Customer: Neurovalens Ltd
Information Type: Statistical Analysis Plan

Title:	Statistical Analysis Plan for 243973
Device:	VeSTAL Weight Loss Device
Effective Date:	15JUN2022

Description: Statistical Analysis Plan for Study 243973 to Demonstrate the Safety and Efficacy of VeSTAL Weight Loss Device in Human Subjects

Subject: Randomized Study of VeSTAL in Subjects who are Overweight

Author's Name, Title and Functional Area: Ruth Farquhar, Statistical Consultant, Exploristics
Meabh McCurdy, Statistician, Exploristics
Claire Watkins, Statistical Consultant, GEM Programming Solutions

Approved by:

L. McCrum
_____ 06 / 15 / 2022
Leigh-Ann McCrum Date
Clinical Research Manager
Neurovalens Ltd

_____ 06 / 15 / 2022
Claire Watkins Date
Statistical Consultant
GEM Programming Solutions Limited

Audit Trail For This Document			
Compiled by/ Affiliation	**Type of Change**	**Date**	**Version**
Meabh McCurdy	Initial Version Issued and Signed	27JUN2019	1.0
Linda Warnock	Updates following FDA advice	28OCT2019	1.1
GEM Programming Solutions statistics team	Changed ordering of secondary endpoints for label claims, alignment with protocol updates, alignment with FDA advice on VeSTAL DM study SAP, and clarifications prior to database lock	01JUN2022	1.2
Effective Date		15JUN2022	

Table of Contents

ABBREVIATIONS

ACTRI	Altman Clinical and Translational Research Institute
ADE	Adverse Device Event
AE	Adverse Event
BOCF	Baseline Observation Carried Forward
BMC	Bone Mineral Content
BMI	Body Mass Index
CRF	Case Report Form
CTM	Clinical Trial Mentors
DXA	Dual Energy X-Ray Absorptiometry
EPARC	Exercise and Physical Activity Resource Center
EU	European Union
FDA	Food and Drug Administration (United States)
FFQ	Food Frequency Questionnaire
G	Grams
HbA1c	Glycated Hemoglobin
HDL	High Density Lipoprotein
HEI	Healthy Eating Index
ITT	Intention To Treat
IWQOL	Impact of Weight on Quality Of Life
Kg	Kilograms
LOCF	Last Observation Carried Forward
LS	Least Squares
MMRM	Mixed Model with Repeated Measures
PP	Per Protocol
SAE	Serious Adverse Device Event
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Safety Analysis Set
SAS®	SAS® Statistical Analysis System
SD	Standard Deviation
SOP	Standard Operating Procedure
TDE	Texas Diabetes and Endocrinology
TFLs	Tables, Figures and Listings
UCSD	University of California San Diego
USA	United States of America
UU	University of Ulster
VAT	Visceral Adipose Tissue
VeNS	Vestibular Nerve Stimulation
WST	We Slim Together

1. INTRODUCTION

Neurovalens are conducting a randomized, double blind, sham controlled clinical trial to evaluate the efficacy and safety of non-invasive electrical vestibular nerve stimulation (VeNS), together with a lifestyle modification program, as a method of reducing excess body weight. The study will be randomized using a parallel assignment in a 1:1 active to control allocation, whereby the active group will use the vestibular nerve stimulation (VeNS) device, combined with a lifestyle modification program, and the control group will receive a sham device, combined with a lifestyle modification program. The study aims to recruit a total of 200 participants, across two or three sites, with a dropout allowance of 10% to generate a minimum of 90 active treatment and 90 control subjects. The study will be double blinded, meaning that both the subjects and outcome assessors will be blind to device assignment until the end of the visit at the 6 month timepoint.

For the purposes of the final analysis, the treatment assignment will be unblinded to all study personnel such that inferences can be made as to the efficacy and safety of the vestibular nerve stimulation (VeNS) device, in comparison to a sham control. The study will last approximately 6 months in total for eachparticipant. The final analysis will be conducted at the 6 month timepoint. All subjects will cease involvement in the trial after the 6 month blinded period.

This document outlines the process followed by the study team to conduct the final analysis, along with the planned efficacy and safety analyses designed to fulfill the study objectives. As specified in the study protocol, it is noted that the criteria to meet the primary objective, to establish the clinical performance of the VeSTAL device stimulation effect in subjects who are overweight, differ between the EU and USA regulatory clearance pathways. As a result, this document clearly defines the proposed analyses and success criteria for both the EU and USA regulatory clearance pathways in the relevant sections. The secondary efficacy and safety objectives are shared by both regulatory pathways and therefore will follow the same analysis plan once the primary objective has been evaluated.

Note that the version 1.1 of this SAP that was originally discussed and agreed with FDA was authored by Exploristics Ltd; GEM Programming Solutions Ltd, who will be conducting the analyses, made further updates and clarifications in version 1.2 of this SAP prior to database lock.

2. STUDY OBJECTIVE(S) AND ENDPOINT(S)

2.1 Study Objective(s)

The primary objective of the study is to establish the clinical performance of the VeSTAL device stimulation effect in subjects who are overweight. The clinical performance will be evaluated in terms of the efficacy of the VeSTAL device, in comparison to the sham control.

Secondary objectives of the study will include evaluations of the safety and tolerability of the VeSTAL device, relative to the sham control group, in terms of the occurrence of adverse events, change from baseline in vital signs, hearing tests and treatment exposure.

2.2 Study Endpoint(s)

2.2.1 Primary Efficacy

The primary efficacy endpoints, which will be analyzed at 6 months in order to quantify the efficacy of the VeSTAL device, differ between regulatory clearance pathways.

For the purpose of the ***USA regulatory clearance pathway***, the efficacy of the VeSTAL device will be quantified by two primary endpoints:

1. Mean percentage weight loss in comparison to the sham control
2. Proportion of participants who lose at least 5% body weight

For the purpose of the ***EU regulatory clearance pathway***, the efficacy of the VeSTAL device will be quantified based on a single primary endpoint:

1. Mean percentage weight loss in comparison to the sham control

2.2.2 Secondary Efficacy

The secondary efficacy endpoints will be consistent between regulatory clearance pathways and will include further analyses of percentage weight loss, fat loss, subject reported outcomes for quality of life, diet and activity, certain blood assays and changes in concomitant medications. These secondary endpoints will be analyzed at the 6 month timepoint.

