STATISTICAL ANALYSIS PLAN

*Ba*riatric surgery vs. *M*edical care for o*b*esity and polycystic ovar*i*an syndrome related i*n*fertil*i*ty:

The BAMBINI randomised-controlled clinical trial

Principal Investigator

Professor Alexander Dimitri Miras

Sponsor

Imperial College London

23rd November 2022

Department of Investigative Medicine

Imperial College London

6th Floor Commonwealth Building

Imperial College London at Hammersmith Campus

Du Cane Road, London W12 0NN, UK

Tel: +44 (0)20 8383 3242

Fax: +44 (0)20 8383 8320

TABLE OF CONTENTS

1. Abbreviations 4
Study Design 5
2.1 Interim Analysis 5
2.2 Final Analyses and Reporting 5
2.2.1 Final Analysis 5
2. Study Objectives and Endpoints 6
3.1 Study Objective 6
3.2 Study Endpoints 6
3.2.1 Primary Efficacy Endpoint 6
3.2.2 Secondary Efficacy Endpoints 6
5 Analysis Populations 7
5.1 Intention to Treat Population (ITT) 7
5.2 Per-Protocol Population (PP) 7
5.3 Safety Population 7
6 General Issues for Statistical Analysis 7
6.1 Analysis Software and GENERAL METHODS OF ANALYSIS 7
6.2 Disposition of Subjects and Withdrawals 7
6.3 Methods for Missing Data 8
6.4 Protocol Violations 8
6.5 Multiple ENDPOINT ADJUSTMENT 8
7 Demographics and Other Baseline Characteristics 8
7.1 Demographics 8
7.2 Baseline Medical History 8
7.3 Baseline LABORATORY MEASUREMENTS 9
8 Efficacy Analyses 9
8.1 Primary Outcome 9
8.2 Secondary efficacy outcomes 9
9 Safety Analyses 10
9.1 Primary Safety outcome 10
9.2 Secondary Safety outcomes 10
9.3 Adverse Events 10
9.3.1 All Adverse Events 10

9.3.2 Adverse Events Leading to Withdrawal 10
9.3.3 Serious Adverse Events 11
10 Other Planned Analyses 11
10.1 Planned Subgroup Analyses 11
11 REFERENCES 11

1. ABBREVIATIONS

Abbreviation	Definition
PCOS	Polycystic ovary syndrome
BMI	Body mass index
VSG	Vertical sleeve gastrectomy
SAP	Statistical analysis plan
RCT	Randomised controlled trial
HbA1c	Glycated haemoglobin
HOMA-IR	Homeostatic Model Assessment for Insulin Resistance
TT	Total testosterone
SHBG	Sex hormone binding globulin
FAI	Free androgen index
DHEAS	Dehydroepiandrosterone sulfate
AMH	Anti-Mullerian Hormone
LH	Luteinising hormone
FSH	Follicle-stimulating hormone
OGTT	Oral glucose tolerance test
QoL	Quality of Life
HADS	Hospital anxiety and depression score
PCOS-Q	Polycystic ovary syndrome-questionnaire
FGS	Ferriman-Gallwey score
ITT	Intent-To-Treat Population
PP	Per-Protocol Population
GLM	Generalised linear model
CSR	Clinical study report
SAE	Serious Adverse Event
TEAE	Treatment emergent adverse event
AE	Adverse Event
SOC	System organs class
PT	Preferred Term
CRF	Case Report Form

STUDY DESIGN

This is an open-label randomised-controlled clinical trial to investigate the safety and efficacy of obesity surgery vs medical care for women with polycystic ovary syndrome (PCOS), obesity and oligo- or amenorrhoea.

Patients will be randomised at a ratio of 1:1, stratified by body mass index (BMI) and trial site to:

- Medical care: a combination of lifestyle modification and pharmacotherapy *or*
- Obesity surgery: standard laparoscopic vertical sleeve gastrectomy (VSG).

Key Inclusion criteria

- Pre-menopausal women ≥18 years old
- BMI ≥ 35 kg/m^2 with obesity-related complications
- Diagnosis of PCOS based on international evidence-based guidelines for the assessment and management of PCOS 2018

Key Exclusion criteria

- Type 1 or Type 2 diabetes mellitus
- Specific contraindications to obesity surgery
- Previous obesity surgery
- Inability to maintain adequate contraception
- Medications affecting reproductive function (e.g. oral steroids, hormonal contraceptives) at screening or 3 months previously.
- Other causes of anovulation (e.g. untreated hypothyroidism, adrenal or pituitary disorders)
- Current pregnancy or breastfeeding

Key outcomes

Reproductive

- Weekly serum progesterone
- Monthly profile of reproductive hormones
- Self-reported menses

Metabolic

- Oral glucose tolerance test measuring glucose and insulin secretion and sensitivity indices. This will take place pre-intervention, at 26 weeks and 52 weeks post-intervention.
- Monthly metabolic profile: fasting glucose, insulin, and HOMA-IR.

