modifying, the study prematurely. If this criterion were to be adopted, it would have the practical advantage that the exact number of interim analyses would be of little importance, so no fixed schedule is proposed.

A separate DMC reporting template will be drafted and agreed by the DMC including an agreement on which outcomes will be reported at interim analyses. The statistical methods stated in this SAP will be followed for the outcomes included in the DMC report, where possible.

4.10.<Internal> Pilot Progression Rules

Not applicable; an internal pilot was not part of the trial design.

4.11. Timing of final analysis

The final analysis for the trial will occur after all randomised women have completed primary and major secondary outcomes (up to 12 months for the patient completed questionnaires and potentially up to 18 months for endometrial biopsies) and the corresponding outcome data has been entered onto the trial database and validated as being ready for analysis. This is

provided that the trial has not been stopped early for any reason (e.g. DMC advice or funding body request); if this is the case the analysis will be completed once the existing randomised participants have completed the study (up to 18 months follow-up).

4.12. Timing of other analyses

Not applicable.

4.13. Trial comparisons

All references in this document to 'group' refer to UPA or LNG-IUS.

5. Statistical Principles

5.1. Confidence intervals and p-values

A p-value will be reported from a two-sided test at the 5% significance level for the primary outcome (in the primary population (A), see section 5.3) and Serious Adverse Events only (safety outcomes can be subject to statistical testing without adjustment for multiple testing as adjustment for multiplicity is counterproductive for considerations of safety[8]). All estimates of differences between groups for all outcomes will be presented with two-sided 95% confidence intervals, unless otherwise stated (see further point in section 5.2).

5.2. Adjustments for multiplicity

No correction for multiple testing will be made; cautious interpretation of secondary outcome confidence intervals will be necessary due to the possibility of multiplicity.

5.3. Analysis populations

As a result of the two USMs, four analysis populations have been defined (see Appendix C2 for a graphical description).

Primary population (A):

Data from questionnaire responses received before the date of the first USM (13/2/2018) in Phase 1 of the trial as well as those received between 20/12/2018 and 17/3/2020 (date of the second USM in Phase 2 following trial restart). This will include all participants, regardless of adherence to treatment, as per intention to treat (ITT) principles.

Secondary (sensitivity) populations (B1 and B2) – adherent participants only (see section 5.4 for definition):

1. Both groups: data from questionnaire responses received before 31/5/2018 in Phase 1 (to allow for completion of existing treatment course) as well as those received between 20/12/2018 and 17/3/2020 in Phase 2.
2. UPA: data from questionnaire responses received before 31/5/2018 in Phase 1 (to allow for completion of existing treatment course) as well as those received between 20/12/2018 and 17/3/2020 in Phase 2; LNG-IUS: data from all questionnaire responses until trial end.

In addition, a profile of responses over time for a selection of outcomes in the UPA group (single-group, observational **population C**) will be presented to highlight any impact of stopping treatment.

Previously agreed time windows for responses at each assessment time will still be enforced (see section 5.5).

Analysis populations will apply to a differing set of outcomes (see sections 9.5 and 9.6 for details as well as Appendix C3). All outcomes will be analysed for the primary analysis population (A), but for the other sensitivity populations (B1, B2 and C) it was considered prudent to limit these to the most important clinical outcomes to reduce the possibility of over-interpretation of data.

5.4. Definition of adherence

Adherence to allocated intervention will be monitored by participant self-report. Compliant participants in the LNG-IUS group will be those women who have had the LNG-IUS fitted following allocation and have not had the device removed at the corresponding assessment time-point. This will be self-reported on the follow-up questionnaire (LNG-IUS='Yes'). In the UPA group compliant participants will be those women who have confirmed to still be taking their allocation (UPA='Yes') and have done so at least 'Every day' or 'Most days (5-6 per week on average). This will again be recorded on the self-reported follow-up questionnaire.

5.5. Handing protocol deviations

A protocol deviation is defined as a failure to adhere to the protocol such as errors in applying the inclusion/exclusion criteria, the incorrect intervention being given, incorrect data being collected or measured, follow-up visits outside the visit window or missed follow-up visits. We will apply a strict definition of the ITT principle and will include all participants as per the primary analysis population described in section 5.3 in the analysis, in some form, regardless of deviation from the protocol.[9] This does not include those participants who have specifically withdrawn consent for the use of their data in the first instance; however these outcomes will be explored as per other missing responses (see section 9.3).

Three month follow-up questionnaires will be considered valid provided they have been

completed prior to the six month assessment time point (six months post-randomisation). Six month follow-up questionnaires will be considered valid provided they have been completed prior to the twelve month assessment time point (twelve months post-randomisation). Twelve month follow-up questionnaires will be considered valid provided they have been completed before 18 months post-randomisation. If the three and six month questionnaires are completed too late to be considered valid for that particular assessment time and subsequent questionnaires have not been returned they will be considered valid for the subsequent time-point (e.g. a six month form completed at thirteen months but with twelve month form missing will be considered valid for the twelve month time-point).

5.6. Unblinding

Not applicable, UCON is an open-label study.

6. Trial population

6.1. Recruitment

A flow diagram (as recommended by CONSORT[10]) will be produced to describe the participant flow through each stage of the trial. This will include information on the number (with reasons) of losses to follow-up (drop-outs and withdrawals) over the course of the trial. A template for reporting this is given in Appendix D1.

6.2. Baseline characteristics

The trial population will be tabulated as per Appendix D2 for **all randomised participants as well as analysis population A (for the purpose of Baseline characteristics, Population A will be based only on those participants that returned MMAS score at 12 months)**. Categorical data will be summarised by number of participants, counts and percentages. Continuous data will be summarised by the number of participants, mean and standard deviation if deemed to be normally distributed or number of participants, median and interquartile range if data are skewed, and ranges if appropriate. Tests of statistical significance will not be undertaken, nor confidence intervals presented.[11]

7. Intervention(s)

7.1. Description of the intervention(s)

Not applicable, as they were prescribed as per the protocol.

7.2. Adherence to allocated intervention

A cross-tabulation of allocated intervention by the adherence categories stated in section 5.4

will be produced (proportions and percentages) by time-point. A template for reporting adherence is given in Appendix D3.

8. Protocol deviations

Frequencies and percentages by group will be tabulated for the protocol deviations as per Appendix D4. This will not include any deviations related to enforced or non-enforced adherence.

9. Analysis methods

Intervention groups will be compared using suitable regression models to adjust for all covariates as specified in section 9.1, where possible.

9.1. Covariate adjustment

In the first instance, intervention effects between groups for all outcomes will be adjusted for the minimisation parameters listed in section 4.6. Categorised continuous variables (age and BMI) will be treated as continuous variables in this adjustment and all factors will be treated as fixed effects. Where repeated assessments (longitudinal data) are being taken into account in the model, parameters for participant, treatment group, time and baseline response (as a continuous variable) will be included. Time will be assumed to be a categorical (fixed) variable and all assessment times will be included in the model. To allow for a varying treatment effect over time, a time by treatment interaction parameter will also be included in the model.

If covariate adjustment is not possible (e.g. the model does not converge), centre will be dropped from the model in the first instance. If convergence of the model remains problematic, alternative models will be explored (e.g. an adjusted Poisson regression model with robust standard errors[12]). If this also fails to converge, unadjusted estimates will be produced. It will be made clear in the final report why this occurred (e.g. not possible due to low event rate/lack of model convergence).

9.2. Distributional assumptions and outlying responses

Distributional assumptions will be assessed visually to check if the proposed analysis method is appropriate. If responses are considered to be particularly skewed and/or distributional assumptions violated, the impact of this will be examined through sensitivity analysis; this will consist of transformation of responses prior to analysis (e.g. log transformation) in the first instance. If extreme values are apparent and considered to be affecting the integrity of the analysis, a sensitivity analysis consisting of removing the outlying response(s) and repeating the analysis will be performed. Output from these analyses, if performed, will be described and presented alongside the original analysis (or included, e.g. in appendices) with the

excluded values clearly labelled. See section 9.10 for further details regarding sensitivity analyses.