2.2.2.1 Labelling Claims

Statistical comparisons between the VeSTAL and Sham control device treatment groups will be performed on percentage weight loss in the follow-up period, percentage fat loss, DXA measurements, dietary outcomes, quality of life measures, and blood parameters. These endpoints will be used to support the following seven labeling claims which will be tested sequentially, after the primary endpoints, in the following order:

At the 6 month timepoint

1. Percentage visceral adipose tissue (VAT) loss
2. Percentage change in low density lipoprotein (LDL)
3. Percentage fat loss
4. Percentage change in lean muscle mass
5. Change in Atherogenic index
6. Change in Systemic inflammation
7. Change in Total calorie intake
8. Change in Quality of Life total score

VAT loss, fat loss and lean muscle mass are measured by body DXA scanning. The atherogenic index will be determined by the ratio of total cholesterol to HDL. Systemic inflammation and quality of life

are measured by high-sensitivity C-reactive protein and the IWQoL questionnaire respectively. Total calorie intake will be measured using the two-day 24-hour dietary recall.

2.2.2.2 Other secondary endpoints

The efficacy of the VeSTAL device will be evaluated further by performing statistical comparisons between treatment groups on the following secondary efficacy endpoints:

- Percentage trunk fat loss
- Percentage bone mineral content (BMC) gain

The questionaire data, inclusive of the two-day 24-hour dietary recall and modifiable activity questionnaire, will be summarized per treatment group and time point in support of the efficacy findings. Additionally, changes in cardiovascular and initiation of diabetic medications will be summarized in terms of an increase, decrease or no change in concomitant medication. Medications will also be summarized by treatment according to potential impact on weight status. These will be classified as medications that could potentially cause either weight gain or weight loss as deemed by the Chief Investigator.

2.2.3 Safety

The safety of subjects will be assessed through summaries and descriptive comparisons between treatment groups on the following safety data:

- Laboratory Data
    - Fasting glucose
    - Glycated hemoglobin (HbA1c)
- Vital Signs
    - Blood pressure
    - Heart rate
    - Waist and hip circumferences
    - Body Mass Index (BMI)
- Adverse Event and Quality of Life Data
    - Frequency of Adverse Events (AE) and Adverse Device Events (ADE)
    - Adverse event monitoring questionnaire
    - Outcome of hearing loss assessments

2.2.4 Tolerability

Tolerability of treatment will be summarized and compared descriptively between treatment groups in terms of:

- Duration of Exposure
- Device usage data will be available from the devices permitting a treatment response analysis.
- Mentor support group usage (hours per week)

2.2.5 Blinding Assessment

In order to ensure the randomization of subjects to blinded treatment has been effective, a blinding assessment will be performed as part of the final analysis.

Subjects will be asked at baseline, 1 and 6 months to state which treatment arm they think they have been assigned to. Post unblinding of treatment at the final analysis, the true randomization list will be compared with the subject responses.

2.3 Statistical Hypothesis

2.3.1 Primary

As the primary efficacy endpoints differ between regulatory clearance pathways, the corresponding statistical hypotheses, tested at 6 months, are also specific to the regulatory pathway.

For both the ***USA and EU regulatory clearance pathways***, there will be one statistical hypothesis test to determine the efficacy of the VeSTAL device in comparison to the sham control:

- There is no significant difference between the mean percentage weight loss of participants using the VeSTAL device and the sham control, i.e. $\bar{x}_v - \bar{x}_s >= 0$, where $\bar{x}_v$ and $\bar{x}_s$ are the mean percentage weight loss of participants using the VeSTAL device and sham control respectively.

For the purpose of the ***USA regulatory clearance pathway***, the study will be deemed to have met the efficacy criteria if this hypothesis is rejected, and two numerical criteria are achieved:

- The difference between the mean percentage weight loss of participants using the VeSTAL device and the sham control is at least 2% as observed from the confidence intervals, (i.e. equivalent to assessing the hypothesis $\bar{x}_v - \bar{x}_s >= 2$).
- The lower bound of the 95% confidence interval of the response rate in the VeSTAL device group exceeds 50%, with response defined as a loss of at least 5% of body weight i.e. $p_v - ½ (C_u - C_L) \geq 50\%$, where p_v is the proportion of participants who lose at least 5% body weight using the VeSTAL device. C_L and C_U are the lower and upper bounds of the 95% confidence interval.

For the purpose of the ***EU regulatory clearance pathway***, the study will be deemed to have met the efficacy criteria if the hypothesis is rejected, and the following numerical criterion is achieved:

- The difference between the mean percentage weight loss of participants using the VeSTAL device and the sham control is at least 2% as observed from the mean estimates.

2.3.2 Support Labeling Claims

The secondary endpoints detailed in Section 2.2.2.1 will be used to support labeling claims and will be tested at the 6 month time point in a sequential order after the primary endpoints. These endpoints will assess the difference between treatment groups by considering the change from baseline to the 6 month time point using a linear model.

The first secondary endpoint that will be tested at 6 months is VAT. An increase in VAT loss, as measured by the whole body DXA scan, will be considered successful if significance is declared with the following statistical hypothesis:

- There is no significant difference between the mean percentage VAT loss of participants using the VeSTAL device and the sham control, i.e. $\bar{x}_v - \bar{x}_s = 0$, where $\bar{x}_v$ and $\bar{x}_s$ are the mean percentage VAT loss of participants using the VeSTAL device and sham control respectively.

The second secondary endpoint to be tested is LDL. A decrease in LDL will be considered successful if significance is declared with the following statistical hypothesis:

- There is no significant difference between the mean percentage change from baseline in LDL of participants using the VeSTAL device and the sham control, i.e. $\bar{x}_v - \bar{x}_s = 0$, where $\bar{x}_v$ and $\bar{x}_s$ are the mean percentage change from baseline in LDL of participants using the VeSTAL device and sham control respectively.

The third secondary endpoint to be tested is fat loss. An increase in fat loss, as determined by the whole body DXA scan, will be considered successful if significance is declared with the following statistical hypothesis:

- There is no significant difference between the mean percentage fat loss of participants using the VeSTAL device and the sham control, i.e. $\bar{x}_v - \bar{x}_s = 0$, where $\bar{x}_v$ and $\bar{x}_s$ are the mean percentage fat loss of participants using the VeSTAL device and sham control respectively.

