STATISTICAL ANALYSIS PLAN
FOR PROTOCOL BTG-001933-01

Sponsor: Biocompatibles UK Ltd and Siemens Healthcare.

Protocol: BTG-001933-01

Title: Bariatric Embolization of ArTeries with imaging visibLe EmbolicS (BEATLES)

SAP Author: Richard E Thompson

Plan Version: SAP – Version 1.0

Plan Date: 23 October 2019

NCT04197336

Table of Contents

1.0 Introduction

This document provides details of the statistical considerations, analyses, and reports planned for the Bariatric Embolization of ArTeries with imaging visibLe EmbolicS (BEATLES) trial. BEATLES is an investigator-initiated, prospective, double-blinded, randomized, sham-controlled study that will assess the impact of bariatric embolization on the systemic levels of obesity-related hormones and, as a consequence, on weight loss. The goal of this study is to help treat obesity combining a lifestyle program and a minimally invasive, angiographic approach.

All data collection procedures and statistical analyses in this SAP report are considered as final, and were finalized and approved by the Clinical Investigators and Sponsor prior to data lock. In addition, the data collection procedures and statistical analyses detailed in this SAP report may modify and are given precedence over the analytical plans outlined in the clinical protocol; however, any major modifications or change in the primary endpoint and/or its analysis will also be reflected in a protocol amendment.

2.0 CONSORT Diagram

A CONSORT diagram will be created as the trial progresses, and presented at each Data Safety Monitoring Board (DSMB) meeting and at other meetings as requested by the DSMB. This diagram will present a breakdown of patient disposition, and information presented in the figure will include: 1) the number of patients screened; 2) enrolled and randomized numbers; 3) patients who failed screening (along with a breakdown of reasons for trial exclusion); 4) patients removed from the analyses who were enrolled and/or randomized, but found to be ineligible post-randomization; and 5) patients who dropout or were lost-to-follow-up, and reasons for dropping out or lost, if known.

3.0 Study Design and Objectives

3.1 Design Overview

The BEATLES study is a prospective, phase 2, single-site, double-blinded, randomized, sham-controlled trial. Patients who meet all of the inclusion / exclusion criteria, and who consent to participate in the trial, will be randomized at a 1:1 ratio to receive either bariatric embolization surgery under CT guided imaging, or receive a sham procedure, which mimics the actual embolization surgery. Those patients randomized to the control arm will be given subcutaneous lidocaine and receive a skin nick on the wrist or groin, whichever area is pre-determined in the pre-anesthesia care unit by the operating physician. The sham arm is necessary in this trial to provide a control against which weight loss due to embolization can be compared. All patients in both arms will receive weight management support.

3.2 Randomization

If all eligibility criteria for the study have been met by a given patient, and the patient consents to participate in the study, then this patient will be randomized into one of the two treatment groups via a computer-generated randomization scheme. Additionally, stratified randomization by diabetic status will be utilized to remove its potential confounding effect in comparing weight loss between treatment arms. Upon randomization/enrollment, each patient will be assigned a unique numeric patient identity code. Once randomized, the patient will be schedule for their assigned procedure, and will follow the steps of the procedure as outlined in the clinical protocol.

3.3 Blinding

Prior to the start of their assigned procedure, patients will be blind-folded and have their hearing damped so that they will be aware of their assigned procedure. It will not be possible to blind the surgical team to treatment given. However, the clinical team involved in patient evaluation post-surgery will have no knowledge of the procedure undertaken by any patients. In addition, the data analyses will be performed with only partially unmasked treatment assignments known to the data analysts (i.e. the analysts will have 'A' / 'B' treatment assignments without knowledge of which procedure corresponds to group 'A' and which to group 'B').

3.4 Study Objectives

The primary objective of this trial is to determine the efficacy of the bariatric arterial embolization (BAE) procedure using radiopague 100-200 µm microspheres (BTG-001933) for weight loss in obese patients. The secondary objective is to determine the safety of the BAE procedure and use of radiopague 100-200 µm microspheres in obese patients.

3.5 Outcome Measures

<u>Primary Efficacy Endpoint</u>

The primary outcome is the change in body weight at 12 months' post-procedure from baseline weight.

<u>Secondary Efficacy Endpoints</u>

Secondary efficacy endpoints include the following: 1) Percent change in body weight at 12 months' post-procedure from baseline; 2) The achievement of > 5% reduction in body weight at 12 months' post-procedure; 3) Percent excessive weight loss (EWL) at 12 months' post-procedure from baseline; and 4) Change in MRI biomarkers liver fat proton

density fat fraction (PDFF), visceral adipose tissue volume, subcutaneous adipose tissue volume, and muscle fat infiltration.

