STATISTICAL ANALYSIS PLAN

Evaluation of the effect of Duodenal Mucosal Resurfacing (DMR) using the Revita System in the Treatment of Type 2 diabetes (T2D)

C-30000

Sponsor:

Fractyl

May 3, 2019

Boston Biomedical Associates

100 Crowley Drive, Suite 216

Marlborough, MA 01752

TABLE OF CONTENTS

1 ABBREVIATIONS

Abbreviation	Definition
AE	Adverse Event
ALT	Alanine Aminotransferase
ANOVA	Analysis of Variance
AST	Aspartate Aminotransferase
AUC	Area Under the Curve
BBA	Boston Biomedical Associates
BMI	Body Mass Index
CP	Conditional Power
CRF	Case Report Forms
CSR	Clinical Study Report
DBP	Diastolic Blood Pressure
DMR	Duodenal Mucosal Resurfacing
FDA	United States Food and Drug Administration
FIB-4	Fibrosis-4 Index for Liver Fibrosis
FIM	First-in-Man
FPG	Fasting Plasma Glucose
FPI	Fasting Plasma Insulin
HbA1c	Glycated Hemoglobin
HDL	High Density Lipoprotein
IS	Insulin Sensitivity
ITT	Intent-To-Treat Population
LDL	Low Density Lipoprotein
LOCF	Last Observation Carried Forward
mITT	Modified Intent-To-Treat Population
MMTT	Mixed Meal Tolerance Test
MRFF	Magnetic Resonance Fat Fraction
PP	Per-Protocol Population
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SBP	Systolic Blood Pressure
SOC	System Organ Class
SSR	Sample Size Re-estimation
TEAE	Treatment Emergent Adverse Event
T2D	Type 2 Diabetes
UACR	Urine Albumin/Creatinine Ratio
UADE	Unanticipated Adverse Device Effect

2 SUMMARY

TITLE	Evaluation of the effect of Duodenal Mucosal Resurfacing (DMR) using the Revita System in the treatment of Type 2 diabetes (T2D)
PREFACE	This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for Fractyl Laboratories Protocol C-30000 (Evaluation of the effect of DMR using the Revita System in the treatment of Type 2 diabetes (T2D)). This study is being completed to assess the safety and efficacy of the Fractyl Revita System for the treatment of T2D. The following documents were reviewed in preparation of this SAP: • Clinical Research Protocol C-30000, issued 12OCT2017 • Case report forms (CRFs) issued YYXX2017 for Protocol C-30000
PURPOSE	The purpose of this SAP is to outline the planned analyses in support of the Clinical Study Report (CSR) for protocol C-30000. Exploratory analyses not necessarily identified in this SAP may be performed to support the clinical development program. Any post-hoc, or unplanned, analyses not identified in this SAP will be clearly identified in the respective CSR.
STUDY OBJECTIVES	Training Phase Objective: Provide training on the intervention procedure for the endoscopist as well as verify the safety profile before each site begins the randomized protocol. All subjects enrolled as training cases are treated with the Fractyl DMR procedure and followed per protocol for 48 Weeks. Phase 1 (0 - 24 Weeks, Double-Blind) Objective: To study the effect of DMR on glycemic and mechanistic endpoints 24 weeks post-procedure in subjects with T2D. Phase 2 (24 - 48 Weeks, Open-Label) Objective: To study the effect of DMR on glycemic endpoints for assessment of durability for patients who received DMR at Visit 3 (visit of initial randomized procedure).
ENDPOINTS AND ANALYSES	**Primary Efficacy: 1.** The change from baseline at 24 weeks in glycated hemoglobin (HbA1c) in DMR vs. Sham; **2.** The absolute change from baseline at 12 weeks in MR-PDFF in patients with baseline MR-PDFF >5%, DMR vs. Sham. **Primary Safety:** Incidence rates of device or procedure related Serious Adverse Events (SAEs), Unanticipated Adverse Device Effects (UADEs), and Adverse Events of Special Interest (AESIs) in DMR and Sham through 24 Weeks post treatment initiation. **Secondary Endpoints:** • HbA1c change from baseline through Week 24 (Visit 9) simultaneously across visits, DMR vs. Sham • The relative MR-PDFF change from baseline to Week 12 in patients with baseline MR-PDFF > 5%, DMR vs. Sham • Proportion of randomized-DMR-treated subjects with an HbA1c improvement from baseline at 24 weeks (Visit 9) that maintain an HbA1c improvement at 48 weeks