The fourth secondary endpoint to be tested is lean muscle mass. An increase in lean muscle mass, as determined by the whole body DXA scan, will be considered successful if significance is declared with the following statistical hypothesis:

- There is no significant difference between the mean percentage change in lean muscle mass of participants using the VeSTAL device and the sham control, i.e. $\bar{x}_v - \bar{x}_s = 0$, where $\bar{x}_v$ and $\bar{x}_s$ are the mean percentage change in lean muscle mass of participants using the VeSTAL device and sham control respectively.

The fifth secondary endpoint to be tested is the atherogenic index. A decrease in the atherogenic index, as determined by the ratio of total cholesterol to HDL, will be considered successful if significance is declared with the following statistical hypothesis:

- There is no significant difference between the mean change from baseline for the ratio of total cholesterol to HDL of participants using the VeSTAL device and the sham control, i.e. $\bar{x}_v - \bar{x}_s = 0$, where $\bar{x}_v$ and $\bar{x}_s$ are the mean change from baseline in the ratio of total cholesterol to HDL of participants using the VeSTAL device and sham control respectively.

The sixth secondary endpoint to be tested is systemic inflammation. A decrease in systemic inflammation, as measured by high-sensitivity C-reactive protein, will be considered successful if significance is declared with the following statistical hypothesis:

- There is no significant difference between the mean change from baseline for the high-sensitivity C-reactive protein of participants using the VeSTAL device and the sham control, i.e. $\bar{x}_v - \bar{x}_s = 0$, where $\bar{x}_v$ and $\bar{x}_s$ are the mean change from baseline in the high-sensitivity C-reactive protein of participants using the VeSTAL device and sham control respectively.

The seventh secondary endpoint that will be tested at 6 months is total calorie intake. A decrease in total calorie intake, as measured by the two-day 24-hour dietary recall, will be considered successful if significance is declared with the following statistical hypothesis:

- There is no significant difference between the mean change from baseline for the total calorie intake of participants using the VeSTAL device and the sham control, i.e. $\bar{x}_v - \bar{x}_s = 0$, where $\bar{x}_v$ and $\bar{x}_s$ are the mean change from baseline of the total calorie intake of participants using the VeSTAL device and sham control respectively.

The eighth secondary endpoint that will be used to support labeling claims at 6 months is quality of life. An increase in quality of life, as determined using the IWQoL questionnaire, will be considered successful if significance is declared with the following statistical hypothesis:

- There is no significant difference between the mean change from baseline for the total QOL score of participants using the VeSTAL device and the sham control, i.e. $\bar{x}_v - \bar{x}_s = 0$, where $\bar{x}_v$ and $\bar{x}_s$ are the mean change from baseline in total QOL scores of participants using the VeSTAL device and sham control respectively.

3. STUDY DESIGN

The study is designed as a randomized double-blind sham controlled clinical trial to evaluate the efficacy and safety of vestibular nerve stimulation (VeNS), combined with a lifestyle modification program, compared to a sham control and lifestyle modification program, as a means of reducing excess weight.

The study will be randomized using a parallel assignment with a 1:1 allocation between the active VeSTAL device and the sham control. A total of 200 subjects are planned to be recruited across study sites, with a dropout allowance of 10% to generate a minimum of 90 active treatment and 90 control subjects. The anticipated recruitment across the sites are as follows:

- University of San Diego (UCSD) Altman Clinical and Translational Research Institute (ACTRI): a total of 106 subjects who pass through screening and are randomized.
- UCSD Exercise and Physical Activity Resource Center (EPARC; second trial site at UCSD campus): a total of 100 subjects who pass throughs screening and are randomized.
- University of Ulster (UU): a total of 94 subjects who pass through screening and are randomized.
- Texas Diabetes and Endocrinology (TDE): a total of 40 subjects who pass through screening and are randomized.

Data will be collated from all sites and analyzed as one dataset at the 6-month timepoint. The study is planned to last approximately 6 months for each subject, with the subjects given the headset device to use daily in the home environment.

4. PLANNED ANALYSES

The final analyses will be performed after the study database has been cleaned, locked and the true treatment allocations have been unblinded. Furthermore all randomized subjects will have received their device for a sufficient amount of time to allow 6 month data to be recorded and therefore contribute sufficient data for analysis.

The analyses will commence by summarizing the subject disposition, in particular, reporting the number of subjects in the analysis populations and any subjects that withdrew prior to study reporting or completion. The subject population will be described further by summarizing the demographic and baseline characteristics of subjects .

The efficacy analyses will proceed to investigate primarily any between group differences in the percentage weight loss from baseline adjusting for gender and baseline weight. Further analyses will investigate differences in the primary efficacy endpoints between groups, without adjusting for baseline covariates, repeated measures over time and a binary assessment of weight responders with fat loss >=5%. Multiple imputation will be the primary method for missing data, with LOCF and BOCF as sensitivity methods. The secondary endpoints VAT, LDL,fat loss, lean muscle mass, atherogenic index, sytemmic inflammation, total calorie intake and total QoL score will be analysed sequentially to form the basis of labeling claims.

Other remaining secondary endpoints will be summarised and analysed.

The safety of subjects will be summarized and descriptively compared between groups. The safety data will include a variety of endpoints including adverse events, hearing loss assessments, laboratory measurements, vital signs, and AE questionnaire responses.

The tolerability of device treatment will again be summarized and descriptively compared between groups for treatment exposure (or usage), adherence, and mentor support group usage.

Lastly the blinding of subjects will be assessed by comparing the subject provided responses of which treatment device they think they have been assigned to, compared to what treatment device they actually received.

5. SAMPLE SIZE CONSIDERATIONS

The primary outcomes are percent total body weight loss from the start of treatment (as a percentage of baseline total body weight), and the proportion of subjects who lose at least 5% of their baseline total body weight. These outcomes will be assessed at the 6-month mark.

Power calculations are based on simulation (for full details see supporting documents EXP18006 Simulation Plan v4.1 and EXP18006 Simulation Report v1.0). Simulation assumptions were informed by pilot data and data from the Modius device as follows:

- 90Kg median baseline weight in both groups (log-normal distribution with CV=24%),
- 6.6% mean percent weight loss in the active group at 6 months (SD 3.9%),
- 3.1% mean percent weight loss in the control group at 6 months (SD 3.9%),
- correlation between baseline weight and weight loss at 6 months of 0.4.