2.1 INTERIM ANALYSIS

There will be no interim analysis for this study.

2.2 FINAL ANALYSES AND REPORTING

2.2.1 FINAL ANALYSIS

Our study will involve collecting data from all participants for 52 weeks. Once this follow-up period is complete, the database will be locked, and no further modifications will be made to it. The final analysis of the primary and secondary outcomes, as outlined in the protocol and this statistical analysis plan (SAP), will only be performed using the locked database. In addition to presenting results at the main

time point of interest (52 weeks), findings will also be presented at the 26-week mark, along with the baseline measurements.

2. STUDY OBJECTIVES AND ENDPOINTS

3.1 STUDY OBJECTIVE

To perform an randomised controlled trial (RCT) comparing the safety and efficacy of obesity surgery vs medical care for women with PCOS, obesity and oligo or amenorrhea.

3.2 STUDY ENDPOINTS

3.2.1 PRIMARY EFFICACY ENDPOINT

Primary efficacy outcome:

Number of ovulatory cycles within the 52-week follow-up period. Ovulation will be defined as a rise in serum progesterone ≥16 nmol/L.

3.2.2 SECONDARY EFFICACY ENDPOINTS

For each secondary endpoint, temporal changes, mean levels and peak levels will be analysed as appropriate:

- Anthropometric measurements – body composition, waist circumference, body weight
- Metabolic outcomes – glycated plasma haemoglobin (HbA1c), plasma liver function tests, plasma lipid profile
- Arterial blood pressure
- Reproductive hormones – serum luteinising hormone (LH), follicle-stimulating hormone (FSH), oestradiol, sex hormone binding globulin (SHBG), testosterone, free androgen index (FAI), dehydroepiandrosterone sulfate (DHEAS), androstenedione, Anti Mullerian hormone (AMH)
- Number of reported menses
- Glucose concentrations at the oral glucose tolerance test (OGTT)
- Hospital Anxiety and Depression Scale (HADS) score
- Multidimensional Health Profile: Health Functioning questionnaire score
- PCOS Health-Related Quality of Life score
- Modified Ferriam-Galwey hirsutism score
- Ludwig visual score
- Savin alopecia scale score
- Cardiff Acne Disability index
- Number of medications

4 Sample Size

Based on the available data on the effect size of lifestyle interventions[1] and obesity surgery[2] on ovulation we estimated that women in the standard medical care group will have a mean of 7 and women in the obesity surgery group a mean of 10 ovulatory cycles in the 52-week follow-up period. With a standard deviation of 3.3 around both means, we will need 33 women in each group to have a 95% power to detect statistically significant differences between the groups at α of 0.05. We will recruit 40 patients in each group to account for a 15-20% drop-out rate based on rates in similar trials we have conducted in this field (LIPOS UTN: U1111-1126-3292 and DOMINO ISRCTN76278694).

5 Analysis Populations

5.1 Intention to Treat Population (ITT)

We will recruit 80 women, with each group comprising forty participants randomised at a 1:1 ratio. The intention-to-treat (ITT) population will include all patients who were initially assigned to either medical care or obesity surgery, regardless of any deviations from the trial protocol or the actual treatment they received. The analysis of this population will be based on their original assignment group.

5.2 Per-Protocol Population (PP)

The Per-Protocol (PP) analysis population includes the patients who completed their trial follow up at 52 weeks with no major deviations to the treatment protocol.

5.3 Safety Population

We expect the rate of crossovers in our study to be low. We will analyse the safety population using the "as-treated" approach, which means that participants will be categorised based on the treatment groups they actually received. The safety population will comprise all randomised participants who underwent obesity surgery or medical care, and had at least one outcome measurement or clinical visit after randomisation.

6 General Issues for Statistical Analysis

6.1 Analysis Software and GENERAL METHODS OF ANALYSIS

We will use SPSS (version 27, IL, USA) for all analyses. Two-sided P-values < 0.05 will be considered statistically significant.

- *Primary outcome:* Number of ovulatory cycles
    - The primary outcome was considered a Poisson distributed variable. Thus, we will fit a generalised linear model (GLM) with the Poisson distributional family and a log link function.

