Imperial College London

STATISTICAL ANALYSIS PLAN

Investigation of the metabolic effects of DuOdenal resurfacing on insulin resistant woMen wIth polycystic ovariaN syndrOme

The DOMINO Trial

Principal Investigator

Dr Alexander Dimitri Miras

Sponsor

Imperial College London

14th February 2019

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

Abbreviation	Definition
AE	Adverse Event
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
BBA	Boston Biomedical Associates
BMI	Body Mass Index
CRF	Case Report Forms
CSR	Clinical Study Report
DHEAS	Dehydroepiandrostenedione
DMR	Duodenal Mucosal Resurfacing
FSH	Follicle-Stimulating Hormone
HbA1c	Glycated Hemoglobin
HOMA-IR	Homeostatic Model Assessment – Insulin Resistance
ITT	Intent-To-Treat Population
LH	Luteinizing Hormone
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-To-Treat Population
NHS	National Health System
NIH	National Institutes of Health
PCOS	Polycystic Ovary Syndrome
PP	Per-Protocol Population
PT	Preferred Term
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SHBG	Sex Hormone Binding Globulin
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
UADE	Unanticipated Adverse Device Effect
USS	Ultrasound Scan

2. STUDY DESIGN

This is a randomised double-blinded sham-controlled prospective investigation of women with PCOS, insulin resistance and oligo/amenorrhoea.

- 1:1 randomised, double blinded (subject and endocrinologist) trial comparing DMR treatment to sham procedure
- In addition to the allocated procedure, all patients will be provided with standard NHS lifestyle advice for 6 months.

Key Inclusion criteria
- Diagnosis of PCOS based on the NIH criteria
- Insulin resistance as defined by a 2 hour oral glucose tolerance test glucose concentration of 7.8 mmol/l and/or HOMA-IR $\geq$ 3.0
- BMI $\geq$ 30 Kg/m^2

Key Exclusion criteria
- Other causes of anovulation
- More than 6 menses in the previous 12 months
- medications affecting insulin sensitivity
- pregnant or breastfeeding at screening or 6 months previously
- Type 2 diabetes mellitus

Key Assessments
Reproductive
- Hormone profile: plasma/serum reproductive hormones
- Weeks 0 to 24: Self-reported menses
- Weeks 12 to 24:
    - Weekly USS to track development of ovarian follicles
    - Measure serum progesterone 7-10 days later (i.e. mid-luteal phase)

Metabolic
- Oral glucose tolerance test to measure glucose and indices of insulin secretion and sensitivity. Will take place pre-intervention, within 2 weeks and then at 3 months post-intervention
- Insulin clamp: hepatic and peripheral insulin sensitivity will be measured using the gold-standard two-step euglycaemic hyperinsulinaemic clamp combined with a constant infusion of [6, 6 $^2H_2$] glucose pre and at 3 months post intervention

- Unblinding to occur after the last patient reaches the 24 week visit.

2.1 INTERIM ANALYSIS

There will be no interim analysis for this study.

2.2 FINAL ANALYSES AND REPORTING

2.2.1 FINAL DOUBLE-BLIND ANALYSIS

The data through 24 weeks will be locked and all final planned analyses of primary and key secondary endpoints identified in the protocol and in this SAP, will be performed after the last subject has completed the 24-week visit.

3. STUDY OBJECTIVES AND ENDPOINTS

3.1 STUDY OBJECTIVE

To demonstrate the efficacy of the Fractyl DMR Procedure using the Revita System compared to a sham procedure for the treatment of women with PCOS, insulin resistance and oligo/amenorrhoea.

0 - 24 Week Double-Blind Phase Objective: To study the effect of DMR on mechanistic and clinical endpoints 24 weeks post-procedure.

3.2 STUDY ENDPOINTS

3.2.1 PRIMARY EFFICACY ENDPOINT

Primary Efficacy:

1. The change from baseline in total insulin sensitivity at 12 weeks. Total insulin sensitivity is the sum of hepatic and peripheral insulin sensitivity.

2. The number of menses during 24 weeks

3.2.2 SECONDARY EFFICACY ENDPOINTS

- Change in hepatic insulin sensitivity from baseline at 12 weeks
- Change in peripheral insulin sensitivity from baseline at 12 weeks
- Number of ovulatory cycles defined by an increase in serum progesterone and / or ultrasound evidence of ovulation followed by menstrual bleeding between weeks 12-24
- Change in fasting and post-prandial concentrations of glucose, insulin and c-peptide from baseline at 2 weeks
- Change in fasting and post-prandial concentrations of glucose, insulin and c-peptide from baseline at 12 weeks
- % body weight loss from baseline at 24 weeks

3.2.3 EXPLORATORY ENDPOINTS

There will be no formal treatment comparisons on the following exploratory endpoints.