Primary Safety Outcome

The total adverse events (AEs) from procedure to 12 months' post-procedure will used to assess the safety of the BAE procedure and the use of radiopaque 100-200 μm microspheres in obese patients.

Secondary Safety Outcome

The total number of events until day 30 post-procedure will be considered the secondary safety outcome.

Exploratory Outcomes

The following outcomes will be considered as exploratory and potentially supportive of future research studies: 1) Change in blood pressure; 2) Change in lipid/lipoprotein profiles; 3) Change in obesity related hormones (ghrelin, LEAP2, leptin, glucagon-like peptide-1, peptide YY, etc.); 4) Changes in markers of obesity-related complications; 5) Food intake; 6) Hungry and satiety measures; 7) Quality of life as measured by the SF-36 and IWQOL; 8) Endoscopy results; 9) Ghrelin expression in pre- and post-operative fundal biopsy samples; 10) Changes in taste perception; 11) Image analyses of bead distribution and fundal coverage; 12) Embolization endpoints of occlusion and recanalization; 13) Change in muscle volume; 14) Change in muscle fat infiltration; 15) Change in visceral adipose tissue; 16) Change in CHD propensity; 17) Change in DT2 propensity; and 18) Differing treatment response in patients with adverse muscle composition.

3.6 Follow-up

Patient measures will be obtained post-procedure at weeks 1 and 2, and months 1, 3, 6, and 12.

3.7 Pre- and Intra-Procedure Characteristics

Tables of patient summary data will be created for important pre-operative (baseline) characteristics. The variables to be included in this table include: age; race; ethnicity; smoking status; child-bearing potential; comorbidities; and pre-operative weight (in kilograms), body mass index (BMI), percent body fat, excess body weight, diabetic status, blood pressure, laboratory and hematologic variables, and weight circumference.

Additional tables will be created that summarizes intra-operative variables, including procedure time, procedural complications, fluoroscopy time, radiation time, radiation dose, embolic

volume, peri-procedural medications, use of mechanical protection devices, and technical success/failure.

3.8 Power Calculation

Sample sizes per randomized treatment arm for the proposed study are based on providing sufficient power for testing the primary hypothesis that BAE leads to larger mean weight loss from baseline than a sham procedure.

Based on previous data, it is anticipated that the mean weight loss from baseline at 12-months post-randomization in the BAE arm will be at least 8 kg.[1,2,3] In contrast, we expect that participants in the sham arm will lose 3.3 kg at month 12 post-surgery (corresponding to 3% weight loss from a baseline weight of 110 kg). Also, a common standard deviation for the change in weight from baseline at 12 months of 7 kg is assumed based on the data from the prior BEAT obesity trial. Using a two-sample t-test and a two-sided alpha of 0.2, 21 participants in each arm will provide 80% power. The sample size is increased to 27 participants in each arm (i.e. a total of 54 randomized participants) to take account of an expected 20% drop-out rate by the time of the 12-month assessment. Since this is a phase 2 study, it is appropriate to use a two-sided alpha greater than the standard 0.05 value.

4.0 Analytical Populations

4.1 Full Analysis Set

The Full Analysis Set (FAS) to be used in the analyses of the efficacy outcomes will comprise all study patients who are randomized, undergo and complete the procedure to which they were assigned (embolization or sham surgery), and have at least one post-procedure outcome measure. The FAS will include any multiple imputed values for missing data from these patients.

4.2 Safety Analysis Set

Patients in the Safety Analysis Set (SAS) will include all individuals who are randomized and who underwent the procedure to which they were assigned. Patients in the SAS may include those individuals who do not complete any post-procedure, follow-up visits (i.e. have no post-procedure outcome measures). In addition, patients in the safety data set may include patients who underwent partial embolization or partial sham procedure due to the procedure being stopped short for any reason.

4.3 Completer Analysis Set

The Completer Analysis Set (CAS) will comprise all patients who were randomized, underwent and completed their assigned procedure, and who completed all post-procedure follow-up

visits (e.g. have at least the primary efficacy measures at post-procedure time points of weeks 1 and 2, and months 1, 3, 6, and 12). Analysis of the treatment effectiveness using CAS will be considered as sensitivity analysis to that performed using the FAS. However, if the efficacy outcome measures are missing completely at random (MCAR) or missing at random (MAR), then estimates of the treatment effect will be unbiased, and should produce results similar to those obtained using the FAS data with multiple imputation (MI) for the missing outcome measures.[4]

5.0 Statistical Analyses of Outcome Measures

5.1 General Exploratory Analyses

All statistical analyses will be performed using SAS® for Windows, version 9.4 or later. As a first step, descriptive statistics will be performed, and the normality of continuous variables evaluated using the Shapiro-Wilk test and through exploratory plots (e.g. histograms and Q-Q plots). Non-normal variables will be transformed to normality, using appropriate data transformations, or will be analyzed using non-parametric statistical methods. Frequency distributions, tabulations, and bar charts will be created to quantify categorical and binary variables. These exploratory analyses will help us assess data distributions, and also help identify any unusual, outlying or suspect observations.