	• Proportion of randomized-DMR-treated subjects with an MR-PDFF > 5% at baseline and MR-PDFF improvement from baseline at 12 weeks (Visit 7) that maintain an MR-PDFF improvement at 48 weeks • Fasting Plasma Glucose (FPG) change from baseline at 24 weeks DMR vs. Sham • FPG change from baseline to Week 24 simultaneously across visits, DMR vs. Sham • Weight change from baseline at 24 weeks DMR vs. Sham • In randomized-DMR-treated subjects with an HbA1c improvement from baseline at 24 weeks, average HbA1c improvement from baseline at 48 weeks • In randomized-DMR-treated subjects with an MR-PDFF > 5% at baseline and MR-PDFF improvement from baseline at 12 weeks, average MR-PDFF improvement at 48 weeks • HOMA-IR change from baseline at 24 weeks DMR vs. Sham **Exploratory Endpoints:** • Change from baseline in the following at 12 weeks in DMR vs. Sham o MMTT change from baseline (glucose AUC through 2 hours) o Change in MMTT measure of Insulin secretion o Change in MMTT measure of Insulin resistance o MR-LIC Liver Iron Content • Change from baseline in the following by visit out to 24 weeks in DMR vs. Sham o FPI o Fasting C-peptide o Weight o ALT o AST o FIB-4 o Ferritin o UACR o eGFR o Triglycerides o HDL o TG/HDL • Change from baseline in the following at each of 4 weeks, 12 weeks, and 18 weeks, DMR vs. Sham: o HbA1c o FPG • Proportion of randomized-DMR-treated subjects with an abnormal MR-LIC at baseline and MR-LIC improvement from baseline at 12 weeks that maintain an MR-LIC improvement at 48 weeks.

	• In randomized-DMR-treated subjects with an abnormal MR-LIC at baseline and MR-LIC improvement from baseline at 12, average MR-LIC improvement at 48 weeks. • Training Cohort only: o Change in SBP by ABPM from baseline in training cohort, at 12 weeks o Change in DBP by ABPM from baseline in training cohort, at 12 weeks
STUDY DESIGN	Randomized double-blind (subject and endocrinologist) sham-controlled prospective multicenter clinical investigation of subjects with T2D sub-optimally controlled on 2 oral anti-diabetic medications. • Up to 15 Investigational Sites in European Union and global geographies • Maximum of 50 training cases (up to 5 in each site) and up to 120 randomized subjects (60 per each of DMR and Sham) • 1:1 randomized, double blind (subject and endocrinologist) trial comparing DMR treatment to sham procedure • 4 week oral anti-diabetic medication run-in to assess stability of blood glucose control in conjunction with medication compliance and nutritional counseling • Oral diabetic medications held constant from start of run in period through 24 Week endpoint with predefined rescue algorithm for hypo and hyper glycemia • Unblinding to occur at 24 Weeks and: o Sham treatment arm to cross over to receive DMR treatment at 24 Weeks with background medications held constant from 24 - 48 Weeks of follow up o DMR treatment arm to be managed according to current diabetes standard of care for 24 - 48 Weeks of follow up • Mechanism of action assessments, conducted in a subset of Study Sites, include: ambulatory blood pressure monitoring (ABPM) in training case only, Mixed Meal Tolerance Test (MMTT), Urine Micro Albumin, and Radiological Hepatic Status (MR-PDFF, MR-LIC) • DMR: Subject follow-up visits will occur at 7 and 14 Days (by phone) and 4, 12, 18, 24, 36 and 48 Weeks (in clinic), and 15, 21, 30 and 42 weeks (by phone) post procedure. Sham: Subject follow-up visits will occur at 7 and 14 Days (by phone) and 4, 12, 15(phone),18, and 24 weeks after the initial procedure. Sham patients who complete the crossover DMR procedure at Visit 3C(7-14 days after Week 24) follow up will occur at 7 and 14 Days (by phone) and 4, 12, 18, 24,36 and 48 Weeks (in clinic), and 15, 21, 30 and 42 weeks (by phone) post crossover DMR procedure.
INTERIM ANALYSES	There is no interim analysis for this study.

FINAL ANALYSES	The data through 24 weeks will be locked and all final planned analyses of primary and key secondary endpoints of the double-blind phase, identified in the protocol and in this SAP, will be performed after the last subject has completed the 24-week visit. A final double-blind report will be completed at this time. The data (up to 48 weeks) will be locked, and all final, planned, analyses of open-label 48-week endpoints identified in the protocol and in this SAP will be performed after the last DMR subject has completed the 48-week visit. A final report summarizing all results through 48 weeks will be completed at this time.