For the primary outcome (MMAS score at 12 months): upon inspection of pooled data as part of data validation processes by the trial statistician ahead of a compiling a DMC report, a high degree of skew in the responses was observed. The trial statistician concluded that a more formal reserve method for analysis should be included in the updated version of the analysis plan if the regression residuals for the originally planned method (a linear regression) are considered too skewed to proceed with. This change has been approved by an independent statistician reviewer, blind to accruing data, approving the new version of the SAP. This analysis method will consist of a generalised estimating equation (GEE) model[13] with cumulative logit link (for ordered categorical data, see below for categories) that will take into account all assessment times (i.e. correlated longitudinal data). A general 'independent' covariance structure will be assumed. Cumulative odds ratios and 95% confidence intervals for the treatment group parameter will be produced; a chi-squared test will be used to test the statistical significance (p-value produced) of the estimated treatment group parameter from the model. Responses will be categorised in the following manner: ≤50, 51-75, 76-99, =100. These categories have been used in similar trials of heavy menstrual bleeding with MMAS as the primary outcome.[14] The proportional odds assumption will be investigated by examining the odds ratios for the treatment group parameter from three binary splits of the data (0–99 vs 100, 0–75 vs 76–100, and 0–50 vs 51–100) using a similar generalised estimating equation (GEE) model but with logit link. If the parallel regression assumption is thought to be violated then we will prioritise the 0-99 vs 100 analysis as the primary cut-off point. This analysis is described in section 9.6.

9.3. Handling missing data

In the first instance, analysis will be completed on received data only with every effort made to follow up participants even after protocol violation to minimise any potential for bias. To examine the possible impact of missing data on the results, and to make sure we are complying with the intention-to-treat principle, sensitivity analysis will be performed on the primary outcome measure (MMAS at 12 months) in the primary analysis population (A). This will consist of simulating the missing responses using a multiple imputation approach.[15] A chained equations (Fully Conditional Specification) approach, incorporating a linear (or ordinal regression as appropriate) model to impute missing responses will be utilised. The imputation model will be compatible with the analytical model in terms of the parameters included. Twenty simulated data-sets will be produced. The regression analysis (as described below) will then be performed on each set of data with the results combined using Rubin's rules to obtain a single set of results (treatment effect estimate and confidence interval).

9.4. Data manipulations

The Trial Statistician will derive all responses from the raw data recorded in the database.

Primary Outcome**:**

- **MMAS questionnaire summary score at 12 months.**

CRF: Follow-up Questionnaire
Question:

- SECTION A: Impact of your periods on your day to day life

To obtain the summary score, the scores from each response (from all six domains) will be added together (each response carries a unique weight[1]).

If some participants decline to complete the MMAS on the grounds they are no longer having periods their score will be assumed to be maximum (MMAS=100). This will not include any participant who were not having periods because they were pregnant, reached menopause or declined to give a reason (indicated on the follow-up questionnaire), these will be treated as missing responses.

Baseline Characteristics:

- **Any fibroids>2cm**

Fibroids >2cm:
CRF: Ultrasound Form
Question:

- 3 - Dimensions of Largest Fibroids
  - Were fibroids seen on ultrasound? = 'Yes'
    AND EITHER
  - Dimensions of largest fibroid (cm): Longitudinal = '>2cm'
    OR
  - Dimensions of largest fibroid (cm): Transverse = '>2cm'
    OR
  - Dimensions of largest fibroid (cm): Anteroposterior = '>2cm'

Fibroids ≤2cm:
CRF: Ultrasound Form
Question:

- 3 - Dimensions of Largest Fibroids
  - Were fibroids seen on ultrasound? = 'Yes'

AND

- Dimensions of largest fibroid (cm): Longitudinal = '≤2cm'
  AND
- Dimensions of largest fibroid (cm): Transverse = '≤2cm'
  AND
- Dimensions of largest fibroid (cm): Anteroposterior = '≤2cm'

No Fibroids:
CRF: Ultrasound Form
Question:

- 3 - Dimensions of Largest Fibroids
  - Were fibroids seen on ultrasound? = 'No'
    OR
  - Were fibroids seen on ultrasound? = Missing/blank

Adherence**:**

- **Adherence to allocated intervention.**

For participants allocated to LNG-IUS group:
CRF: Follow-up Questionnaire
Question:

- SECTION E: Questions about any treatments you have had or are now taking
  - 1. What medical treatment(s) do you take for your heavy menstrual bleeding OR as contraception? Indicate as many as applicable = LNG-IUS (Coil) = Yes

For participants allocated to UPA group:
CRF: Follow-up Questionnaire
Question:

- SECTION E: Questions about any treatments you have had or are now taking
  - 1. What medical treatment(s) do you take for your heavy menstrual bleeding OR as contraception? Indicate as many as applicable = Ulipristal tablets = Yes
    AND
  - Question: Please can you describe how frequently you take (or did take) the tablets = 'Every Day' **OR** 'Most days (5-6 days per week on average)'

For Adherence to allocated intervention table:

- **No – due to USM (personal preference)**

Only relevant to those participants allocated to UPA group:
CRF: Follow-up Questionnaire
Question:

- SECTION E: Questions about any treatments you have had or are now taking
    - 1. What medical treatment(s) do you take for your heavy menstrual bleeding OR as contraception? Indicate as many as applicable = Ulipristal tablets = No

AND

- Date completed on Follow up questionnaire between 13/2/18 and 31/5/2018

- **No – due to USM (enforced)**

Only relevant to those participants allocated to UPA group:
CRF: Follow-up Questionnaire
Question:

- SECTION E: Questions about any treatments you have had or are now taking
    - 1. What medical treatment(s) do you take for your heavy menstrual bleeding OR as contraception? Indicate as many as applicable = Ulipristal tablets = No

AND

- Date completed after 31/5/2018 as participant told to come off OR any participants randomised after 20/12/18 and returned a questionnaire after 17/3/20 (Second USM again told to come off).

Secondary Outcomes**:**

- **MMAS scores measured at the Baseline, 3 Months and 6 Months.**

CRF: Follow-up Questionnaire
Question:

- SECTION A: Impact of your periods on your day to day life

To obtain the summary score, the scores from each response (from all six domains) will be added together (each response carries a unique weight[1]).

If some participants decline to complete the MMAS on the grounds they are no longer having periods their score will be assumed to be maximum (=100). This will not include any participant who were not having periods because they were pregnant, reached menopause or declined to give a reason (indicated on the follow-up questionnaire), these will be treated as missing responses.

- **Menstrual bleeding (PBAC[2]).**

CRF: Follow-up Menstrual Blood Loss Diary

Where diary is filled in with some missing responses, set these missing responses to 0 to allow the total score to be calculated.

Multiply the numbers given by scores stated below. Add the scores (tampon and sanitary towels) together in order to obtain total score.

<u>For version 1</u>
The scores for tampons assigned:
lightly stained = 1
moderately soiled = 5
completely saturated with blood = 10
Flooding = 5
1p sized clot = 1
50p sized clot = 5
Golf ball sized clot = 5

The scores for sanitary towels assigned:
lightly stained = 1
moderately soiled = 5
completely saturated with blood = 20
Flooding = 5
1p sized clot = 1
50p sized clot = 5
Golf ball sized clot = 5

<u>For version 2</u>
The scores for tampons assigned:
lightly stained = 1
moderately soiled = 5
completely saturated with blood = 10
Flooding = 5
<2cm sized clot = 1
2cm sized clot = 1
3cm sized clot = 5
>3cm sized clot = 5

The scores for sanitary towels assigned:

lightly stained = 1
moderately soiled = 5
completely saturated with blood = 20
Flooding = 5
<2cm sized clot = 1
2cm sized clot = 1
3cm sized clot = 5
>3cm sized clot = 5

If some participants decline to complete the Menstrual Blood Loss Diary on the grounds they are no longer having periods their score will be assumed to be equal to 0 (i.e. no bleeding). This will not include any participant who were not having periods because they were pregnant, reached menopause or declined to give a reason (indicated on the follow-up questionnaire), these will be treated as missing responses.