- Change from baseline to week 24/ For each endpoint, temporal changes, mean levels and peak levels will be analysed:
    - Plasma lipid concentration
    - Plasma liver function tests
    - Arterial blood pressure
    - Serum LH
    - Serum FSH
    - Serum Oestradiol
    - Serum SHBG
    - Serum Testosterone
    - Serum free androgen index
    - Serum DHEAS
    - Serum Androstenedione
    - Energy expenditure
    - Apnoea hypopnoea index
    - Body composition

- Change from week 12 to week 24/ For each endpoint, temporal changes, mean levels and peak levels will be analysed:
  - Endometrial thickness
  - Ovarian volume
  - Follicle number
  - Diameter of largest follicle in each ovary

4 SAMPLE SIZE

Assumptions of effect size for the primary efficacy endpoints in the treatment arm was derived from previous publications in which insulin sensitising medications were administered in similar groups of women [1, 2]. The assumption made was that the insulin sensitising effect of DMR would be similar to the effect observed with those medications.

It is assumed that:

(a) a difference in mean change in total insulin sensitivity between treatment and control of 5.6 μmol/kg.min at 12 weeks with equal variance in both groups (standard deviation of 4.5). Total insulin sensitivity is the sum of hepatic and peripheral insulin sensitivity.

(b) a difference in the number of menses between treatment and control of 1.0 over 24 weeks with equal variance in both groups (standard deviation of 1.0)

The weighted Hochberg procedure to adjust for multiple endpoints is described in Section 6.5. Under this procedure 24 randomised subjects (12 per group) provides at least 94% power that the benefit of DMR treatment over sham will be found for at least one primary endpoint when testing using an overall one sided 0.050 significance level, and provides at least 88% power that the benefit of DMR treatment over sham will be found for at least one primary endpoint when testing using an overall one sided 0.025 significance level.

Thirty patients will be randomised to account for potential patients lost to follow up prior to the primary endpoint assessment. Patients who for technical reasons cannot have the DMR will be replaced.

5 ANALYSIS POPULATIONS

5.1 INTENT TO TREAT POPULATION (ITT)

The intent-to-treat (ITT) population for this study includes all randomised subjects.

5.2 MODIFIED INTENT TO TREAT POPULATION (MITT)

The mITT population includes all randomised subjects in whom the study procedure (DMR or sham) is attempted and who have a baseline measurement for at least one primary endpoint. The procedure is attempted when all endoscopic exclusion criteria are verified, the catheter is introduced into the subject, and at least one ablation is performed. Subjects will be analysed according to their randomised group assignment. The mITT population is the primary analysis population for both the primary and secondary efficacy endpoints.

5.3 PER-PROTOCOL POPULATION (PP)

The Per-Protocol (PP) analysis population includes the subset of mITT subjects who received the treatment to which they were randomised, and excludes any subjects with major protocol deviations, which include those DMR cases that did not undergo the full DMR procedure. The full details on "major protocol deviations" that lead to patients being excluded from the PP population are discussed later in this SAP. This is a secondary analysis population for efficacy.

5.4 Safety Population

This analysis population includes all treated subjects, and these subjects are analysed by actual treatment received.

6 General Issues for Statistical Analysis

6.1 Analysis Software

Analysis data sets, statistical analyses and associated output will be generated using SPSS Software version 24 or later or Prism version 6.0 or later.

Variables are presented by treatment group using various descriptive statistics. Nominal and ordinal variables for each time period are presented using frequencies and percent of patients in each category. For variables collected at multiple follow-up time periods, tables which include appropriate descriptive statistics of change from baseline are presented by treatment group at each follow-up interval.

Distributions of each continuous variable are assessed prior to analysis and examined for normality. Statistical tests for the efficacy endpoints will be carried out using the 5.0% one-sided significance level, as well as using the 2.5% one-sided significance level unless otherwise specified.

6.2 Disposition of Subjects and Withdrawals

All subjects who provide written informed consent will be accounted for. The number and percentage of ITT and mITT subjects who discontinued the study prior to Week 24 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 and mITT 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

For the measures of insulin sensitivity there are only assessments at baseline and at Week 12, i.e., there are no post-baseline values that could be used for imputation. Therefore, in analysis of hepatic insulin sensitivity change will be missing if either the baseline value or the Week 12 value is missing. Likewise, change in peripheral insulin sensitivity will be missing if either the baseline or Week 12 value is missing. For the first primary endpoint of change from Baseline to Week 12 in Total Insulin Sensitivity, this will be missing if either of the two baseline, or either of the two Week 12 values are missing.