Next, initial confirmatory analyses will be used to determine whether there are significant differences in baseline demographic and clinical variables between treatment arms, as well as between any comparable outside groups. These confirmatory analyses will include t-tests for continuous variables deemed to be normally distributed, or the non-parametric Wilcoxon Rank Sum test for non-normally distributed continuous variables. Binary and categorical variables will be analyzed using the Chi-Square goodness-of-fit test, or the Fisher's Exact Test, appropriate for categorical variables with small cell sizes. A lack of differences between treatment arms will be considered as evidence of adequate randomization; however, we expect that some baseline covariates will be imbalanced by treatment arm, given the small size of this study, due to chance.

5.2 Main Statistical Hypothesis

The null and alternative hypotheses for the primary efficacy outcome is given as:

Where Δwt is the change in weight (kg) at 12-months post-procedure minus baseline weight. In this trial, the null hypothesis will be rejected with type 1 alpha = 0.20.

5.3 Analysis of the Primary Efficacy Measure

The primary efficacy outcome for the proposed trial is total weight loss from baseline weight at 12-months post-randomization. Differences in the weight loss from baseline between the treatment arms will be compared using the two-sample t-test. In addition, we will also consider an ANCOVA model that looks at the group difference in mean weight loss from baseline, adjusting for baseline covariates that might be imbalanced by treatment arm. We will use the Wilcoxon Rank Sum test to compare the two treatment arms if the data are found to be non-normal as determined by exploratory plots.
[1]

5.4 Analysis of the Safety Measures

The primary safety endpoint will be the differences in adverse events at 12-months post procedure. The statistical analysis on the secondary outcome for safety will consider differences in the adverse events (AEs) at 30-days post-procedure.

5.5 Analysis of Secondary Efficacy Measures

Statistical assessments of the change in the continuous body-mass measures at 12-months from baseline (e.g. percent body-fat, excess weight loss) will follow the analyses described above for the primary efficacy measure. The binary outcome of > 5% weight loss will be statistically compared by treatment arm using the chi-square goodness-of-fit-test. Risk differences and relative risks, along with corresponding 95% confidence intervals, will be also calculated to quantify this group difference.

[How do we analyze the change in the multi-compartment composition?]

5.6 Analyses of Exploratory Measures

Several exploratory endpoints will be considered in the proposed trial as stated above, including blood pressure values, lipid profiles, serum obesity hormones, food intake, hunger assessments, QoL measures, and changes in taste perception. These endpoints will be quantified using a variety of assessment tools with both continuous and categorical scores. Statistical differences by treatment in these assessments will include the two-sample t-tests for continuous scores that are normally distributed, or the non-parametric Wilcoxon Rank Sum test for non-normally distributed continuous scores, while the binary and categorical scores will be analyzed using the Chi-Square goodness-of-fit test.

5.7 Sub-group Analyses

All analyses on the primary, secondary, and safety measures will be performed by gender, race, and ethnicity to determine if the treatment effect or safety differs by these important

demographic groups. Patient-specific clinical or "severity" factors/categories and treatment categories will also be explored. Planned explorations include diabetic status, age, compliance with weight management, baseline ghrelin levels, and post-procedure imaging characteristics.

5.8 Interim Analyses

An interim analysis of safety and technical utility of the embolic device (image visibility) and procedural assessments such as blinding, will be done after the 5 run-in patients complete the embolization procedure. Study reports on device/ procedure related complications are to be sent to the FDA following treatment of the five (5) run-in patients, and then after every 10 additional patients are randomized and treated (blinded).

No interim analysis for efficacy or futility are planned.

5.9 Multiplicity

The primary analysis of efficacy is tested at the two-sided alpha level of 0.20. In addition, secondary analyses will consider the alpha = 0.20 two-sided type I error. Secondary analyses will be examined in the order specified, and will not be corrected for multiplicity of testing. For safety outcomes, multiplicity will not be considered since this study is not powered on safety.