- **Cycle regularity**

CRF: Follow-up Questionnaire
Question:
- SECTION B: About your periods and pelvic pain at present;
  - Question 2: How regular is your cycle?

The responses from this question will fulfil the outcome.

- **Duration of period**

CRF: Follow-up Questionnaire
Question:
- SECTION B: About your periods and pelvic pain at present;
  - Question 3: What is the average duration of your period?

The responses from this question will fulfil the outcome.

- **Visual analogue scales (for pelvic pain during periods, intercourse and at other times)**

<u>Pelvic pain during periods</u>
CRF: Follow-up Questionnaire
Question:
- SECTION B: About your periods and pelvic pain at present;
  - Question 4: Do you experience pelvic pain during your periods? = Yes

The number indicated on the scale will fulfil the outcome.

Pelvic pain during intercourse

CRF: Follow-up Questionnaire

Question:

- SECTION B: About your periods and pelvic pain at present;
  - Question 5: Do you experience pelvic pain during intercourse? = Yes

The number indicated on the scale will fulfil the outcome.

Pelvic pain at other times

CRF: Follow-up Questionnaire

Question:

- SECTION B: About your periods and pelvic pain at present;
  - Question 6: Do you experience pelvic pain at any other times (other than during period or during intercourse)? = Yes

The number indicated on the scale will fulfil the outcome.

- **Uterine Fibroid Symptom and Quality of Life (UFS-QoL) instrument[3] (only given to women diagnosed with fibroids).**

CRF: Follow-up Questionnaire

Question:

- SECTION C: Impact of fibroids on your day to day life

Domains are scored as followed:

Concern=AnxDur+SoilUnd+StainBed+Extra+SoilOut;
Activities=uAnxTrav+uPhys+uDecEx+uDiffAct+uIntSoc+uPlan+uEmbarr;
EnergyMood=uTired+uDrows+uSad+uDown+uExhaus+uIrrat+uWeak;
Control=uNotCon+uLessProd+uHealth+uFuture+uControl;
SelfCon=uWeight+uStomach+uSize;
SexualFunc=uDesire+uRelation;

HRDQL Domain score:

HRQL=Concern+Activities+EnergyMood+Control+SelfCon+SexualFunc;

Symptom Severity Domain score:
SymptomRaw=uHeavy+uClot+uDur+uLen+uTight+uUriDay+uUriNight+uFat;

Total score of each domain:

HRQLConcern=((25-Concern)/20)*100;

HRQLAct=((35-Activities)/28)*100;
HRQLEnergy=((35-EnergyMood)/28)*100;
HRQLControl=((25-Control)/20)*100;
HRQLSelfCon=((15-SelfCon)/12)*100;
HRQLSexual=((10-SexualFunc)/8)*100;
HRQLScore=((145-HRQL)/116)*100;
SymptomScore=((SymptomRaw-8)/32)*100;

- **Sexual Activity Questionnaire[4]**

CRF: Follow-up Questionnaire
Question:

- SECTION D: Sexual Activity

Scores are calculated for only those participants that answered 'yes' to the following question:
3. Do you engage in sexual activity with anyone at the moment?

Domains are scored as follows:

PLEASURE = q1+q2+q4+q7+q8+q10;

DISCOMFORT = q5+q6;

HABIT= q9;

The weightings applied to each question are as follows:

Pleasure
(Questions 1, 2, 4, 7, 8, 10):
Very much = 3
Somewhat = 2
A little = 1
Not at all = 0

Pleasure
(Questions 8):
5 times or more = 3
3-4 times = 2
1-2 times = 1
Not at all = 0

The 'pleasure' score ranges between 0 and 18, a low score representing low pleasure.

Discomfort
(Questions 5, 6) :
Very much = 0
Somewhat = 1
A little = 2
Not at all = 3

The 'discomfort' score ranges between 0 and 6, a low score representing high discomfort.

Habit
(Question 9):
Much more = 3
Somewhat more = 2
About the same = 1

- **Generic Quality of Life[5] (EQ-5D-5L)**

CRF: Follow-up Questionnaire
Question:

- SECTION F: General Quality of Life

The EQ5D (5 level) will be scored using the Crosswalk index value calculator (found in the following folder: K:\BCTU\BCTU\Statistics\SOPs\SAS Code Depository\Euroqol EQ5D\EQ5D-5L\Mapping EQ-5D-5L to EQ-5D-3L)

The EuroQol Group coordinated a study that administered both the 3-level and 5-level versions of the EQ-5D, in order to develop a "crosswalk" between the EQ-5D-3L value sets and the new EQ-5D-5L descriptive system, resulting in crosswalk value sets for the EQ-5D-5L

A scientific publication by Van Hout et al. (2012) describing the mapping methodology behind the study in detail is published in Value In Health Journal (https://pubmed.ncbi.nlm.nih.gov/22867780/ - van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard AS. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012 Jul-Aug;15(5):708-15. doi: 10.1016/j.jval.2012.02.008. Epub 2012 May 24. PMID: 22867780).

- **Satisfaction with treatment outcome measured on a 5-point Likert scale.**

CRF: End of Study Form

Question:

- SECTION 6 – Patient Satisfaction with Treatment
    - Question 6.1: Since taking the treatment, how did the patient rate their satisfaction with treatment outcomes?

The responses from this question will fulfil the outcome.

- **Participant rating of effect of treatment on HMB over 12 months measured on a 4-point Likert scale.**

CRF: End of Study Form
Question:

- SECTION 6 – Patient Satisfaction with Treatment
    - Question 6.2: Compared to 12 months ago, when you started the treatment, would you say that your heavy menstrual bleeding has: (please tick one):

The responses from this question will fulfil the outcome.

- **Whether participant is willing to recommend the treatment to a friend. (yes/no)**

CRF: End of Study Form
Question:

- SECTION 6 – Patient Satisfaction with Treatment
    - Question 6.3: Would the participant recommend the treatment to a friend?

The responses from this question will fulfil the outcome.

- **Surgical intervention (hysterectomy, endometrial ablation and other gynaecological surgery).**

CRF: SECTION E: Questions about any treatments you have had or are now taking
Question:

- SECTION E: Questions about any treatments you have had or are now taking
    - Question 6: Since you last completed a UCON questionnaire (upon entering the trial before 3mth/ 3mth /6mth /1 year) have you been to hospital? = Yes
        - Question: If 'YES', was this for a gynaecological (women's health) reason = Yes
            - Question: If YES, you have been to hospital for a gynaecological reason since you last completed a UCON questionnaire, what was this for?= 'hysterectomy' OR 'endometrial ablation' OR 'other'

The responses from this question will fulfil the outcome.

- **Adherence to trial treatments and reasons for changing treatment, as reported by the participant.**

Time to first treatment change

CRF: Follow-up Questionnaire

Question:

- SECTION E: Questions about any treatments you have had or are now taking
  - Question 2: Have the treatment(s) that you take for your heavy menstrual bleeding changed since you last completed a UCON questionnaire (upon entering the trial / 3mth/ 6mth/ 1 year ago)? = 'Yes'
  - Question: If YES, when did you change treatment?