3 STUDY DESIGN

This is a randomized double-blind (subject and endocrinologist) sham-controlled prospective multicenter clinical investigation of subjects with T2D sub-optimally controlled on 2 oral anti-diabetic medications.

- Up to 15 Investigational Sites in European Union and global geographies
- Maximum of 50 training cases (up to 5 in each site) and up to 120 randomized subjects (60 subjects per each of DMR and Sham). The training phase aims to provide training on the intervention procedure for the endoscopist as well as verify the safety profile before each site begins the randomized protocol. All subjects enrolled as training cases are treated with the Fractyl DMR procedure and followed per protocol for 48 Weeks. Training phase patients are analyzed separately from patients entering the double-blind phase.
- 1:1 randomized, double blind (subject and endocrinologist) trial comparing DMR treatment to sham procedure
- 4 week oral antidiabetic medication run-in to assess stability of blood glucose control in conjunction with medication compliance and nutritional counseling
- Oral antidiabetic medications held constant from start of run in period through 24 Week endpoint with predefined rescue algorithm for hypo and hyper glycemia management
- Unblinding to occur at 24 Weeks and:

    o Sham treatment arm to cross over to receive DMR treatment at 24 Weeks with background medications held constant from 24 - 48 Weeks of follow up

    o DMR treatment arm to be managed according to current diabetes standard of care for 24 - 48 Weeks of follow up

- Mechanism of action assessments, conducted in a subset of Study Sites, include: ambulatory blood pressure monitoring (ABPM) in training cases only, Mixed Meal Tolerance Test (MMTT), Urine Micro Albumin, and Radiological Hepatic Status (MR-PDFF, MR-LIC).

- DMR: Subject follow-up visits will occur at 7 and 14 Days (by phone) and 4, 12, 18, 24, 36 and 48 Weeks (in clinic), and 15, 21, 30 and 42 weeks (by phone) post procedure. Sham: Subject follow-up visits will occur at 7 and 14 Days (by phone) and 4, 12, 15(phone),18, and 24 weeks

after the initial procedure. Sham patients who complete the crossover DMR procedure at Visit 3C(7-14 days after Week 24) follow up will occur at 7 and 14 Days (by phone) and 4, 12, 18, 24,36 and 48 Weeks (in clinic), and 15, 21, 30 and 42 weeks (by phone) post crossover DMR procedure.

4 Sequence of Planned Analyses

4.1 Interim Analysis for Sample Size Recalculation

There will be no interim analysis for this study.

4.2 Final Analyses and Reporting

4.2.1 Final Double-Blind Analysis

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

4.2.2 Final Open Label Analysis

The data (up to 48 weeks) will be locked, and all final, planned, analyses of open-label 48-week endpoints identified in the protocol and in this SAP will be performed after the last subject randomized to DMR has completed the 48- week visit. A final report summarizing all results through 48 weeks will be completed at this time.

4.2.3 Final Report

As discussed directly above, two final reports will be completed: the first after all subjects complete the double-blind portion of the study, and the second after all DMR-randomized subjects complete their final 48-week visit. In each report, any post-hoc, exploratory analyses completed to support planned study analyses, which were not identified in this SAP, will be documented and reported as necessary. Any results from these unplanned analyses will also be clearly identified as post-hoc analyses.

5 STUDY OBJECTIVES AND ENDPOINTS

5.1 STUDY OBJECTIVE

To demonstrate the efficacy and safety of the Fractyl DMR Procedure using the Revita System compared to a sham procedure for the treatment of uncontrolled type 2 diabetes.

Training Phase Objective: The training phase aims to provide training on the intervention procedure for the endoscopist as well as verify the safety profile before each site begins the randomized protocol. All subjects enrolled as training cases are treated with the Fractyl DMR procedure and followed per protocol for 48 Weeks. Patients treated in the training phase (training cases) are not part of the double-blind phase and open-label phase analyses described below.

Phase 1 (0 - 24 Week Double-Blind Phase) Objective: To study the effect of DMR on glycemic and mechanistic endpoints 24 weeks post-procedure in subjects with T2D.

Phase 2 (24 - 48 Weeks Open-Label Phase) Objective: To study the effect of DMR on glycemic endpoints for assessment of durability for patients who received DMR at Visit 3 (randomized intervention visit). Patients randomized to Sham at the start of Phase I will be allowed to crossover to DMR but will not be included in the analysis of durability of DMR; only patients randomized to DMR at the start of Phase I will be included in the assessment of durability of DMR.