Under these assumptions, with a sample size of 180 participants randomized 1:1 active: placebo, there is greater than 80% power to demonstrate both the weight loss mean and categorical acceptance criteria at the 6 month time point.

Under an assumption of no difference between active and control but the same assumptions around SDs and correlations there is <2% probability overall of a type-I error with this sample-size.

Anticipating a drop-out rate of 10%, we will plan to recruit n = 200 total subjects overall in the study across all sites.

6. DATA MANAGEMENT

6.1 Data Collection and Preparation

The daily use of the device at home will be recorded automatically via the device application. Weight and body fat measurements will be recorded by study personnel at each site, who will enter data from source documents into the protocol-specific Case Report Form (CRF). All subject data will be transcribed from the subject level CRF and saved in excel format. For the purposes of the final analysis, the clinical study database will be cleaned, with particular attention given to the critical data and locked to prevent any further data entry or adjustment prior to the study reporting of 6 month data. Further details of the data management process are provided in the study Data Collection and Management SOP (Data Collection and Management for Vestibular Stimulation to Trigger Adipose Loss (VeSTAL) Clinical Trial (Study References 243973 and 181094), Ref NVDC020119, Revision 2).

6.2 Unblinding of Data

The clean, locked, blinded data will be transferred to the lead statistician in password protected excel files, one per site, via Exploristics's file sharing server. Likewise, the randomization list, in the form of a password protected excel file, will be requested from Neurovalens's randomization representative and shared with the lead statistician via Exploristics's file sharing server. Passwords for both documents will be shared separately via phone, to protect the security of the data.

The statistician, supported by a lead programmer, will import the data from Exploristics's secure server into SAS® and merge with the randomization list to unblind the data, such that it can be made analysis ready.

7. ANALYSIS POPULATIONS

The efficacy analyses will be performed on the Intention to Treat (ITT) population, which will include all randomized subjects. Subjects in the ITT Population will be classified according to the treatment as randomized.

Similarly, the safety and tolerability analyses will be performed on the Safety Analysis Set (SAS) population, which will include all randomized subjects who received either the VeSTAL or sham control devices (i.e., recorded at least one session of use defined as usage on at least one day).

However, subjects in the Safety Analysis Set will be classified according to the treatment received, which may on occasions differ from that randomized.

Should any major protocol deviations occur on study, the Per Protocol (PP) population will be defined which excludes any subjects who violate the protocol in terms of a major deviation. Of particular attention will be those subjects who are non-compliant with device usage. For example, to be included in the PP population, subjects must use the device for at least an average of three and a half hours a week for at least four of the first 6 months. If the PP population contains less than 95% of the ITT population then efficacy analyses will be repeated on the PP population.

8. TREATMENT COMPARISONS

All treatment comparisons will be performed between the VeSTAL and the sham control devices' treatment groups. Statistical comparisons will be performed under the hypothesis of no significant difference between groups, unless otherwise specified. The level of significance for the primary efficacy endpoints will not be adjusted for multiple testing as detailed in section 9.4. However, as the endpoints VAT loss, LDL, fat loss, lean muscle mass, atherogenic index, systemic inflammation, calorie intake and quality of life will be analyzed with the intention of making labeling claims, the level of significance for each endpoint will be adjusted for sequential testing (detailed in section 9.4). All remaining statistical comparisons, including the secondary efficacy endpoints will declare a significant difference between groups at the 5% significance level.

9. GENERAL CONSIDERATIONS FOR DATA ANALYSES

In general, continuous data will be summarized by the number of subjects in the analysis population (N), mean, standard deviation, minimum, median and maximum. The mean and median will be presented to one more decimal place than the original data, the minimum and maximum to the same

number of decimal places as the original data and the standard deviation presented to two more decimal places than the original data.

In summaries of categorical data, frequencies and percentages will be used. Subjects with missing values will be summarized in a 'Missing' category. Generally, percentage will be displayed to one decimal place and based on the number of subjects in the analysis population.

For statistical comparisons, p-values will be presented to 3 decimal places with p-values less than 0.001 will be presented as <0.001 in Tables, Figures and Listings (TFLs). Unless otherwise specified, confidence intervals will be two-sided with 95% confidence limits, and presented to the same number of decimal places as the point estimate.

If the assumptions for any planned linear model are not achieved, then a non-linear model such as one including quadratic terms will be investigated. If, after this, a satisfactory model still cannot be found then non-parametric methods will be applied.

All analyses will be performed using SAS® v 9.4 (or higher).

9.1 Dependent Variables (Endpoint Variables)

The efficacy endpoints detailed in sections 2.2.1 and 2.2.2 will serve as the dependent variables for the primary and secondary efficacy objectives of the final analysis respectively. Such endpoints will be utilized to meet the primary study objective, to establish the clinical performance of the VeSTAL device stimulation effect in subjects who are overweight.

9.2 Independent Variables

Independent variables will be comprised of demographic randomisation stratification factors and baseline values for the relevant endpoint and will be utilised as covariates when investigating differences between treatment groups for the dependent primary efficacy endpoints.

Such endpoints will include:

- Gender (stratification factor)
- Baseline value, e.g. weight (kg), body fat (kg)

9.3 Examination of Subgroups

The following subgroups will be examined:

- Region [poolability analysis] (US, non-US)
- Gender (male, female)
- Age (<55 years, ≥55 years)
- Race (white, non-white)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino/not reported)

Only subgroups with at least 10 patients per arm will be statistically evaluated. Some categories may be redefined or cut-offs changed as necessary to facilitate this.

Descriptive statistics (based on observed data) will be produced by subgroup for the primary efficacy (continuous weight and responders), key secondary efficacy for labelling claims, and primary safety (adverse event and adverse device effects) outcomes.

Interaction tests will be performed for the primary efficacy outcome where the subgroup main effect and a treatment*subgroup interaction term will be added to the primary model for percentage weight loss. The p-value for the interaction term will be presented.

Individual subgroup analyses will be performed for the primary efficacy outcome model and a table and forest plot will be presented showing the individual subgroup results (LS means by treatment, difference in LS means and 95% confidence intervals), to evaluate the magnitude and nature of any differential treatment effects between subgroups.