CRF: CARF

Question:

- PART 11: Allocated Treatment
  - Question: Date of LNG-IUS Fitting

OR

CRF: Follow-up Menstrual Blood Loss Diary

Question:

- PART 11: Allocated Treatment
  - Question: Date Ulipristal Acetate Started

  OR
  - Question: Date LNG-IUS (Coil) fitted

Time to first treatment change = 'If YES, when did you change treatment?' – 'Date Ulipristal Acetate Started' OR 'Date LNG-IUS (Coil) fitted'

Reasons for changing treatment, as reported by the participant

CRF: Follow-up Questionnaire

Question:

- SECTION E: Questions about any treatments you have had or are now taking
  - Question 2: Have the treatment(s) that you take for your heavy menstrual bleeding changed since you last completed a UCON questionnaire (upon entering the trial / 3mth/ 6mth/ 1 year ago)? = 'Yes'
  - Question: If YES, why have you changed your treatment?

The responses from this question will fulfil the outcome.

- **Serious Adverse Events (SAEs) and reactions reported by participants, principally those that are serious and detailed in the respective**

CRF: Serious Adverse Event Form
Question:
- SECTION 2. EVENT DETAILS
  - Question: Description of SAE
  - Question: Seriousness Criteria

- **Clinical measurements via pelvic ultrasound**

Uterine volume
CRF: Ultrasound Form
Question:
- SECTION 2: Dimensions of Uterus
  - Question: Volume of uterus

Evidence of adenomyosis
CRF: Ultrasound Form
Question:
- SECTION 2 - Dimensions of Uterus
  - Question: Was there evidence of adenomyosis? = 'Yes'

Presence of fibroids
CRF: Ultrasound Form
Question:
- SECTION 3 - Dimensions of Largest Fibroids
  - Question: Were fibroids seen on ultrasound? = 'Yes'

Largest fibroid volume
CRF: Ultrasound Form
Question:
- SECTION 3 - Dimensions of Largest Fibroids
  - Question: Volume of largest fibroid

Endometrial thickness
CRF: Ultrasound Form
Question:
- SECTION 4 – Endometrial thickness

- Question: Endometrial thickness

Endometrial appearance (regular/ irregular)
CRF: Ultrasound Form
Question:

- SECTION 4 – Endometrial thickness
  - Question: Endometrial appearance

Evidence of ovarian cysts
CRF: Ultrasound Form
Question:

- SECTION 5 – Ovarian cysts
  - Question: > 2cm ovarian cyst seen? = 'Yes'

- **Clinical measurement via endometrial biopsy: primary diagnosis (normal/ benign/hyperplasia/malignant) and further sub-diagnosis if non-normal including presence or absence of PAEC (PRM-associated endometrial change) or other non-physiological changes**

CRF: Local Endometrial Biopsy Evaluation Form
Question:

- SECTION: Endometrial Biopsy Details
  - Question: Biopsy Result (please tick only one)

AND

CRF: Local Endometrial Biopsy Evaluation Form
Question:

- SECTION: Repeat Endometrial Biopsy Details
  - Question: Biopsy Result (please tick only one)

- **Clinical measurement via blood samples: liver function (including alanine transaminase (ALT) and asparate aminotransferase (AST) and other tests according to local protocols) once mandated by the USM**

CRF: Liver Function Test (LFT) Form
Question:

- SECTION 2: LFT Results (for each of the four LFTs associated with the Course No. stated in Section

- **Other blood sample measurements: serum haemoglobin and oestradiol levels.**

Serum haemoglobin
CRF: End of Study Form
Question:

- SECTION 3 – Blood Test
  - Question 3.1: Blood sample taken: Hb = 'Yes'
  - Question: Hb Result

Oestradiol
CRF: End of Study Form
Question:

- SECTION 3 – Blood Test
  - Question 3.1: Blood sample taken: Oestradiol = 'Yes'
  - Question: Oestradiol Result

9.5. Analysis methods – primary outcome(s)

A template for reporting the primary outcome is given in Appendix D5. **Analysis will be conducted in populations A, B1 and B2 with population A considered the primary population. Responses over time will be presented for population C.**

Adjusted mean differences between group means and associated 95% confidence intervals at the 12 month time point will be calculated using a mixed linear regression model for repeated measures.[16] An F-test will be used to test the statistical significance (p-value produced) of the estimated treatment group parameter generated from the restricted maximum likelihood estimates. See section 9.1 for parameters to be included in the model.

9.6. Analysis methods – secondary outcomes

A template for reporting the secondary outcomes is given in Appendix D6. Point estimates and 95% confidence intervals will be presented for all estimates; p-values will not be presented. **Analysis will be conducted in populations A for all outcomes as well as B1, B2 and C where stated below.**

Analysis will be performed as described in section 9.5 for the following patient reported summary scores: MMAS at 3 and 6 months (see also section 9.2), VAS scores, UFS-QoL scores, SAQ scores, EQ-5D-5L index score, EQ-5D-5L health state.

For the proportion of patients reporting any surgical intervention, a logistic regression model will be used to generate adjusted odds ratios along with 95% confidence intervals. Patients having a surgical intervention will initially be those undergoing either a hysterectomy or an ablation. The same analysis will then be repeated, including any other gynaecological surgery

that patients might undergo; events will only be counted once for those with more than one surgery recorded, but it will be clarified in the tables where this has occurred (**analysis also conducted in populations B1 and B2**).

For MMAS scores we will analyse the proportion of patients who have score equal to 100 (maximum score; no symptoms) using a generalised estimating equation (GEE) model with logit link that will take into account all assessment times (correlated longitudinal data). An 'independent' covariance structure will be assumed. Odds ratios and 95% confidence intervals for the treatment group parameter will be produced (**also in populations B1 and B2**).

For the menstrual bleeding loss, the PBAC scores will be dichotomized as i) amenorrhoea (bleeding score=0) and 'any bleeding' (bleeding score >0), and ii) non-heavy (bleeding score<=100) and heavy (bleeding score>100) (**also conducted in populations B1 and B2; responses over time will also be presented for population C**). For cycle regularity the responses will be dichotomized as i) regular (patients who reported "Regular, I know when to expect my period" and "Fairly regular, my periods starts within a few days of when I expect") and ii) irregular ("Irregular, I cannot predict when my period will start" and "I have bleeding on and off all the time"). These outcomes will be analysed in a similar fashion to the repeated binary responses described in the previous paragraph.

Duration of period ("1-3 days", "4-6 days", "more than 6 days") will be analysed using the methodology described in section 9.2 for repeated ordinal data. Cumulative odds ratios and 95% confidence intervals will be generated. Statistical significance of the treatment group parameter will be determined (p-value generated) through examination of the associated chi-squared statistic.

Satisfaction with treatment and participant rating of treatment are both measured on a Likert scale at a single time point and will be analysed using ordinal logistic regression. Cumulative odds ratios and 95% confidence intervals will be produced. Whether the participant is willing to recommend the treatment to a friend is measured at a single time point only and will be analysed as per the number of surgical interventions (i.e. will be treated as a binary outcome). Odds ratios and 95% confidence intervals will be produced.

Continuous outcomes measured via pelvic ultrasound/blood samples such as uterine volume, largest fibroid volume, endometrial thickness, serum haemoglobin and oestradiol levels are measured once at the end of the study and will be analysed using a linear regression model. Difference between group means and associated 95% confidence intervals will be produced. Data will be checked for skew and if necessary subject to suitable transformation (e.g. log-transformation and presented with geometric means). The remainder of the clinical measurements via pelvic ultrasound will be analysed in a similar fashion to the binary outcomes measured at a single time point described above. (**analysis also conducted in populations B1 and B2**).

Endometrial biopsy and liver function output (**UPA only; timing related to treatment length so treatment populations do not apply**) will be tabulated using appropriate summary statistics.

9.7. Analysis methods – exploratory outcomes and analyses

Any data that does not form a pre-specified outcome will be presented using simple summary statistics by intervention group (i.e. numbers and percentages for binary data and means (or medians) and standard deviations (or inter-quartile ranges) for continuous normal (or non-normal) data.