5.2 STUDY ENDPOINTS

5.2.1 PRIMARY EFFICACY ENDPOINT

Primary Efficacy: 1. The change from baseline at 24 weeks in glycated hemoglobin (HbA1c) in DMR vs. Sham; **2.** The absolute change from baseline at 12 weeks in MR-PDFF in patients with baseline MR-PDFF >5%, DMR vs. Sham.

5.2.2 PRIMARY SAFETY ENDPOINT

Incidence rates of device or procedure related Serious Adverse Events (SAEs), Unanticipated Adverse Device Effects (UADEs), and Adverse Events of Special Interest (AESIs) in DMR and Sham through 24 Weeks post treatment initiation.

5.2.3 SECONDARY EFFICACY ENDPOINTS

- HbA1c change from baseline through Week 24 (Visit 9) simultaneously across visits, DMR vs. Sham
- The relative MR-PDFF change from baseline to Week 12 in patients with baseline MR-PDFF > 5%, DMR vs. Sham
- Proportion of randomized-DMR-treated subjects with an HbA1c improvement from baseline at 24 weeks (Visit 9) that maintain an HbA1c improvement at 48 weeks
- Proportion of randomized-DMR-treated subjects with an MR-PDFF > 5% at baseline and MR-PDFF improvement from baseline at 12 weeks (Visit 7) that maintain an MR-PDFF improvement at 48 weeks
- Fasting Plasma Glucose (FPG) change from baseline at 24 weeks DMR vs. Sham
- FPG change from baseline to Week 24 simultaneously across visits, DMR vs. Sham

- Weight change from baseline at 24 weeks DMR vs. Sham
- In randomized-DMR-treated subjects with an HbA1c improvement from baseline at 24 weeks, average HbA1c improvement from baseline at 48 weeks
- In randomized-DMR-treated subjects with an MR-PDFF > 5% at baseline and MR-PDFF improvement from baseline at 12 weeks, average MR-PDFF improvement at 48 weeks
- HOMA-IR change from baseline at 24 weeks DMR vs. Sham

5.2.4 EXPLORATORY ENDPOINTS

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

- Change from baseline in the following at 12 weeks in DMR vs. Sham
  - MMTT change from baseline (glucose AUC through 2 hours)
  - Change in MMTT measure of Insulin secretion
  - Change in MMTT measure of Insulin resistance
  - MR-LIC Liver Iron Content

- Change from baseline in the following by visit out to 24 weeks in DMR vs. Sham
  - FPI
  - Fasting C-peptide
  - Weight
  - ALT
  - AST
  - FIB-4
  - Ferritin
  - UACR
  - eGFR
  - Triglycerides
  - HDL
  - TG/HDL

- Change from baseline in the following at each of 4 weeks, 12 weeks, and 18 weeks, DMR vs. Sham:
  - HbA1c
  - FPG

- Proportion of randomized-DMR-treated subjects with an abnormal MR-LIC at baseline and MR-LIC improvement from baseline at 12 weeks that maintain an MR-LIC improvement at 48 weeks.

- In randomized-DMR-treated subjects with an abnormal MR-LIC at baseline and MR-LIC improvement from baseline at 12, average MR-LIC improvement at 48 weeks.

- Training Cohort only:
  - Change in SBP by ABPM from baseline in training cohort, at 12 weeks
  - Change in DBP by ABPM from baseline in training cohort, at 12 weeks

6 SAMPLE SIZE

6.1 TRAINING PHASE

For each study site, between 3 and 5 training cases are performed, with up to 15 sites participating. The number of training cases per site is based on training requirements and not statistical considerations. A maximum of 50 cases will be conducted as part of the Training Phase.

6.2 RANDOMIZATION PHASE

The primary endpoints are 1) the change from baseline at 24 weeks in HbA1c and 2) The absolute change from baseline at 12 weeks in MR-PDFF in patients with baseline MR-PDFF > 5% , DMR vs Sham.

Assumptions of effect size for the primary efficacy endpoint of change from baseline at 24 weeks in HbA1c in the treatment arm are derived from the Revita 1 study (Fractyl Protocol C-20000). The treated subjects saw a mean change difference in HbA1c of -1.0 at 24 weeks with a standard deviation of 1.0. Based on previous experience, the estimated mean (± sd) sham effect at 24 weeks is -0.3 ± 1.0.

Assumptions are similarly obtained for the primary efficacy endpoint of absolute change from baseline at 12 weeks for MR-PDFF in patients with baseline MR-PDFF > 5%, with an assumed difference between treatment and sham MR-PDFF means of 4.0 and a standard deviation assumption of 6.5 per treatment group.