For number of menses during the 24 weeks after randomization, patients that do not have data recorded for any of Week 21-24 will have imputation carried out as follows: Let W represent the last

week for which menses (yes, no) was recorded, let Y represent W rounded up to a multiple of 4 (so that Y can take the values 4, 8, 12, 16, 20, or 24), then X the number of menses recorded in the first W weeks after randomization is multiplied by (24/Y) prior to analysis to allow for the missing data on the last (24-W) weeks since randomization, e.g., if a patient's last week at which menses (yes, no) are recorded is Week 15 (so that W=15, Y=16) and where two menses (so that X=2) have been recorded since randomization, then the value X=2 is replaced by the imputed value of 3.0 (=2x[24/16]).

Last observation carried forward (LOCF) imputation will be used for fasting and post-prandial glucose, insulin, and c-peptide excursions at Week 12, by carrying forward the corresponding Week 2 value if the Week 12 value is missing.

6.4 PROTOCOL VIOLATIONS

Protocol violations will be summarised in the 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
- Device or equipment not used per protocol
- Device not returned to sponsor
- DMR Procedure/laboratory assessment incomplete or not done
- DMR Procedure/laboratory assessment not done per 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 and un-blinding, all protocol violations will be reviewed in a blinded manner and patients who have had major violations will be noted and excluded from the per-protocol population.

6.5 MULTIPLE ENDPOINT ADJUSTMENT

The trial will be viewed as positive if statistical significance is obtained on either of the two primary endpoints. Multiple endpoint adjustment will be carried out using a weighted extension of the Hochberg[3] procedure so as to control the overall type 1 error at the required level. Within this procedure a weight w=0.8 will be used for the first primary endpoint (Prim1 = change from baseline to Week 12 in total insulin sensitivity), and a weight of 1-w =0.2 will be used for the second primary endpoint (Prim2 = number of menses in the first 24 weeks after randomization).

Let P1 denote the one-sided p-value for Prim1 and let P2 denote the one-sided p-value for Prim2, where in each case the one-sided alternative hypothesis represents the beneficial effect of DMR over Sham. The weighted Hochberg procedure will be first carried out using an overall α =0.050 one-sided. It would proceed as follows:

(i) if both one-sided p-values (P1 and P2) are at most 0.050 then statistical significance is demonstrated for both primary endpoints;

(ii) if (i) is not satisfied but P1 is at most 0.040 then statistical significance for that endpoint only is demonstrated;

(iii) if (i) is not satisfied but P2 is at most 0.010 then statistical significance for that endpoint only is demonstrated; or

(iv) if none of (i)-(iii) are satisfied then neither primary endpoint is statistically significant at this overall $\alpha = 0.050$ one-sided significance level.

A symbol "*" will be used to denote statistical significance for the primary endpoints using the overall $\alpha = 0.050$ one-sided significance level in the procedure described above.

The weighted Hochberg procedure would then be repeated but using an overall $\alpha = 0.025$ one-sided, i.e., comparing vs. 0.025 in (i), comparing vs. 0.020 in (ii) and comparing vs. 0.005 in (iii). A symbol "**" will be used to denote statistical significance for the primary endpoints using the overall $\alpha = 0.025$ one-sided significance level.

If a primary endpoint qualifies to be flagged as both "**" and "*" then it will be flagged using "**", i.e., will be flagged to denote the strongest significance level.

There will be no adjustment for the multiple secondary endpoints.

7 Demographics and Other Baseline Characteristics

7.1 Demographics

Demographics will be summarised by randomised treatment group for the mITT and the safety analysis populations. There will be no formal statistical comparisons between treatment groups on demographic variables. The continuous variables will be summarised by treatment group using sample size, mean, standard deviation, minimum and maximum. For the categorical variables of gender, race, and ethnicity the number and percentage of patients in each category will be presented for each randomised treatment group.

7.2 Baseline Medical History

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

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 mITT analysis set. If the baseline value is missing for a given variable and patient, the screening value will be used in its place prior to calculating the descriptive statistics.

8 Efficacy Analyses

8.1 Primary Efficacy Variables

The primary efficacy variables are:

1. The change from baseline at 12 weeks in total insulin sensitivity

2. The number of menses during 24 weeks.

The primary analysis for change from baseline to Week 12 in total insulin sensitivity endpoint will be performed in the mITT analysis population comparing treatment groups with an Analysis of Covariance (ANCOVA) model with terms for baseline total insulin sensitivity and treatment. A secondary analysis will be carried out on rank of change from baseline to Week 12 in total insulin sensitivity using ANCOVA with terms for rank of baseline total insulin sensitivity and treatment.