9.8. Safety data

The number and percentage of participants experiencing any adverse events, serious adverse events (SAEs) and suspected unexpected serious adverse reactions (SUSARs) will be presented by intervention group. Statistical significance will be determined by chi-squared test. No other formal analysis is anticipated due to the low anticipated frequency of events. The total number of SAEs in each group will also be given along with a descriptive table of the events. A template for reporting this safety data is given in Appendix D7 (**analysis conducted in populations A, B1 and B2**).

9.9. Planned subgroup analyses

Due to the reduced sample size as a result of the Urgent Safety Measures subgroup analysis will not be conducted as they are likely to be highly underpowered and not informative.

9.10.Sensitivity analyses

Sensitivity analyses will be limited to the primary outcome and will consist of:

- Assessment of effect of missing data on the primary outcome as described in section 9.3; **this will be conducted in population A**
- Heterogeneity of treatment effect over recruitment period (phase 1 and 2 in Appendix C2) will be explored by including a treatment by phase interaction parameter to the primary analysis model; **this will be conducted in population A**
- Including scores for those questionnaires returned late outside of the agreed window (see section 5.5); **this will be conducted in populations A, B1 and B2**

10. Analysis of sub-randomisations

Not applicable.

11. Health economic analysis

No health economic analysis is planned for this trial.

12. Statistical software

SAS software, version 9.4 (or higher) or STATA version 14 (or higher) will be used for all analyses.

13. References

1. Shaw RW, Brickley MR, Evans L, Edwards MJ. Perceptions of women on the impact of menorrhagia on their health using multi-attribute utility assessment. Br J Obstet Gynaecol.1998; 105(11):1155-9.
2. Higham JM, O'Brien PM, Shaw RW. Assessment of menstrual blood loss using a pictorial chart. Br J Obstet Gynaecol. 1990; 97(8):734-9.
3. Spies JB, Coyne K, Guaou Guaou N, Boyle D, Skyrnarz-Murphy K, Gonzalves SM. The UFSQOL, a new disease-specific symptom and health-related quality of life questionnaire for leiomyomata. Obstet Gynecol. 2002; 99(2):290-300.
4. Thirlaway K, Fallowfield L, Cuzick J. The Sexual Activity Questionnaire: a measure of women's sexual functioning. Qual Life Res. 1996; 5(1):81-90.
5. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011; 20(10): 1727-36.
6. Gupta J, Kai J, Middleton L, Pattison H, Gray R, Daniels, J for the ECLIPSE Trial Collaborative Group. Levonorgestrel Intrauterine System versus Medical Therapy for Menorrhagia. N Engl J Med 2013; 368:128-137.
7. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV et al. Design and analysis of randomised clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer. 1976;34:585-612.13. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. BMJ. 2011; 342:d40.
8. Huque MF, Dmitrienko A and D'Agostino R. Multiplicity issues in clinical trials with multiple objectives. Statistics in Biopharmaceutical Research. 2013 5:4,321-337.
9. Gupta SK. Intention-to-treat concept: A review. Perspect Clin Res. 2011; 2(3): 109-112.
10. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010; 340:c332.
11. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. Lancet. 1990; 335:149–153.
12. Zou G. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004;159(7):702-6.
13. Liang KY, Zeger SL. Longitudinal Data Analysis using Generalized Linear Models. Biometrika. 73:13-22.

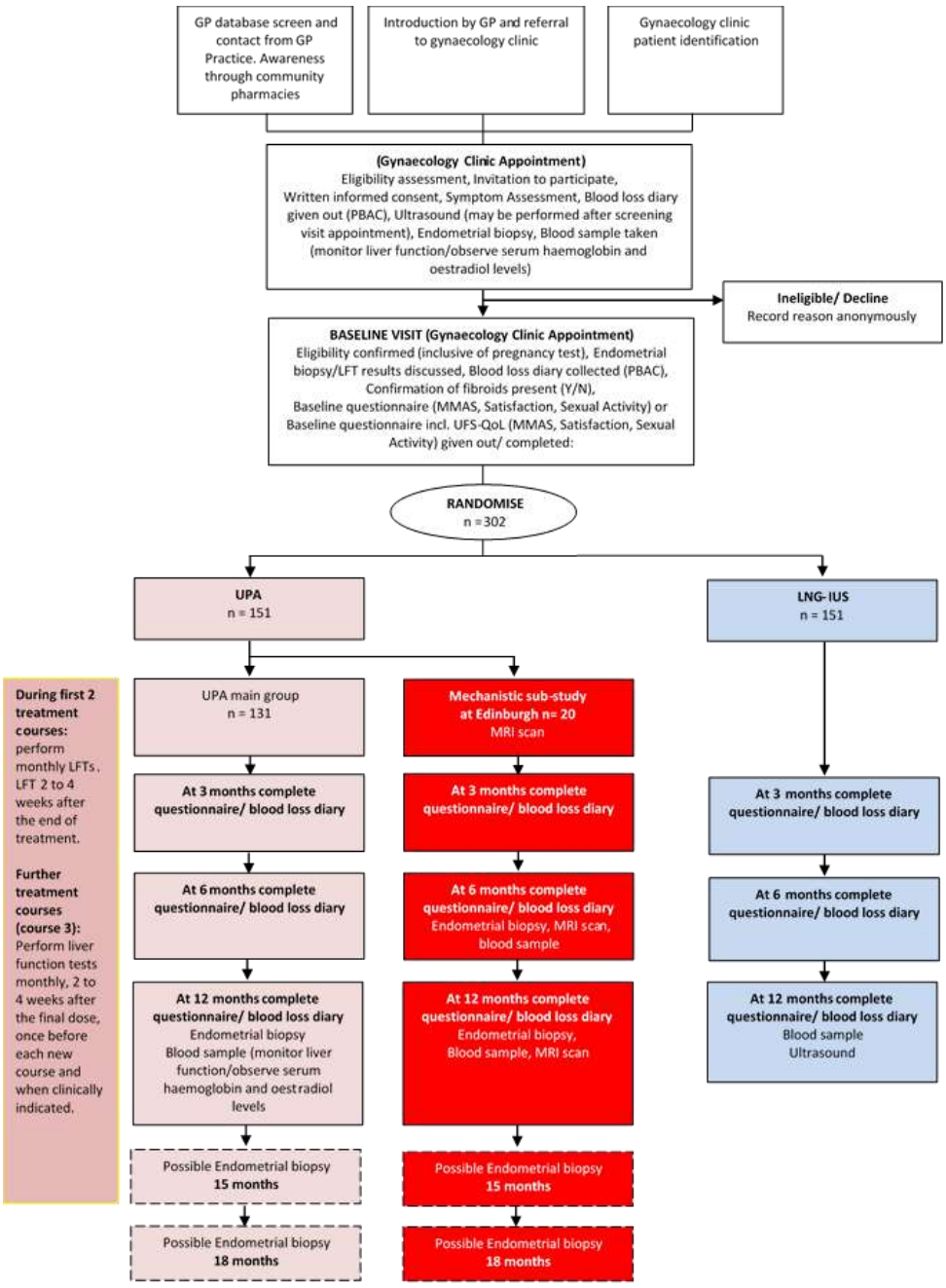
14. Cooper K, Breeman S, Scott NW, Scotland G, Clark J, Hawe J, et al. Laparoscopic supracervical hysterectomy versus endometrial ablation for women with heavy menstrual bleeding (HEALTH): a parallel-group, open-label, randomised controlled trial. Lancet. 2019 Oct 19;394(10207):1425-1436.
15. Rubin DB. Multiple imputation for nonresponse in surveys. Wiley, 1987.
16. Brown H, Prescott R. Applied Mixed Models in Medicine, 2nd Edition. Wiley, 2006.