Under (a) the assumption of a difference in mean change in HbA1c between treatment and control of 0.7 at 24 weeks with equal variance in both groups (standard deviation of 1.0) of the C-30000 study; (b) the assumption of a difference in mean change in MR-PDFF between treatment and control of 4.0 at 12 weeks and a standard deviation of 6.5 per treatment group; (c) approximately 3% of randomized subjects will not be evaluable for HbA1c and approximately 70% of randomized subjects will have baseline MR-PDFF >5% and be evaluable for 12-week MR-PDFF; and (d) the correlation between the two primary endpoints is 0 or very small, then 90 randomized subjects (45 per group) provides at least 90% power that the benefit of treatment over sham will be found for at least one primary endpoint using the Hochberg procedure controlling the experimentwise significance level at a one-sided 0.05 value.

A maximum of 120 patients may be randomized to account for potential patients lost to follow up prior to the primary endpoint assessment.

Table 2: Planned Sample Size

Study Cohort	Fractyl DMR	Sham Arm	Total
Training	up to 50	N/A	**Up to 50**
Miminum Randomized	45	45	**90**
Maximum Randomized	60	60	**120**
Combined Training and Randomized (minimum - maximum)	**95 – 110**	**45 – 60**	**140 - 170**

7 ANALYSIS POPULATIONS

7.1 TRAINING SUBJECTS

The training population for this study includes all subjects enrolled as training cases at the site. These cases are described separately from randomized subjects in the final study reportS and are presented primarily using descriptive statistics. These subjects will not be included in the randomized primary efficacy analysis.

7.2 INTENT TO TREAT POPULATION (ITT)

The intent-to-treat (ITT) population for this study includes all randomized subjects and excludes the training subjects.

7.3 MODIFIED INTENT TO TREAT POPULATION (MITT)

The mITT population includes all randomized 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 analyzed according to their randomized group assignment. The mITT population is the primary analysis population for both the primary and secondary efficacy endpoints. The mITT population specifically excludes the training subjects.

7.4 PER-PROTOCOL POPULATION (PP)

The Per-Protocol (PP) analysis population includes the subset of mITT subjects who received the treatment to which they were randomized, and excludes any subjects with major protocol deviations, which include those Fractyl 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. The PP population specifically excludes the training subjects. This is a secondary analysis population for efficacy.

7.5 SAFETY POPULATION

This analysis population includes all subjects in whom the treatment (DMR or sham) was initiated, and these subjects are analyzed by actual treatment received. This analysis population is the primary analysis population used for all safety endpoint analyses.

7.6 CROSSOVER POPULATION

Control subjects who elect to crossover to DMR treatment at 24 weeks will be analyzed separately beginning from the point at which active treatment is initiated. Safety endpoints will be summarized for this cohort separately.

8 General Issues for Statistical Analysis

8.1 Analysis Software

Analysis data sets, statistical analyses and associated output generated by Boston Biomedical Associates, LLC (BBA) will be generated using SAS® Software version 9.4 or later or R version 3.3.2 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. Interval and ratio variables for each time period are presented using means and standard deviations and/or medians, quartiles and ranges, as appropriate. 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. Data with interval or ratio scales to be analyzed that are not normal are transformed in an appropriate manner (e.g., log transformation), or are analyzed using nonparametric statistics. Statistical tests are performed using two-sided significance levels of 5% unless otherwise specified.

8.2 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 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. The number and percentage of subjects who prematurely discontinued prior to Week 48 will be presented in a similar manner, by original randomized treatment group; percentages will be based on the number of subjects entering the open-label phase.

8.3 Methods for Missing Data

All efforts will be made to prevent the occurrence of missing data. The sites will be instructed to identify and obtain contact information for the primary care giver for the subject. Complete site training and regular monitoring will also help to minimize missing data. Nevertheless, it is anticipated that withdrawals will occur and hence there will be missing data on primary and secondary efficacy endpoints. Further, some patients may take rescue medication prior to the end of the study; given that rescue medication may cause misestimation of the true effect of randomized treatment, for the primary analysis on the primary endpoints, all endpoint data measured after the use of rescue medication will be set to missing and treated in the analysis as missing data. For the purpose of this analysis, rescue medications will be defined as increase to oral anti-diabetic medications.