- *Secondary efficacy outcomes:*
    - Continuous outcomes: We will fit mixed-effects linear models (random intercept models). In the fixed-effects parts of the models, we will include group (1 = obesity surgery, 0 = medical care) and time as a continuous outcome. Number of reported menses within the 52-week follow-up period: We will use an identical approach compared to the primary outcome, namely a Poisson regression model via GLM and bootstrap Student's t test.

- *Secondary safety outcomes:* Safety outcomes will be evaluated via Fisher's exact tests at 52 weeks.

6.2 Disposition of Subjects and Withdrawals

All subjects who provide written informed consent will be accounted for. The number and percentage of ITT subjects who discontinued the study prior to Week 52 will be presented by treatment group, overall and by reason of discontinuation (adverse event, discontinued by investigator, withdrawn consent/request to terminate, lost-to-follow-up, death, other). Percentages will be based on the number of ITT subjects.

6.3 Methods for Missing Data

All efforts will be made to prevent the occurrence of missing data. Nevertheless, it is anticipated that withdrawals will occur and hence there will be missing data on primary and secondary efficacy endpoints. As outlined above, we will use Poisson regression, incorporating all available data, even for participants lost to follow-up. In addition, we will use mixed-effects modelling, which naturally accounts for missing data assuming that data are missing at random. The number of participants with missing data per variable and reasons will be reported as recommended[3].

6.4 Protocol Violations

Protocol violations will be summarised in the clinical study report (CSR). This summary will include the number and percent of subjects with each violation type. Major violations in this study may be those that are related to:

- Informed consent deviation
- Inclusion/Exclusion criteria, if such protocol violation is likely to impact one of the two primary endpoints
- Participant not complying with trial protocol

The main reason for assessing the incidence of major violations during the study is to determine which patients are in the per-protocol population (the per-protocol population excludes “major” violations). Prior to database lock, all protocol violations will be reviewed and patients who have had major violations will be noted and excluded from the per-protocol population.

6.5 Multiple ENDPOINT ADJUSTMENT

There will be no adjustment for the multiple secondary endpoints. All secondary outcomes will be considered exploratory.

7 Demographics and Other Baseline Characteristics

7.1 Demographics

Demographics will be summarised by randomised treatment groups. There will be no formal statistical comparisons between treatment groups on demographic variables. We will present variables with a normal distribution as mean (standard deviation). Variables with a skewed distribution will be summarised as median (interquartile range). Categorical variables will be presented as numbers (percentages).

7.2 Baseline Medical History

The medical history of all ITT and the safety analysis population subjects will be summarised in a table by treatment group. Specifically, the number and per cent of subjects who currently have the condition will be presented for each condition.

7.3 BASELINE LABORATORY MEASUREMENTS

A table presenting descriptive statistics (sample size, mean, standard deviation, median, min and max) of laboratory variables by treatment group at baseline will be provided for the ITT analysis set. If the baseline value is missing for a given variable and patient, the screening value will be used before calculating the descriptive statistics.

8 EFFICACY ANALYSES

8.1 PRIMARY OUTCOME

Number of ovulatory cycles

- We will fit a GLM with the Poisson distributional family and a log link function. The Poisson model will be the primary analysis. Results will be presented as incidence rates and incidence rate ratios (95% confidence interval). Standard techniques will be employed to verify the model assumptions, which include residual analysis, overdispersion testing, and influence diagnostics. In case of substantial overdispersion, we may utilise a negative binomial model. As a sensitivity analysis, we will conduct a bootstrap t-test for independent samples. The analysis will include all participants who have complete follow-up data. We will use bias-corrected and accelerated (BCa) intervals via 10.000 replicates.

8.2 SECONDARY EFFICACY OUTCOMES

- Number of reported menses within the 52-week follow-up period: We will use an identical approach compared to the primary outcome, namely a Poisson regression model via GLM and bootstrap Student's t test.

The following secondary efficacy outcomes will be compared between treatment groups using a mixed-effects regression modelling:

- Anthropometric measurements – body composition, waist circumference, body weight
- Metabolic outcomes – glycated plasma haemoglobin (HbA1c), plasma liver function tests, plasma lipid profile
- Arterial blood pressure
- Reproductive hormones – serum luteinising hormone (LH), follicle-stimulating hormone (FSH), oestradiol, sex hormone binding globulin (SHBG), testosterone, free androgen index (FAI), dehydroepiandrosterone sulfate (DHEAS), androstenedione, Anti Mullerian hormone (AMH)
- Number of reported menses
- Glucose concentrations at the oral glucose tolerance test (OGTT)
- Hospital Anxiety and Depression Scale (HADS) score
- Multidimensional Health Profile: Health Functioning questionnaire score
- PCOS Health-Related Quality of Life score

- Modified Ferriam-Galwey hirsutism score
- Ludwig visual score
- Savin alopecia scale score
- Cardiff Acne Disability index
- Number of medications

We will fit mixed-effects linear models (random intercept models). In the fixed-effects parts of the models, we will include group (1 = obesity surgery, 0 = medical care) and time as a continuous outcome. We will use likelihood ratio tests to determine if there is an improvement in the model fit when including a treatment-by-time interaction, as evaluated using the Akaike information criterion. We will use a cutoff of a difference greater than 2 in AIC to indicate a significant improvement in the model.