Appendix A: Deviations from SAP

This report below follows the statistical analysis plan dated *<insert effective date of latest SAP>* apart from following:

Section of report not following SAP	Reason
<insert section >	<insert, e.g. exploratory analyses request by TMG>

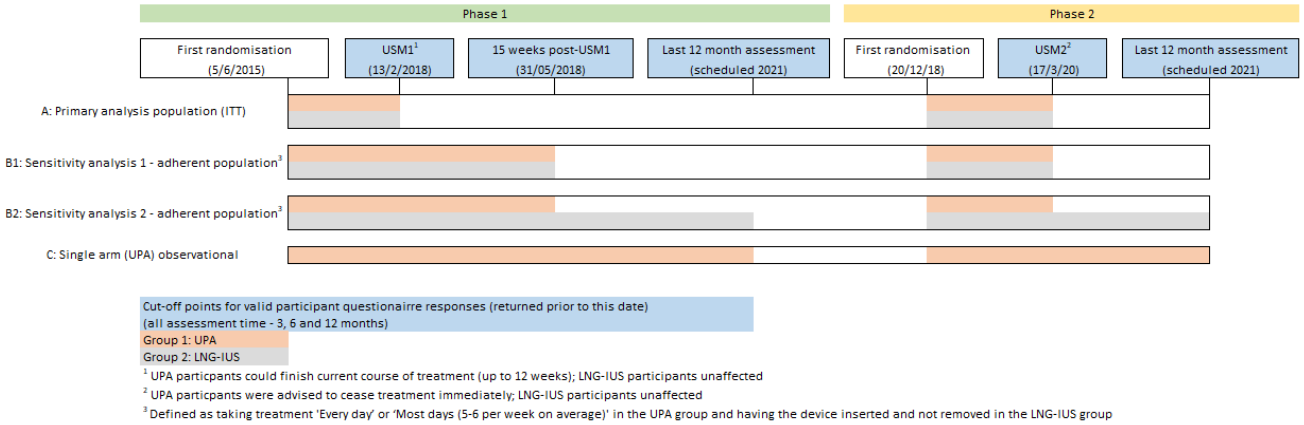
Appendix B: Trial schema



Appendix C1: Schedule of assessments

Timepoint	Screening (1)	Baseline (2)	3 months (approx) (3)	6 months (approx) (4)	12 months (approx) (5)	Post-treatment 1 (6)	Post-treatment 2 (7)
Written informed consent	X						
Liver function tests***	X		X	X	X		
Patient questionnaires (MMAS, UFS-QOL, EQ-5D-5L, ICE-CAP, SAQ)		X	X	X	X		
Other patient reported outcomes (compliance, adverse events, willingness to recommend to a friend, rating of treatment, satisfaction of treatment)					X		
Menstrual bleeding diary	X		X	X	X		
Blood sample, to observe haemoglobin and oestradiol levels (not safety bloods)	X				X		
Ultrasound pelvic assessment	X				X		
Endometrial biopsy	X				X UPA only **		
Endometrial biopsy – additional for women in UPA group who exhibit PAEC						((X)) UPA only	((X)) UPA only
Follow up outpatient appointment to discuss post-trial treatment options						(X) UPA only	
Liver Function Test (if indicated)			X***				

Appendix C2: Analysis populations



Cut-off points for valid participant questionairre responses (returned prior to this date)
(all assessment time - 3, 6 and 12 months)
Group 1: UPA
Group 2: LNG-IUS

[1] UPA particpants could finish current course of treatment (up to 12 weeks); LNG-IUS participants unaffected

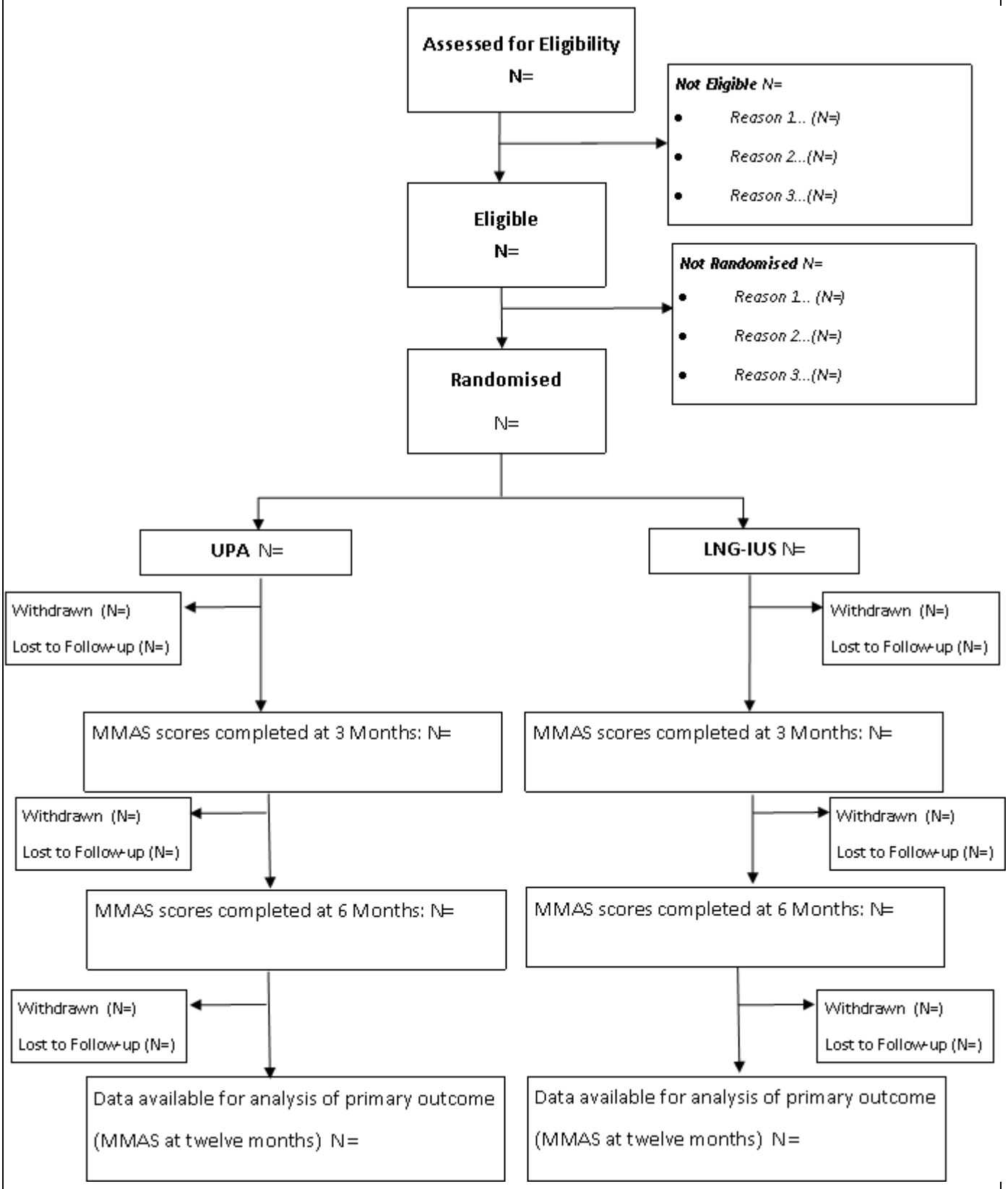
[2] UPA particpants were advised to cease treatment immediately; LNG-IUS participants unaffected

[3] Defined as taking treatment 'Every day' or 'Most days (5-6 per week on average)' in the UPA group and having the device inserted and not removed in the LNG-IUS group

Appendix C3: Outcome measured to be examined by population

	A (Primary)	B1 (Sensitivity 1)	B2 (Sensitivity 2)	C (Observational)
MMAS	✓	✓	✓	✓
MMAS sensitivity analysis for missing data	✓			
MMAS sensitivity analysis heterogeneity of effect between recruitment phases	✓			
MMAS sensitivity analysis for late data	✓	✓	✓	
MMAS score=100 (maximum)	✓	✓	✓	
PBAC (including amenorrhea and heavy bleeding outcomes)	✓	✓	✓	✓
Cycle regularity	✓			
Cycle duration	✓			
VAS score	✓			
UFS-QoL scores	✓			
SAQ scores	✓			
EQ-5D scores	✓			
Surgical interventions	✓	✓	✓	
Treatment satisfaction/rating of treatment	✓			
Clinical measurements (including uterine volume and blood sample measures)	✓	✓	✓	
Biopsy output	UPA group only – timing related to treatment length			
SAEs	✓	✓	✓	
Liver function data	UPA group only – timing related to treatment length			

Appendix D1: CONSORT flow diagram



Appendix D2: Baseline characteristics

All randomised participants and population A (for the purpose of Baseline characteristics, Population A will be based only on those participants that returned MMAS score at 12 months).