If any subgroup interaction term shows $p<0.1$ or there is a difference between subgroups that is judged to be of a potentially clinically relevant magnitude then a multivariate model will also be fitted including additional covariates that may explain possible outcome differences between subgroups. Those additional covariates will include the other subgroup main effects and their interactions with treatment, and the interaction of baseline weight with treatment. The model parameter estimates, 95% CIs and p-values will be presented including the interaction terms.

Similar subgroup analyses will only be conducted for key secondary efficacy outcomes (for labelling claims) if there is evidence of a differential effect for the primary outcome.

9.4 Multiple Comparisons and Multiplicity

The primary efficacy endpoint, percentage total body weight loss from baseline will be assessed at the 6-month timepoint and therefore will not require any adjustment for multiplicity or multiple comparisons. As a result, to control the type I error rate, the threshold of significance will be set at the 5% alpha level, such that statistical significance will require $p<0.05$ (two-sided).

As the secondary endpoints analyzed with the intention of making a labelling claim will be assessed sequentially, with a predefined ordering, there will be no adjustments made to the alpha level for multiple testing.

If a significant difference is achieved for the primary endpoint at 6 months, the study will continue to assess the secondary endpoints in the following order: VAT loss, LDL, fat loss, lean muscle mass, atherogenic index, systemic inflammation, calorie intake and quality of life. Each endpoint can only be assessed if the criterion for the previous endpoints have been met and all endpoints will be assessed using a nominal alpha level of significance set at 5%.

10. DATA HANDLING CONVENTIONS

10.1 Derived and Transformed Data

All subjects in the ITT population and Safety Analysis Set will have been randomized. Baseline will be considered as the latest value obtained prior to their treatment start (or randomisation, for any subjects randomised but not treated). Endpoint data will be reported using nominal times related to the treatment start date.

Percentage weight loss at any nominal time will be calculated as the percentage difference between the weight measurements recorded at baseline, minus the respective time point's weight measurement per subject. Response rate at any time will be calculated as the proportion of subjects who lose at least 5% body weight, with percentage weight loss defined as above.

Similarly, percentage change at any nominal time for other variables will be calculated as the percentage difference between the measurements recorded at baseline, minus the respective time point's measurement per subject.

Prior to performing statistical comparisons, the continuous efficacy endpoints will be evaluated for conformance to parametric assumptions, for example by assessing their distribution, and may be transformed e.g. log transformed if appropriate. If the parametric assumptions cannot be met through transformation, a non-parametric alternative may be used instead.

The laboratory and vital signs data will be recorded using the USA units at each site. As a result, it is not expected that any conversions will be required to standardise data into consistent units.

10.2 Premature Withdrawal and Missing Data

All subjects enrolled into the study who meet the requirements for the ITT or Safety Analysis Set population will be included in the final analyses. This will include those subjects who withdrew from the study prematurely or have missing data.

For the large part, subjects will contribute what data they have available without imputation. However, for the primary efficacy endpoint of total body weight loss, three approaches to missing or incomplete data will be taken, under the assumption that the data is missing at random:

1. Multiple imputation will be applied by means of a multiple imputation model which is composed of three steps:

    1.1. The missing data are imputed multiple times to generate multiple sets of data. The imputed values are predicted total body weight loss values obtained from a linear regression model with observed total body weight loss data as the dependent endpoint, and the same covariates of interest that are used to adjust the primary endpoint (treatment group, gender and baseline weight). By providing 75 imputations per missing value, the variability between imputed values represents the uncertainty over the missing value.

    1.2. The 75 complete datasets are analyzed using the planned statistical analysis as per section 12.1.1. The results from the 75 complete datasets are then combined to produce an estimate for the primary endpoint, which reflects the uncertainty due to the missing values.

2. Last Observation Carried Forward (LOCF) will be applied for all subjects without data, inclusive of those who have withdrawn from the trial and subjects who have not reached the time point.

3. Baseline Observation Carried Forward (BOCF) will be applied for all subjects without data, inclusive of those who have withdrawn from the trial and subjects who have not reached the time point.

The multiple imputation method will be considered primary, with the analyses using LOCF and BOCF imputation considered supportive. These methods will be applied to missing data at 6 months and also at 3 months for use in the repeated measures model. A separate multiple imputation model will be fitted for each time point.

For the binary weight loss endpoint, subjects who have missing data and withdrew due to adverse events or a change in medication will be considered non-responders in all analyses. Other subjects with missing data will use their imputed weight value to determine response status.

Additionally, a sensitivity analysis will be performed with the binary weight loss endpoint. This analysis will include a best-case scenario, assuming subjects with missing values (and that did not withdraw due to adverse events or change in medication) have all lost over 5% body weight if randomized to VeSTAL device and not lost over 5% body weight if randomized to sham control, and a worst-case scenario, assuming no such subjects with missing values have lost over 5% body weight if randomized to VeSTAL device and all lost over 5% body weight if randomized to sham control. If the best and worst case scenario imputations provide conflicting conclusions i.e. straddle p=0.05 then a tipping point analysis will be used to determine the proportion of values imputed with the best or worst case scenario at which the level of imputation changes the significance of the binary weight loss criterion (as described in Yan et al., 2009).

For adverse events which have missing or partial start dates, and it is unclear as to whether the AE (or ADE) is treatment emergent, the worse-case scenario will be assumed to categorise the AE (or ADE) as treatment emergent.

11. DESCRIPTIVE STATISTICS

11.1 Subject Disposition

A descriptive summary of subject disposition will be provided to summarize the number of subjects screened, randomized, treated and eligible for the ITT or Safety Analysis Set populations. Additionally, the number of subjects who have completed or discontinued treatment at the 6-month time point will be summarized, along with their reasons for discontinuation. In particular, those subjects who discontinue treatment or study follow up due to adverse events will be reported.

11.2 Demographic and Baseline Characteristics

The demographic and baseline characteristics listed in section 9.2 will be summarized for the ITT population according to the general considerations described in Section 9. The baseline characteristics of each treatment group will be observed and compared descriptively, with summary statistics provided per treatment group.

12. ANALYSIS CORRESPONDING TO THE STUDY OBJECTIVES

12.1 Efficacy

The clinical performance of the VeSTAL device, in comparison to the sham control device will be evaluated using efficacy analyses performed on the ITT analysis population and categorised as

primary, supportive and secondary efficacy. The primary efficacy analysis results, in combination with the supportive efficacy analysis results will be utilized to consider the success of the study.