6.0 Missing Data

The clinical investigation team will make substantial efforts to ensure complete collection of data for all patients, and to accrue minimal loss to follow-up to optimize evaluation of the primary efficacy outcome. Our overall approach is to engage participants early on and keep them engaged through personal contact with the local center and access to information and resources that are relevant to them, their injury and their treatment. Furthermore, we do not expect the rates of missing data will be differential across treatment groups. As with most prospective studies, missing data will be unavoidable (even with excellent follow-up).

In order to impute the missing outcome data, analytic procedures will be utilized to 'fill-in' the missing body weights at 12 months' post-procedure for those patients who do not have data on this primary endpoint. Specifically, a prediction model using completers (within each treatment group) will be created. In this case, a completer is a patient with data on body weight at 12 months' post-procedure. The prediction model will include age, race, sex, and baseline body weight. We will then use multiple imputation to produce *m* random values for each missing 12 months' post-procedure endpoint based on the prediction model. In this imputation process, the value of *m* will be set to 20, creating 20 "complete" data sets. Results of the two-sample t-test and rank test (and corresponding 95% confidence intervals) will be calculated over all 20 imputed

data sets based on the combination rules as specified by Rubin (1976, 1987)[6,7]. Since only missing data at 12 months' post-procedure will be imputed, we assume that missingness is monotone.

Since the informative nature of missing data cannot be verified from the observed data, we will adopt a sensitivity analysis framework for reporting results. We will analyze data under a variety of modeling assumptions regarding how strongly the missingness mechanism is related to outcomes.[5]

7.0 Standard Calculations

7.1 Age

Patient age will be calculated in years by the following formula:

Age (years) = Integer of [(date of procedure – date of birth)]/365.25

If only the month and year of either the patient's birth or date of procedure is known, then the day will be given by the 15th of the month.

7.2 Change from Baseline

The change for any continuous outcome at any follow-up time (*t*) from baseline is defined as:

Change from baseline (designated as Δ_t) = Outcome measurement at time *t* – Outcome measure at baseline.

7.3 Percent Change from Baseline

The Percent change for any continuous outcome at any follow-up time point (*t*) from baseline is defined as:

Percent change from baseline = (Outcome measurement at time *t* – Outcome measure at baseline) / (Outcome measure at baseline)

8.0 References

[1] Weiss CR, Akinwande O, Paudel K, Cheskin LJ, Holly B, Hong K, Fischman AM, Patel RS, Shin EJ, Steele KE, Moran TH, Kaiser K, Park A, Shade DM, Kraitchman DL, Arepally A. Clinical Safety of

Bariatric Arterial Embolization: Preliminary Results of the BEAT Obesity Trial. Radiology. 2017;283(2):598-608. Epub 2017/02/15. doi: 10.1148/radiol.2016160914. PubMed PMID: 28195823; PMCID: PMC5410948.

[2] Kipshidze N, Archvadze A, Bertog S, Leon MB, Sievert H. Endovascular Bariatrics: First in Humans Study of Gastric Artery Embolization for Weight Loss. JACC Cardiovasc Interv. 2015;8(12):1641-4. Epub 2015/10/24. doi: 10.1016/j.jcin.2015.07.016. PubMed PMID: 26493259.

[3] Syed MI, Morar K, Shaikh A, Craig P, Khan O, Patel S, Khabiri H. Gastric Artery Embolization Trial for the Lessening of Appetite Nonsurgically (GET LEAN): Six-Month Preliminary Data. J Vasc Interv Radiol. 2016;27(10):1502-8. Epub 2016/08/29. doi: 10.1016/j.jvir.2016.07.010. PubMed PMID: 27567998.

[4] Mavuto, M, White, SA, Terlouw, DJ, Mwapasa, V, Kalilani-Phiri, L, Faragher, EB. Is Using Multiple Imputation Better Than Complete Case Analysis for Estimating a Prevalence (Risk) Difference in Randomized Controlled Trails When Binary Outcome Observations are Missing. Trials. 2016; 17: 341. PubMed PMID: 274500666 doi: 10.1186/s13063-016-1473-3.

[5] Rotnitzky, A, Scharfstein, DO, Su, TL, Robins, JM (2001). Methods for Conducting Sensitivity Analysis of Trials with Potentially Non-Ignorable Competing Causes of Censoring. Biometrics. 2016; 57: 103-112. PubMed PMID: 11252584.

[6] Rubin, DB. Multiple Imputation for Non-response in Surveys. New York: John Wiley & Sons, Inc. 1987.

[7] Rubin, DB. Inference and Missing Data. Biometrika. 1976; 63: 581–592.