9 Safety Analyses

9.1 Primary Safety outcome

The primary safety outcome will be the incidence risk of treatment-emergent serious adverse events (SAEs) through the 52-week follow-up period. We will evaluate this outcome via a Fisher's exact test. Results will be presented as numbers of patients with the event (percentage).

9.2 Secondary Safety outcomes

We will use Fisher's exact tests to analyse all secondary safety outcomes (described in more detail below). The results will be reported numbers of patients with the event and percentages by treatment group.

9.3 Adverse Events

We will assess the number of participants with any treatment-emergent SAEs, any treatment-emergent adverse events (TEAEs), any-cause withdrawals/dropouts and the number of participants per severity grade. The Clavien-Dindo classification for grading the severity of complications will be used. Pregnancy rates will also be assessed and considered an adverse event.

9.3.1 All Adverse Events

The total number of TEAEs and the number and percentage of subjects with at least one TEAE will be presented by treatment groups. A TEAE is an event starting or worsening in severity at or after initiation of the index procedure for the randomised treatment. For subject counts, subjects experiencing a given event more than once will be counted only once for that event.

A listing of all adverse events will include the subject number, adverse event (AE) number, the investigator description of the AE, the AE system organs class (SOC) and preferred term (PT), the severity of AE, whether or not the AE is classified as serious (SAE), the relationship of the AE to the procedure, the action taken, the outcome, and the adjudication status.

9.3.2 Adverse Events Leading to Withdrawal

A summary of number of TEAEs and of the incidence risk rates (number and percentage of subjects) of TEAEs leading to study withdrawal, by SOC and PT will be presented in a similar manner as

discussed above (with the exception of the plot). A data listing of TEAEs leading to withdrawal will also be provided, displaying details of the event(s) captured on the case report form (CRF).

9.3.3 Serious Adverse Events

Summaries of serious TEAEs will be conducted in the same manner as for all TEAEs discussed above.

9.3.4 Deaths

If a death occurs during the trial, relevant information (including study day of death relative to index initiation, cause of death, and adverse event leading to death) will be supplied in a data listing.

10 Other Planned Analyses

10.1 Planned Subgroup Analyses

The planned subgroup analyses aim to explore the differential treatment effect of obesity surgery compared to medical care on the primary outcome:

1. Baseline body mass index (BMI). This analysis will evaluate the treatment effect in patients with different baseline BMI values. Patients will be divided into two groups based on their baseline BMI, with one group having BMI <45 kg/m^2 and the other having BMI ≥45 kg/m^2.
2. Baseline Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) index. This analysis will evaluate the treatment effect in patients with different levels of insulin resistance. Patients will be divided into two groups based on their HOMA-IR value, with one group having HOMA-IR scores <4 and the other having HOMA-IR scores ≥4.
3. Baseline Free Androgen Index (FAI). This analysis will evaluate the treatment effect in patients with different androgen levels. Patients will be divided into two groups based on their FAI value, with one group having an FAI <7 and the other having an FAI ≥7.
4. Age at randomization. This analysis will evaluate the treatment effect on patients of different age groups. Patients will be divided into two groups based on age: one group aged <30 years and the other aged ≥30 years.

For the above-mentioned analyses, we will use the same statistical methods described in section 8.

11 REFERENCES

1. Moran, L.J., et al., *Dietary composition in restoring reproductive and metabolic physiology in overweight women with polycystic ovary syndrome.* J Clin Endocrinol Metab, 2003. **88**(2): p. 812-9.
2. Skubleny, D., et al., *The Impact of Bariatric Surgery on Polycystic Ovary Syndrome: a Systematic Review and Meta-analysis.* Obes Surg, 2016. **26**(1): p. 169-76.
3. Sterne, J.A., et al., *Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls.* Bmj, 2009. **338**: p. b2393.

Approved and signed on 23rd November 2022

……..

Principal Investigator

Professor Alexander Dimitri Miras