12.1.1 Primary Efficacy Analyses

As the primary efficacy endpoints differ between regulatory clearance pathways, the primary efficacy analyses also differ accordingly.

For both the ***EU and USA regulatory clearance pathways***, the primary efficacy endpoint will be summarized using summary statistics for the percentage weight loss at the 6 month time point. The mean percentage weight loss will also be presented graphically, by treatment group, over time. Summaries and graphical outputs will be produced for observed data unless specified otherwise..

For the continuous efficacy endpoint, mean percentage weight loss at 6 months, the difference between treatment groups will be assessed using a linear model controlling for covariates of interest: gender and baseline weight. The least square (LS) mean difference between groups will be assessed for a statistically significantly difference from 0 i.e. p-value <0.05. For the USA regulatory clearance pathway, the lower limit of the confidence interval of the LS Mean difference between groups will be assessed against the 2% target for a successful study, whereas the EU regulatory clearance pathway will assess whether the LS mean difference between groups is significantly different from zero. Analysis using multiple imputation for missing data will be primary, with the LOCF and BOCF imputation methods used for supportive analyses.

12.1.2 Supportive Primary Efficacy Analyses

For the purposes of the ***USA regulatory clearance pathway***, the number and percentage of people achieving a loss of at least 5% of their body weight by the 6 month time point in the VeSTAL device treatment group will be presented with a two-sided 95% exact binomial confidence interval for the response rate. The criterion for success is that the lower bound of the 95% confidence interval of the response rate in the VeSTAL device group exceeds 50%. The analysis will use the multiple imputation for missing data as primary, with the LOCF and BOCF imputation methods as supportive.

A Fisher's exact test will also be used to compare the response rates in each treatment group and to detect whether the difference between groups is statistically significant. This will be repeated using multiple imputation, LOCF and BOCF methods for missing data. For the multiple imputation-based analysis, the median p-value approach will be used to generate the Fisher's exact test p-value (Eekhout 2017) and the exact confidence intervals described above will be produced using the imputation with a p-value equal to or nearest the median p-value.

In addition, a sensitivity analysis will be performed with the binary weight loss endpoint, to assess the impact of imputing the best and worst case scenarios for the binary response, when the total body weight loss at 6 months is missing. As per section 10.2, a tipping point analysis will also be utilised to demonstrate the level of imputation which changes the significance of the binary weight loss criterion.

No supportive primary efficacy analyses are required for the ***EU regulatory clearance pathway***.

12.1.3 Secondary Efficacy Analyses

For both the ***EU and USA regulatory clearance pathways***, secondary efficacy analyses will be performed to investigate the difference between treatment groups (using multiple imputation, LOCF and BOCF for missing data for weight outcomes, observed data for other outcomes) in:

- Mean percentage weight loss;
    - A t-test will be used to detect a statistically significant difference (p-value < 0.05) between the mean percentage weight loss per group (unadjusted analysis).
    - The mean percentage weight loss between groups, having accounted for covariates of interest and repeated measures over time. In order to do so, a mixed effect repeated measures (MMRM) model will be developed using percentage weight loss at 3 and 6 months) and controlling for the prespecified covariates of interest: gender and baseline weight. Subject will be included in the mixed effects model as a random effect with an unstructured covariance structure. If a convergent model cannot be achieved with an unstructured covariance, further covariance structures will be tested sequentially until a satisfactory model has been achieved.
    - Plot of the lsmeans of change from baseline at 3 and 6 months obtained from the MMRM analysis described above with MI for missing data only
- Adjusted response rate;
    - whereby response is defined as a subject achieving a percentage weight loss of at least 5% by 6 months. The adjusted response rate will be obtained by means of a logistic regression model with a binary response indicator as the dependent variable, treatment group as the explanatory variable and adjusting for covariates of interest: gender and baseline weight.
- Mean percentage VAT loss;
    - Summary statistics for percentage VAT loss at 3 and 6 months from baseline by treatment group
    - Plotted graphically by treatment group over time
    - A linear model will be used for the mean percentage VAT loss at 6 months having adjusted for covariates of interest: gender and baseline VAT. The least square (LS) mean difference between groups will be assessed for a statistically significantly difference from 0 i.e. p-value <0.05.
- Mean percentage fat loss;
    - Summary statistics for percentage fat loss at 3 and 6 months from baseline by treatment group
    - Plotted graphically by treatment group over time

- A linear model will be used for the mean percentage fat loss at 6 months having adjusted for covariates of interest: gender and baseline weight. The least square (LS) mean difference between groups will be assessed for a statistically significantly difference from 0 i.e. p-value <0.05.

- Mean percentage change in lean muscle mass;
  - Summary statistics for percentage change in lean muscle mass at 6 months from baseline by treatment group
  - Plotted graphically by treatment group over time
  - A linear model will be used for the mean percentage change in lean muscle mass at 6 months having adjusted for covariates of interest: gender and baseline lean muscle mass . The least square (LS) mean difference between groups will be assessed for a statistically significantly difference from 0 i.e. p-value <0.05.

- Additional fat and mass endpoints;
  - For each of the following endpoints, comparisons will be performed using a linear model, with the efficacy endpoint at 6 months as the dependent variable and treatment group, gender and baseline value as the explanatory variable. Data will also be summarised descriptively over time:
    - Mean percentage trunk fat loss
    - Mean percentage bone mineral content (BMC) gain

- Dietary endpoints;
  - Questionnaire responses, inclusive of those received via the food frequency questionnaire (FFQ) (baseline only) and the two-day 24-hour dietary recall will be summarized per treatment group at baseline and 6 months for the following endpoints:
    - Total Energy Intake (kCal)
    - Carbohydrates (Starchy carbohydrates, sugars and free sugars)
    - Fat and saturated fat
    - Protein
    - Fiber
    - Alcohol

- A statistical comparison will be performed between treatment groups for the change from baseline in the total calorie intake at 6 months, using a linear model, with the dietary endpoint as the dependent variable and treatment group, gender and baseline calorie intake as the explanatory variables. The two-day 24-hour dietary recall will be used to provide the total calorie intake data at the 6-month time point and baseline. Data will be summarised descriptively and plotted graphically over time.