The number and percentage of patients with missing data for each of the primary endpoints and the number and percent of patients taking rescue medication will be presented by treatment group and study visit. The primary analysis of the primary endpoints will use multiple imputation using linear regression to impute missing data (where data collected post-rescue medication use are first set to missing prior to carrying out the imputation) as further described below. Sensitivity analyses will then be conducted where missing data for the primary endpoints (where data collected post-rescue medication use are set to missing) are not imputed (available case analyses). Then as further sensitivity analyses, all imputation methods (multiple imputation, available case) will be repeated where data collected post-rescue medication are NOT set to missing prior to carrying out the analyses.

Multiple imputation (primary approach to analysis of the primary endpoints): The following describes the multiple imputation approach for the first primary efficacy endpoint of the change from baseline at 24 weeks in glycated hemoglobin (HbA1c). The imputation approach will be similarly carried out for the second primary endpoint of the absolute change from baseline at 12 weeks in MR-PDFF in patients with baseline MR-PDFF >5%.

The primary efficacy analysis will be carried out on a dataset where missing Week 24 HbA1c is first multiply imputed for randomized patients with missing HbA1c at week 24. Multiple imputation of HbA1c at week 24 will be carried out as follows:

A. Assuming the missing HbA1c data pattern will not be exactly monotone across visits, then non-monotone missing HbA1c data will be imputed 50 times to make 50 copies of a dataset with a monotone missing HbA1c data pattern; the Monte Carlo Markov Chain approach will be used for this imputation.
B. For each of the 50 datasets imputed as discussed above to have the monotone missing-data pattern, missing week 24 HbA1c for randomized patients will be imputed once using the monotone linear regression approach: the imputation model will include treatment group, study center (centers with fewer than 5 subjects will be pooled with another study center by geographic region to ensure at least 5 subjects per study center), baseline HbA1c, baseline fasting plasma glucose, baseline fasting plasma insulin, change from screening to baseline HbA1c, change from screening to baseline fasting plasma glucose, change from screening to baseline fasting plasma inulin, duration (years) of Type 2 diabetes; also included in the imputation model will be age, sex, BMI, and fasting plasma insulin, fasting plasma glucose and HbA1c measured at scheduled visits prior to week 24.

For each of the 50 imputed datasets, the estimate of the difference in change in HbA1c from baseline to 24 weeks between the treatment group and sham will be estimated, as will the variance of these treatment differences, using analysis of covariance (ANCOVA) adjusting for baseline HbA1c and study center (further details of the ANCOVA are provided in the efficacy analysis section below). These estimates will then be combined across the 50 datasets using SAS PROC MIANALYZE and the appropriate theory for combining imputed datasets to obtain an estimate of the difference in mean change in HbA1c between the treatment and sham, and the standard error of the difference, from which an overall one-sided p-value assessing the significance of the difference between treatments with respect to HbA1c will be calculated.

Available Case (sensitivity analysis):There will be no imputation of missing data; only available (non-missing) data will be used; available case analysis will be conducted using mixed models repeated measures (MMRM) that is detailed later in this SAP, with a focus on estimating the treatment difference at Week 24 from the model.

For secondary endpoints, the only analysis to be carried out will be last-observation-carried forward (LOCF) for missing data, where post-rescue values are first set to missing. There will be no imputation of missing data for exploratory endpoints (post-rescue medication will be set to missing).

8.4 Protocol Violations

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

- Informed consent deviation
- Inclusion/Exclusion
- 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
- Significant changes in anti-diabetic con meds, not per protocol.
- Visit/assessment performed significantly out of window

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 unblinding, 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.

8.5 Multiple Comparisons and Multiplicity

The secondary efficacy endpoints will be compared between treatment groups in the double-blind phase at a one-sided 0.05 level of significance, where the one-sided alternative hypothesis represents a beneficial effect of DMR over sham. There will be no alpha adjustment for multiple comparisons across the secondary endpoints.

There will be no formal treatment comparisons for exploratory endpoints.

8.6 Assessment of Homogeneity Across Regions

To evaluate consistency in treatment effect across geographic regions (Belgium, Brazil, Italy, Netherlands, United Kingdom) for each of the two primary endpoints, summary tables of treatment differences by region will be presented for each primary endpoint using the multiply imputed data where post-rescue medication is set to missing prior to carrying out the imputation. This will be carried out on the mITT analysis population. The significance of treatment-by-region interaction on each of the two primary endpoints will be assessed using analysis of covariance with effects for treatment, region,

baseline value of the outcome, the difference between the screening and baseline value of the outcome (for HbA1c outcome only), and treatment-by-region interaction.