		UPA (N=)	LNG-IUS (N=)	Overall (N=)
Age[1]	≤35 years	N (%)	N (%)	N (%)
	>35 years	N (%)	N (%)	N (%)
	Mean (SD)			
BMI[1]	≤25 kg/m^2	N (%)	N (%)	N (%)
	>25 kg/m^2	N (%)	N (%)	N (%)
	Mean (SD)			
Duration of symptoms (months)[1]	<1 year	N (%)	N (%)	N (%)
	≥1 year	N (%)	N (%)	N (%)
	Median [IQR], n			
Any fibroids>2cm[1]	Yes	N (%)	N (%)	N (%)
	No	N (%)	N (%)	N (%)
Centre[1]	1…	N (%)	N (%)	N (%)
	2…	N (%)	N (%)	N (%)
	3…	N (%)	N (%)	N (%)
	4…	N (%)	N (%)	N (%)
Agreement to enter sub-study[1]	Both MRI	N (%)	N (%)	N (%)
	MRI only	N (%)	N (%)	N (%)
	Biopsy only	N (%)	N (%)	N (%)
	Neither/not applicable	N (%)	N (%)	N (%)
Ethnicity	White	N (%)	N (%)	N (%)
	Mixed	N (%)	N (%)	N (%)
	Asian	N (%)	N (%)	N (%)
	Black	N (%)	N (%)	N (%)
	Other ethnic group	N (%)	N (%)	N (%)
	Not stated	N (%)	N (%)	N (%)
	Missing	N (%)	N (%)	N (%)
Number of times the patient has been pregnant	Median [IQR], n			
Result of pregnancy	Live birth	N (%)	N (%)	N (%)
	Still birth	N (%)	N (%)	N (%)
	Termination	N (%)	N (%)	N (%)
	Miscarriage/ ectopic	N (%)	N (%)	N (%)
	Missing	N (%)	N (%)	N (%)
Route of deliveries	Vaginal	N (%)	N (%)	N (%)
	Caesarean	N (%)	N (%)	N (%)
	Forceps/ ventouse	N (%)	N (%)	N (%)

	Missing	N (%)	N (%)	N (%)
Previous treatments for HMB	Mefenamic Acid/ NSAIDs	N (%)	N (%)	N (%)
	Tranexamic Acid	N (%)	N (%)	N (%)
	Combined Oral Contraceptive	N (%)	N (%)	N (%)
	Progesterone Only Pill	N (%)	N (%)	N (%)
	Norethisterone	N (%)	N (%)	N (%)
	Depo-Provera (medroxyprog acetate)	N (%)	N (%)	N (%)
	Implant (Nexplanon/ Implanon)	N (%)	N (%)	N (%)
	Ulipristal Acetate	N (%)	N (%)	N (%)
	LNG-IUS	N (%)	N (%)	N (%)
	Missing	N (%)	N (%)	N (%)
Previous Surgical treatments	Surgical termination	N (%)	N (%)	N (%)
	Surgical management of miscarriage	N (%)	N (%)	N (%)
	Uterine Acetate	N (%)	N (%)	N (%)
	Missing	N (%)	N (%)	N (%)
Number of fibroids	0	N (%)	N (%)	N (%)
	1	N (%)	N (%)	N (%)
	2	N (%)	N (%)	N (%)
	>2	N (%)	N (%)	N (%)
	Missing	N (%)	N (%)	N (%)
Volume of largest fibroid (ml)	Median [IQR], n			
	Missing	N (%)	N (%)	N (%)

Appendix D3: Adherence to allocated intervention

	Adherent to treatment allocation	UPA (N=)	LNG-IUS (N=)
LNG-IUS fitted/UPA prescription administered	Yes	N (%)	N (%)
	No	N (%)	N (%)
	Missing	N (%)	N (%)
3 months: LNG-IUS retained/compliant to UPA treatment schedule	Yes	N (%)	N (%)
	No – due to USM (personal preference)	N (%)	N (%)
	No – due to USM (enforced)		
	No – other reasons		
	Missing	N (%)	N (%)
6 months: LNG-IUS retained/compliant to UPA treatment schedule	Yes	N (%)	N (%)
	No – due to USM (personal preference)	N (%)	N (%)
	No – due to USM (enforced)		
	No – other reasons		
	Missing	N (%)	N (%)
12 months: LNG-IUS retained/compliant to UPA treatment schedule	Yes	N (%)	N (%)
	No – due to USM (personal preference)	N (%)	N (%)
	No – due to USM (enforced)		
	No – other reasons		
	Missing	N (%)	N (%)

USM=Urgent Safety Measure

Reasons for non-adherence are listed below:

	UPA (N=)	LNG-IUS (N=)
Lack of effectiveness	N (%)	N (%)
Did not control my bleeding	N (%)	N (%)
Irregular bleeding	N (%)	N (%)
Prolonged bleeding	N (%)	N (%)
Coil expulsion	N (%)	N (%)
Pelvic infection	N (%)	N (%)

Disliked treatment	N (%)	N (%)
Tummy upset of nausea	N (%)	N (%)
Disliked taking tablets	N (%)	N (%)
Wanted to get pregnant	N (%)	N (%)
Skin allergy	N (%)	N (%)
Depression/mood swings	N (%)	N (%)
Weight gain	N (%)	N (%)
Thread problems	N (%)	N (%)
Headaches/migraine	N (%)	N (%)
Dizziness	N (%)	N (%)
Hypertension/increased blood pressure	N (%)	N (%)
Pelvic pain	N (%)	N (%)
Other side effects	N (%)	N (%)

Details of first treatment change over 12 months

	UPA (N=)	LNG-IUS (N=)
UPA	N (%)	N (%)
LNG-IUS	N (%)	N (%)
Mefenamic acid	N (%)	N (%)
Tranexamic acid	N (%)	N (%)
Depo-provera injection	N (%)	N (%)
Contraceptive pill	N (%)	N (%)
No treatment	N (%)	N (%)
Other medical treatment[1]	N (%)	N (%)
TOTAL	**N (%)**	**N (%)**

[1] Details

Appendix D4: Protocol deviations

	UPA (N=)	LNG-IUS (N=)
Ineligible patients randomised	N (%)	N (%)
Other protocol deviations	N (%)	N (%)
…	N (%)	N (%)

Appendix D5: Primary outcome results

Populations A, B1 and B2. Responses over time will be presented for population C.

	UPA Mean (SD), n	**LNG-IUS** Mean (SD), n	Mean Difference (95% CI)[2]	p-value
Baseline				
3 months				
6 months				
12 months[3]				

[1] Menorrhagia multi-attribute scale questionnaire; score ranges from 0 (not affected) to 100 (worst affected)

[2] Difference>0 favour UPA.

[3] Primary outcome time-point

Population A (and where stated in Appendix C3 B1 and B2)

Primary outcome (MMAS) sensitivity analysis (12 months only)

	UPA Mean (SD), n	**LNG-IUS** Mean (SD), n	Mean Difference (95% CI)[2]
Sensitivity analysis 1[1]			
Sensitivity analysis 2[1]			
Sensitivity analysis 3[1]			
Sensitivity analysis …			

[1] Difference>0 favour UPA.