The primary analysis for the number of menses during the first 24 weeks after randomization will be performed in the mITT analysis population comparing treatment groups with an Analysis of Covariance (ANCOVA) model on rank of measured number of menses over 24 weeks with terms for rank of number of reported number of menses in the 12 months before randomisation and treatment.

For each of the analyses of the primary endpoints, least square means will be presented together with their SEs and two-sided confidence intervals.

For each primary endpoint corresponding analyses will also be conducted in the PP analysis population.

For each analysis, tables of descriptive statistics of the primary endpoint will include n, mean, standard deviation, least squares mean, standard error of least squares mean, median, quartiles, and minimum and maximum for each treatment group. Two-sided 90% and 95% confidence intervals of the difference between treatment least square means will be presented. Tables will be complimented by graphs as necessary.

8.2 Secondary Efficacy Variables

The following secondary endpoints will be compared between treatment groups on the mITT and PP analysis populations.

- Change in hepatic insulin sensitivity from baseline at 12 weeks
- Change in peripheral insulin sensitivity from baseline at 12 weeks
- Number of ovulatory cycles defined by an increase in serum progesterone and / or ultrasound evidence of ovulation followed by menstrual bleeding between weeks 12-24
- Change in fasting and post-prandial concentrations of glucose, insulin and c-peptide from baseline at 2 weeks
- Change in fasting and post-prandial concentrations of glucose, insulin and c-peptide from baseline at 12 weeks
- Change in free androgen index from baseline at 24 weeks
- % body weight loss from baseline at 24 weeks

Secondary endpoints will be compared between treatments at a one-sided 0.050 level of significance with the direction of the alternative hypothesis favoring DMR over control. There will be no adjustment for the multiple secondary endpoints. Missing data imputation for secondary endpoints is discussed in the Section 6.3 above. Analyses for the number of ovulatory cycles in Weeks 12-24 will be carried out in a similar manner to the analyses of number of menses as described in Section 8.1. Analyses for all other secondary endpoints will be carried out in a similar manner to the analyses of total insulin sensitivity as described in Section 8.1.

Tables will be complimented by graphs as necessary.

8.2.1 Exploratory Endpoints

The exploratory endpoints are as follows, and will be analysed for the mITT analysis population. Results will be presented using appropriate descriptive statistics (by randomised treatment group, where applicable, for endpoints measured during the double-blind phase). There will be no formal statistical comparisons between randomised treatment groups on exploratory endpoints, and there is no imputation of missing data.

- Change from baseline to week 24/ For each endpoint, temporal changes, mean levels and peak levels will be analysed:
  - Plasma lipid concentration
  - Plasma liver function tests
  - Arterial blood pressure
  - Serum LH
  - Serum FSH
  - Serum Oestradiol
  - Serum SHBG
  - Serum Testosterone
  - Serum free androgen index
  - Serum DHEAS
  - Serum Androstenedione
  - Energy expenditure
  - Apnoea hypopnoea index
  - Body composition

- Change from week 12 to week 24/ For each endpoint, temporal changes, mean levels and peak levels will be analysed:
  - Endometrial thickness
  - Ovarian volume
  - Follicle number
  - Diameter of largest follicle in each ovary

9 SAFETY ANALYSES

The analysis of safety data during the 24-week randomized double-blind period will be based on the Safety population, within which patients will be summarized by treatment administered.

9.1 PRIMARY SAFETY VARIABLE

The primary safety endpoint is the incidence rate of the device or procedure related Serious Adverse Events (SAEs) and Unanticipated Adverse Device Effects (UADEs) through the 24 weeks post treatment initiation. The safety endpoint summary will include the number and percentage of subjects in each of the categories. For the AEs occurring during the randomised phase, these numbers and percentages will be presented within each treatment group. A formal hypothesis test comparing treatments is not planned. Further details on adverse event analyses are provided below.

9.2 SECONDARY SAFETY VARIABLES

Physical Examination Vital Signs: Observed measurements and changes in physical exams and vital signs from baseline to post-baseline study time points will be descriptively summarised for each treatment group. For each vital sign, descriptive statistics of each vital sign will be presented at each visit for each treatment group; descriptive statistics of the change from baseline to each visit will also be presented. Listings of abnormal physical examination results will be presented; included in the listing will be subject id, body system where the abnormality occurred, study visit, and the physical

examination results for all visits (i.e., not just the visit where the abnormality occurred) for the given body system.