For each primary endpoint, a treatment-by-region interaction p-value <0.10 will require further investigation of treatment effect by region to assess if regions are poolable. If the p-value <0.10 but treatment effect across regions differs only in magnitude and not direction of effect, the regions will still be pooled for the main analyses of the given endpoint. Otherwise, demographics and other variables will be inspected within each region to assess if difference in demographic characteristics or other variables are the cause of any lack of consistency in treatment effect across regions.

9 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

9.1 DEMOGRAPHICS

Demographics will be summarized by randomized treatment group for the mITT and the safety analysis poopulations. There will be no formal statistical comparisons between treatment groups on demographic variables. The continuous variables of age, height (cm), weight (kg), BMI, SBP (mm/Hg), and DBP (mm/Hg) will be summarized by treatment group using sample size, mean, standard deviation, minimum and maximum. For the categorical variables of sex (male, female), race (White, Black or African American, American Indian/Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Other), and ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not Reported, Unknown), the number and percentage of patients in each category will be presented for each randomized treatment group.

9.2 PRIOR AND CONCURRENT MEDICATIONS

A listing will be provided detailing subjects' medications. In addition, for the mITT analysis set, the number and percent of subjects taking antidiabetic medications through Week 24 will be presented by treatment group overall and within each medication type (WHO ATC3 classification). The number and percent of subjects taking 1, 2, 3 and >3 anti-diabetic medications through Week 24 will also be presented by treatment group. These counts and percentages will then be presented for baseline through Week 4, Week 4 through Week 8, Week 8 through Week 12, Week 12 through Week 16, Week 16 through Week 20, and Week 20 through Week 24.

9.3 DIABETES HISTORY

The diabetes and targeted hypoglycemic events history of all ITT and mITT subjects will be summarized in a table by treatment group. For each diagnosis or symptom, the number and percent of subjects who have a history of the condition will be presented. The number and percent of subjects who have experienced any hypoglycemic events in the past year will be presented. The number of loss of consciousness, admission to hospital, and seizure events in the past year will be summarized by number and percent of subjects with values such as 0, 1, 2, >2.

The number and percent of ITT and mITT subjects reporting hypoglycemia on the daily diary at baseline will be presented by treatment group.

9.4 BASELINE MEDICAL HISTORY

The medical history of all ITT and mITT subjects will be summarized in a table by treatment group. Specifically, for each condition, the number and percent of subjects who currently have the condition will be presented. The list of conditions is captured in the REVITA Medical History CRF.

9.5 BASELINE LABS

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. The laboratory variables are white blood cell count, white blood cell differential, total hemoglobin, hematocrit, platelet count, blood urea nitrogen, calcium, chloride, creatinine, potassium, sodium, albumin, ALT, AST, total bilirubin, alkaline phosphatase, ferritin, amylase, lipase, total cholesterol, HDL, LDL, triglycerides, fasting glucose, HbA1c, C-peptide, fasting insulin, and microalbuminuria. 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.

10 EFFICACY ANALYSES

10.1 PRIMARY EFFICACY VARIABLES

The primary efficacy variables are: 1) The change from baseline at 24 weeks in HbA1c, DMR vs Sham, and 2) in patients with baseline MR-PDFF > 5%, the absolute change from baseline at 12 weeks in MR-PDFF between Fractyl DMR and Sham subjects in the mITT analysis population.

The following are the null and alternative superiority hypotheses for the primary efficacy parameter of change in HbA1c.

$$H_0: \mu_{DMR} - \mu_{Sham} > 0$$

vs.

$$H_a: \mu_{DMR} - \mu_{Sham} \leq 0$$

where μ_{DMR} and μ_{Sham} are the mean change in HbA1c from Baseline to 24 Weeks.

The following are the null and alternative superiority hypotheses for the primary efficacy parameter of change in MR-PDFF.

$$H_0: \gamma_{DMR} - \gamma_{Sham} > 0$$

vs.

$$H_a: \gamma_{DMR} - \gamma_{Sham} \leq 0$$

where γ_{DMR} and γ_{Sham} are the mean change in MR-PDFF from Baseline to 12 Weeks.

The primary analysis for each endpoint will be performed in the mITT analysis population comparing treatment groups with an Analysis of Covariance (ANCOVA) model with Multiple Imputation where post-rescue medication is first set to missing, that will adjust for study region, the baseline value of the outcome, and the difference between the screening and baseline value of the outcome for only the HbA1c endpoint (complete details on the missing data imputation are discussed in the Missing Data section above; details on missing data sensitivity analyses are also discussed in that same section). The Hochberg procedure will be used to control the experimentwise significance level at a one-sided 0.05 level across the two primary endpoints. With the Hochberg procedure, the experimental treatment is considered beneficial over control for both primary endpoints if the treatment comparison one-sided p-value for each endpoint is <0.05; otherwise, if the one-sided treatment comparison p-value <0.025 for one endpoint, treatment is considered beneficial over control for that given endpoint. A similar analysis will be conducted in the PP analysis population.