- Blood endpoints;

  - For the following endpoints, a statistical comparison will be performed between treatment groups at 6 months, using a linear model, with the endpoint as the dependent variable and treatment group, gender and baseline value as the explanatory variables. Data will be summarised descriptively and plotted graphically over time:

    - Percentage change in LDL
    - Change in atherogenic index
    - Change in high-sensitivity C-reactive protein

- Quality of Life endpoint: A statistical comparison will be performed between treatment groups for the change from baseline in the total IWQOL score at 6 months, using a linear model, with the change from baseline in total IWQOL score as the dependent variable and treatment group, gender and baseline score as the explanatory variables. Data will be summarised descriptively over time:

- Modifiable activity questionnaire;

  - Questionnaire responses from the modifiable activity questionnaire will be summarized per treatment group at 0 and 6 months for the following endpoints:

    - Total Physical Activity averaged over the past year in hours per week
    - Total Physical Activity averaged over the past year in MET-hours per week

- Concomitant medication adjustments

  - The changes in cardiovascular and/or initiation of diabetic medications will be summarized in terms of an increase, decrease or no change in concomitant medication, by treatment group. This assessment will be made by suitably qualified members of the study team. (It is not anticipated that any subjects will start the study on anti-diabetic medications as having diabetes mellitus is an exclusion criterion to

enrolment, but it is possible that some may develop the condition during the course of the study).

- Concomitant medications will also be summarized by treatment group according to potential impact on weight status. These will be classified as medications that could potentially cause either weight gain or weight loss as deemed by the Chief Investigator.

In all secondary efficacy analyses which involve the use of linear, logistic or mixed effects repeated measures models, differences between groups will be evaluated using the LS mean difference between groups for continuous data, and an odds ratio between groups for categorical data. In both instances the 95% confidence intervals for the point estimates will be provided, along with the corresponding p-value for the hypotheses of a LS mean difference of 0 and an odds ratio of 1 respectively. Significance will detected using a 5% significance level, i.e. p-value <0.05.

For all efficacy outcomes, descriptive statistics will be produced based on the observed data for the absolute values at baseline and each post-baseline time point, as well as the change from baseline and (if relevant) percentage change from baseline to each time point.

12.2 Safety

The safety of the VeSTAL and Sham devices will be evaluated using descriptive summaries of safety data, with all analyses performed on the Safety Analysis Set.

A summary of treatment emergent Adverse Events (AE) and Adverse Device Events (ADE) will be provided for both the VeSTAL device and sham control treatment groups. Separate summaries will be provided, per treatment group, for any Serious Adverse Events (SAEs or SADEs) or causally related adverse events which occur on study.

Summaries of responses from the Adverse Event Monitoring Questionnaire at 1, 3, 6 months (where available), will be summarized per treatment group and time point. Additionally, the outcomes of hearing and ear canal assessments, performed at baseline, and 6 months will be summarized in order to detect any signs of deterioration which may occur whilst using the devices.

Laboratory data and vital signs as listed in section 2.2.3 will be summarized per treatment group at baseline and per time point on study. Additionally a summary of change from baseline for each of the laboratory tests will be presented over time.

Questionnaire responses for the quality of life – IWQOL-Lite questionnaire will be summarized per treatment group and time point.

12.3 Tolerability

Tolerability analyses will be performed on the Safety Analysis Set, which reflects the true treatment received by subjects. Of interest will be the device usage or exposure to treatment in terms of the time since randomization to study completion, withdrawal or the date of data cut off for subjects ongoing treatment at the time of the final analysis. The total number of device sessions performed over the course of the subject's study participation will also be summarized by treatment group.

It is likely that some subjects will not utilize their device for the recommended 7, or even mandated 5, hours per week, and therefore a treatment response analysis will be performed which will identify the number of subjects per treatment group who fail to meet these requirements at least once on study. The treatment response analysis will further summarize the frequency of times that the mandated hours are not adhered to by calculating the percentage non-adherence as n_t/n_T x 100 whereby n_t is the number of weeks a subject fails to use their device for the mandated 5 hours per week and n_T is the total number of weeks the subject has been on the study. The treatment response analysis will be used to identify groups of subjects which have high (>75%) , medium (50-75%) and low (<50%) treatment adherence (defined as 100% - percentage non-adherence) and their percentage weight loss will be summarized by treatment and adherence group accordingly, at the 6 month time point.

The total number of hours a subject avails of the mentor support group (WST or CTM) over the course of the subject's study participation will be described using summaries of the categorical time spent per week (0, 0-2, 2-4, 4-6, >6 hours), per treatment group.

12.4 Blinding Assessment

For the safety analysis set, a cross tabulation of the number of subjects who received the VeSTAL and sham control devices, against the number of subjects who think they are on 'Sham', 'Active' or who 'Do not know' will be reported for responses provided at baseline, 1 and 6, months. At 6 months this analysis is repeated for the coordinators response as to what device they think the subject is on: 'Sham', 'Active' or 'Do not know'. Both the subjects and coordinators provide reasons for their choice. This information is recorded in a listing.

The assessment will follow the approach outlined by James et al (1996) where an index of blindedness (BI) is created based on a modification of a Kappa statistic to incorporate the 'Do not know' category. The index ranges from 0 to 1 and gives an indication of how good the blind is, where 0 implies complete lack of blinding and 1 implies complete blinding. The weight assigned to the three response options will be 0 for a correct guess, 1 for a 'Do not know' response and 0.75 for an incorrect guess. A table of the actual treatment received against the perceived treatment received showing the numbers and proportions of subjects falling into each of the six categories will be produced alongside the index of blindedness where a result greater than 0.5 indicates a result better than that obtained by random choices. The 95% asymptotic confidence intervals will be included with the interpretation that a suitable level of blindedness has been achieved if the lower bound of the interval is above 0.5. The responses provided at 1 month will take precedence in deciding whether the blinding has been maintained, however should considerable differences exist between responses at each timepoint, further exploratory analysis may be performed, for example to infer whether subject response has an association with device performance or subject adherence.

13. CHANGES OF ANALYSIS FROM PROTOCOL

The analyses contained within this statistical analysis plan supersedes the analyses defined in the study protocol (May 26th 2019, Version 5). For clarity, a summary of deviations from the protocol are summarized, along with justification for the change.

1. All terminology reference to treatment dose has been removed as subjects will not receive a specific dose of treatment, but rather use the device headset as demonstrated. The dose response analysis described in the protocol will be reference as a treatment response analysis within this statistical analysis plan and related documents thereafter. Additionally, the treatment response analysis (section 12.3) will be performed on the safety analysis set to reflect the treatment that subjects actually received as opposed to the Intention to treat population referenced in the protocol which reflects randomized treatment.