Populations A, B1 and B2

Primary outcome (MMAS) proportion of maximum responses (score=100; 12 months only)

	UPA n (%)	**LNG-IUS** n (%)	Odds ratio[1] (95%CI)
3 months			
6 months			
12 months			

[1] Difference>1 favour UPA.

Appendix D6: Secondary outcomes results

Populations A, B1 and B2. Responses over time will be presented for population C.

PBAC bleeding score

	UPA n (%)	**LNG-IUS** n (%)	Odds ratio[1] (95%CI)
Baseline			
Amenorrhea (=0)			
Light (1-10)			
Normal (>10-100)			
Heavy (>100)			
Median score [IQR]			
TOTAL			
3 month			
Amenorrhea (=0)			
Light (1-10)			
Normal (>10-100)			
Heavy (>100)			
Median score [IQR]			
TOTAL			
6 Month			
Amenorrhea (=0)			
Light (1-10)			
Normal (>10-100)			
Heavy (>100)			
Median score [IQR]			
TOTAL			
12 Month			
Amenorrhea (=0)			
Light (1-10)			
Normal (>10-100)			
Heavy (>100)			
Median score [IQR]			
TOTAL			

[1] Relative risk for any bleeding (amenorrhea+light+normal) shown; estimates<1 favour UPA. Relative risk for 'heavy' bleeding also shown; estimates<1 favour UPA

Population A

Cycle regularity questions

	UPA n (%)	**LNG-IUS** n (%)	Odds ratio[1] (95%CI)
Baseline			
Regular			

Fairly regular			
Irregular			
Bleeding on and off			
TOTAL			
3 Months			
Regular			
Fairly regular			
Irregular			
Bleeding on and off			
TOTAL			
6 Month			
Regular			
Fairly regular			
Irregular			
Bleeding on and off			
TOTAL			
12 Month			
Regular			
Fairly regular			
Irregular			
Bleeding on and off			
TOTAL			

[1] Odds ratio for 'irregular' bleeding shown (irregular + bleeding on and off); estimates<1 favour UPA.

Population A

Cycle duration questions

	UPA n (%)	**LNG-IUS** n (%)	Odds ratio[1] (95%CI)
Baseline			
1-3 days			
4-6 days			
More than 6 days			
TOTAL			
3 Months			
1-3 days			
4-6 days			
More than 6 days			
TOTAL			
6 Month			
1-3 days			
4-6 days			
More than 6 days			

TOTAL			
12 Month			
1-3 days			
4-6 days			
More than 6 days			
TOTAL			

[1] Odds ratio from proportional odds model shown; estimates<1 favour UPA (shorted cycle length)

Population A

Questionnaire responses

	UPA Mean (SD), n	**LNG-IUS** Mean (SD), n	Mean Difference (95% CI)
Visual Analogue Scale (VAS)[1]			
Pain during periods			
Baseline			
3 months			
6 months			
12 months			
Pain during intercourse			
Baseline			
3 months			
6 months			
12 months			
Pain at any other time			
Baseline			
3 months			
6 months			
12 months			
UFS-QoL[2]			
Symptom domain			
Baseline			
3 months			

6 months			
12 months			
HRQL domain			
Baseline			
3 months			
6 months			
12 months			
SAQ			
Pleasure domain[3]			
Baseline			
3 months			
6 months			
12 months			
Discomfort domain[4]			
Baseline			
3 months			
6 months			
12 months			
Habit domain[5]			
Baseline			
3 months			
6 months			
12 months			
Euroqol			
EQ-5D-5L[6]			
Baseline			
3 months			
6 months			
12 months			
Health thermometer[7]			
Baseline			
3 months			
6 months			
12 months			
Patient Satisfaction with Treatment (12 months)			
Extremely satisfied			
Unsatisfied			
Neither satisfied or			

Unsatisfied			
Satisfied			
Extremely satisfied			
Participant rating of effect of treatment on HMB (12 months)			
Got much better			
Got a little better			
Not changed much			
Got worse			
Participant recommend the treatment to a friend (12 months)			
Yes			
No			

[1] Scores range from 0 (best outcome) to 10 (worse outcome); scores<0 favour UPA

[2] Uterine fibroid symptom and health-related quality of life questionnaire (only given to women with fibroids); scores range from 0 (worst outcome) to 100 (best outcome); scores>0 favour UPA

[3] Sexual Activity Questionnaire pleasure scores range from 0-18, where low scores are bad and high scores are good; scores>0 favour UPA

[4] Sexual Activity Questionnaire discomfort scores range from 0-6, where low scores are bad and high scores are good; scores>0 favour UPA

[5] Sexual Activity Questionnaire habit scores range from 0-3, where low scores are bad and high scores are good; scores>0 favour UPA

[6] EQ-5D-5L quality of life scores range from -0.59 (worse outcome) to 1.00 (best outcome); scores>0 favour UPA

[7] scores range from 0 (worse outcome) to 100 (best outcome); scores>0 favour UPA

[8] ICECAP-A scores range from 0 (worse outcome) to 1.0 (best outcome)

Populations A, B1 and B2

Surgical interventions over 12 months

	UPA n (%)	**LNG-IUS** n (%)	Relative risk[1] (95%CI)	p-value
Endometrial ablation				
Hysterectomy				
Other gynaecological surgery[2]				
TOTAL				

[1] Relative risk for endometrial ablation+hysterectomy combined and all operations shown; estimates<1 favour UPA.

[2] Details

Clinical measurements via pelvic ultrasound/blood samples at 12 months

	UPA	**LNG-IUS**		
	Mean (SD), n	Mean (SD), n	Mean Difference (95% CI)[1]	p-value
Uterine volume[2] (ml)				

Volume of largest fibroid[2] (ml)				
Endometrial thickness (mm)				
Haemoglobin (g/l)				
Oestradiol levels (pmmol/l)				
	n (%)	n (%)	Relative risk[3] (95%CI)	p-value
Evidence of adenomyosis				
Presence of fibroids				
Irregular endometrial appearance				
Evidence of ovarian cysts (>2cm)				

[1] Estimates<0 favour UPA

[2] Volume=Longitudinal (cm) x Transverse (cm) x Anteroposterior (cm) x 0.523

[3] Estimates<1 favour UPA

Endometrial biopsy output (UPA group only – timing related to stopping treatment; not population specific)

Diagnosis	n (%)
Initial	
Normal	
Insufficient	
Benign	
Hyperplasia	
Malignant	
If benign, evidence of PAEC	
Repeat (further 3 months)	
Normal	
Insufficient	
Benign	
Hyperplasia	
Malignant	
If benign, evidence of PAEC	
Repeat (further 6 months)	
Normal	
Insufficient	
Benign	
Hyperplasia	
Malignant	
If benign, evidence of PAEC	

Liver function output (UPA group only – timing related to stopping treatment; not population specific)

	n (%)	n (%)
	During treatment period	**Post-treatment period**
Number who have had LFT testing		
a) Number with test result outside local normal range in any test[1] at any time		
b) Number with clinically significant results in any test[1] at any time		
c) Number with transaminase levels>3 times upper limit of normal in any test[1] at any time		
d) Number with both b) and c) in any test at any time		

[1]AST, ALT, ALP, Bilirubin, GGT

Appendix D7: Safety

	UPA n (%)	**LNG-IUS** n (%)	**p-value**
Total number of SAEs	n	n	
Total number of participants experiencing an SAE	n (%)	n (%)	<insert p-value>
Total number of SUSARs	n	n	
Total number of participants experiencing an SUSAR	n (%)	n (%)	<insert p-value>

<A line by line listing of each SAE may be appropriate for some studies and can be included in an Appendix to the main report; the table below provides a guide for how the line listing of SAE data could be presented.>

Summary of SAE	**Reason for Reporting**	**Causality**	**Action taken**
UPA			
1 <insert description of SAE>			
2			
3			
4			
LNG-IUS			
1			
2			
3			
4			