Clinical Laboratory Tests: Descriptive statistics of observed measurements in blood chemistry analysis and changes from baseline to each study time point in the double-blind phase will be presented for each treatment group. All laboratory values are compared to normal ranges; for each laboratory value, shift tables of normality status (low/normal/high, or normal/abnormal if the assignment of low and high does not apply) from baseline to each post-baseline visit will be presented for each treatment group. All data will also be presented in listings.

Adverse Events: AEs, SAEs and UADEs will be summarised through 24 weeks for each treatment by number and percentage of patients with at least one adverse event overall. Detailed listings of subjects that experience AEs and SAEs will be provided. The incidence of AEs will also tabulated (frequencies and percentages) by severity and relationship to procedure or device as outlined below. In tabulating the severity of AEs on a per subject basis, the greatest severity will be assigned to a subject should there be more than one occurrence of the same AE with different reported severities. Relationship will be categorised as unlikely, possibly and probably. The highest level of association is reported for subjects with different relationships for the same AE. Details of AE analyses are provided below.

9.3 Adverse Events

9.3.1 All Adverse Events

The number of treatment emergent adverse events (TEAEs) and the number and percent of subjects with at least one TEAE will be presented. 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. For TEAEs occurring in the randomised phase, results will be presented by treatment group.

The proportion of patients with at least one TEAE of special interest (TEAESI) with at least one procedure-related TEAE, and with at least one device related TEAE in the double-blind phase will be plotted by time point (peri-procedure, 0-1 week, 1-4 weeks, 4-8 weeks, …, 20-24 weeks) for each treatment group.

A listing of all adverse events will include the subject number, AE number, days since index procedure, the investigator description of the AE, the severity of AE, whether or not the AE is classified as serious (SAE), the relationship of the AE to the investigational device or 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 rates (number and percentage of subjects) of TEAEs leading to study withdrawal 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.

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 Device and Procedure Related Adverse Events

Summaries of device or procedure related TEAEs will be conducted in the same manner as for all TEAEs discussed above.

9.3.5 *Unanticipated Adverse Device Effects*

Summaries of treatment emergent unanticipated device TEAES will be conducted in the same manner as for all TEAEs discussed above.

9.3.6 *Adverse Events of Special Interest (AESIs)*

- Specific events that may be related to the mechanism of action of the DMR procedure (e.g., hypoglycemia)
- Potential adverse consequences of the procedure (e.g., gastrointestinal adverse events)
- Rare events that may or may not be related to the DMR procedure/device, but are of interest to the Sponsor (e.g. unexplained fever)

Events of Special Interest are:

- Hypoglycemia
- Diarrhoea
- Abdominal pain, nausea, vomiting
- Gastrointestinal bleeding
- Unexplained fever
- Stenosis (GI)

The number of events and incidence of AESIs will be presented.

9.3.7 *Deaths*

If a death occurs during the course of 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 following sections list the planned subgroup analyses. Additional subgroup analyses may be performed for exploratory purposes and will be identified as exploratory in the final report.

Treatment comparisons on the primary efficacy endpoints will be presented for the mITT population patients within each of the following subgroups:

A. Body Mass Index (BMI; (<Median, ≥Median)
B. Baseline total insulin sensitivity (<Median, ≥Median)

The purpose of the subgroup analysis is not to assess significance of the difference between treatments within subgroups, but to assess the consistency of treatment effect across subgroups.

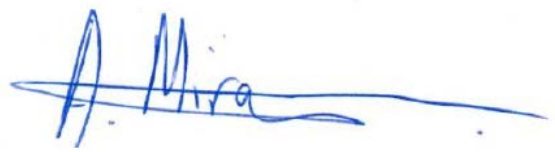
11 REFERENCES

1. Elkind-Hirsch, K., et al., *Comparison of Single and Combined Treatment with Exenatide and Metformin on Menstrual Cyclicity in Overweight Women with Polycystic Ovary Syndrome.* The Journal of Clinical Endocrinology & Metabolism, 2008. **93**(7): p. 2670-2678.

2. Rautio, K., et al., *Endocrine and metabolic effects of rosiglitazone in overweight women with PCOS: a randomized placebo-controlled study*. Hum Reprod., 2006. **21**(6): p. 1400-7. Epub 2006 Feb 24.

3. Hochberg, YA. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; 75:800 - 802.

Approved and signed on 1st June 2019



……..

Principal Investigator

Dr Alexander Dimitri Miras