2. The protocol references using usage time as a covariate in the linear modelling of percentage fat loss from baseline. As a post baseline endpoint, it may be difficult to distinguish the relationship between percentage fat loss and usage time, due to a cause and effect relationship e.g. the subject may discontinue treatment as they think the treatment is not effective, which in turn means that their percentage fat loss is reduced. As a result, only baseline covariates will be used to adjust the percentage fat loss from baseline, in the same way that percentage weight loss from baseline is treated. As an alternative means of investigating the relationship between usage time and percentage weight or fat loss, a descriptive summary will be provided for weight and fat loss by categorical groups of usage adherence.

3. The protocol states that the demographic and baseline characteristics of subjects will be compared statistically between treatment groups. Owing to the sample size, statistical comparisons between treatment groups may highlight potential differences between groups which are not considered clinically different. For example, a 1 year age difference between treatment groups may be statistically significant but not clinically relevant. As an alternative the demographic and baseline characteristics of subjects will be summarized per treatment group and a clinical review of the differences between groups performed. Additionally, the primary efficacy endpoints utilize a linear model with demographic and baseline covariates included. This is considered a more powerful test, having accounted for baseline attributes, and will be performed as the primary analysis, with the t-test performed as a secondary and supportive efficacy analysis.

4. The protocol states that Adjustment of medication, both reduction and new medications will be summarized and included as an efficacy endpoint. The monitoring of medications can be a complex process due to changes to medication dose, frequency, brand etc. As a result, this analysis will summarize any increase, decrease, or no change in medication, as specified by an investigator during the course of study follow up. Additionally, the medication review will focus on diabetic and cardiovascular medications only as they are considered most relevant to the co-morbidities associated with being overweight or obese. Medications will also be summarized by treatment according to potential impact on weight status. These will be classified as medications that could potentially cause either weight gain or weight loss as deemed by the Chief Investigator.

5. The dietary endpoint total calorie intake is intended to form the basis of potential future labeling claims. The total calorie intake endpoint will inform on the quantity of food eaten.

While broadly comparable (see: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5690458/) and producing the same outcome measures, the two-day 24-hour dietary recall and FFQ look at quite different timespans. So, while the two-day 24-hour dietary recall is self-explanatory the FFQ analyses diet over the previous year. The FFQ will only be administered at baseline. Thus, for the analyses that will be used to derive total calorie intake as the basis of potential labelling claims the two-day 24-hour dietary recall will be used.

6. The protocol states all analyses will be conducted using the latest version of R. However, the analyses will be conducted in SAS®.

7. The protocol states that a t-test will be performed for each of the secondary endpoints used to form the basis of labelling claims. However, these endpoints will instead be analysed utilising a linear model adjusted for subject's baseline covariate of the particular endpoint being analysed and gender (stratification factor), which is a more powerful test accounting for baseline differences. T-test will be provided as supportive analyses

8. HEI score had been included as a secondary endpoint to form the basis of labelling claims in the protocol but has since been removed as a study endpoint. This change will be reflected in any future iterations of the protocol.

9. After discussion between the sponsor and the responsible FDA panel chair, it was decided that the study now comes to an end when the 6-month primary endpoint is reached. This means that there will no longer be an unblinded period of 6 months usage at the UCSD sites following this. This was decided in order to allow the UCSD sites to prioritise study visits in the blinded study phase which were strained as a result of the COVID-19 pandemic. All analyses relating to later time points than 6-months have been removed. This change will be reflected in any future iterations of the protocol.

10. An additional mentoring support company called Clinical Trial Mentors (CTM) has been added to assist with the lifestyle modification support that We Slim Together (WST) currently provides if required. This was put in place as a fallback option to allow for the continuation of mentoring support should the COVID-19 pandemic cause WST to be unable to deliver their mentoring service.

11. A new research site called Texas Endocrinology and Diabetes (TDE) in Austin was added to the trial in order to increase the recruitment pool as the COVID-19 pandemic strain

14. REFERENCES

Yan, Lee & Li (2009). Missing Data Handling Methods in Medical Device Clinical Trials. Journal of Biopharmaceutical Statistics, 19: 6, 1085-1098.

Eekhout, van de Wiel & Heymans (2017). Methods for significance testing of categorical covariates in logistic regression models after multiple imputation: power and applicability analysis. BMC Medical Research Methodology, 17:129

James et al (1996). An Index for assessing Blindness in a Multi-Centre Clinical Trial. Statistics in Medicine, 15, 1421-1434.

HELLOSIGN

Audit trail

TITLE	Vestal SAP Signature
FILE NAME	SAPv1.2_final_updated_15Jun2022.docx
DOCUMENT ID	3c01a4358925722e116dd974995698bec50ac5f7
AUDIT TRAIL DATE FORMAT	MM / DD / YYYY
STATUS	Signed

Document history

SENT	**06 / 15 / 2022** 15:08:53 UTC	Sent for signature to Leigh-Ann McCrum (l.mccrum@neurovalens.com) and Claire Watkins (claire.watkins@clarostatconsulting.com) from l.mccrum@neurovalens.com IP: 194.46.129.59
VIEWED	**06 / 15 / 2022** 15:11:04 UTC	Viewed by Leigh-Ann McCrum (l.mccrum@neurovalens.com) IP: 194.46.129.59
SIGNED	**06 / 15 / 2022** 15:11:20 UTC	Signed by Leigh-Ann McCrum (l.mccrum@neurovalens.com) IP: 194.46.129.59
VIEWED	**06 / 15 / 2022** 15:12:01 UTC	Viewed by Claire Watkins (claire.watkins@clarostatconsulting.com) IP: 5.81.111.80

HELLOSIGN

Audit trail

TITLE	Vestal SAP Signature
FILE NAME	SAPv1.2_final_updated_15Jun2022.docx
DOCUMENT ID	3c01a4358925722e116dd974995698bec50ac5f7
AUDIT TRAIL DATE FORMAT	MM / DD / YYYY
STATUS	Signed

Document history

SIGNED	**06 / 15 / 2022** 15:12:24 UTC	Signed by Claire Watkins (claire.watkins@clarostatconsulting.com) IP: 5.81.111.80
COMPLETED	**06 / 15 / 2022** 15:12:24 UTC	The document has been completed.