If a significant beneficial treatment effect is found in both primary endpoints using the Hochberg approach at the experimentwise one-sided 0.05 level of significance, then both primary endpoints will be compared between treatments using the Hochberg procedure at an experimentwise one-sided 0.025 level of significance. If, however, a significant beneficial treatment effect is found in only one endpoint using the Hochberg approach at the experimentwise one-sided 0.05 level of significance (i.e., the endpoint is significant at the one-sided comparisonwise 0.025 level of significance), then that endpoint alone will be compared between treatments at a one-sided 0.0125 level of significance.

For each analyses, tables of descriptive statistics of the primary endpoint will include n, mean, standard deviation, least square mean, standard error of least square mean, median, quartiles, and minimum and maximum for each treatment group. Two-sided 95% confidence intervals of the difference between treatment means and of the difference between treatment least square means will be presented. In addition to tables, there will be graphs for HbA1c. In the first graph, a line graph will be used to show the average HbA1c over time for each treatment group separately. Time will be the x-axis, and the HbA1c level will be the y-axis. The average HbA1c level ±1 standard error bars will be plotted over time with the means connected by a solid line.. In the second set of graphs, the average HbA1c (y-axis) will be displayed in side-by-side bar plots for each treatment group at each study time point (x-axis).

10.2 Secondary Efficacy Variables

The following secondary endpoints will be compared between treatments on the mITT and PP analysis populations. Where applicable, secondary endpoints will be compared between treatments at a one-sided 0.05 level of significance with the direction of the alternative hypothesis favors DMR over control. There will be no alpha-adjustment for multiple comparisons. For secondary endpoints, the only analysis to be carried out will be last-observation-carried forward (LOCF) for missing data, where post-rescue values are first set to missing.

1. HbA1c change from baseline through Week 24 (Visit 9) simultaneously across visits, DMR vs. Sham
2. The relative MR-PDFF change from baseline to Week 12 in patients with baseline MR-PDFF > 5%, DMR vs. Sham

3. Proportion of randomized-DMR-treated subjects with an HbA1c improvement from baseline at 24 weeks (Visit 9) that maintain an HbA1c improvement at 48 weeks
4. Proportion of randomized-DMR-treated subjects with an MR-PDFF > 5% at baseline and MR-PDFF improvement from baseline at 24 weeks (Visit 9) that maintain an MR-PDFF improvement at 48 weeks
5. Fasting Plasma Glucose (FPG) change from baseline at 24 weeks DMR vs. Sham
6. FPG change from baseline to Week 24 (Visit 9) simultaneously across visits, DMR vs. Sham
7. Weight change from baseline at 24 weeks DMR vs. Sham
8. In randomized-DMR-treated subjects with an HbA1c improvement from baseline at 24 weeks, average HbA1c improvement from baseline at 48 weeks
9. In randomized-DMR-treated subjects with an MR-PDFF > 5% at baseline and MR-PDFF improvement from baseline at 12 weeks, average MR-PDFF improvement at 48 weeks
10. HOMA-IR change from baseline at 24 weeks DMR vs. Sham

In addition to tables, there will be a line graph for each key secondary endpoint lab value. Time will be the x-axis, and the endpoint value level will be the y-axis. In each graph, the average lab value level ±1 standard error bars will be plotted over time with the means connected by a solid line. Separate plotted lines will be generated for each treatment, where applicable, with both lines presented on the same graph.

Secondary continuous endpoints (endpoints 2, 5, 7, 10) measured at a given time point in the randomization phase will be tested comparatively using an ANCOVA model to adjust for region and the baseline value of the outcome, and the difference between the screening value and baseline value of the outcome (for endpoints where the screening value is available).

For the secondary endpoints that are compared between treatments over time (i.e., for secondary endpoints 1 and 6 where the treatment comparison is not just at one time point such as 24 weeks, but where the treatments are compared at all visit time points simultaneously), a mixed model repeated measures approach will be used to compare treatments with respect to mean outcome over time; subject will be a random effect, and region, the baseline value of the outcome measure, and the difference between the screening and baseline value of the outcome will be used as a covariate. An unstructured within-patient covariance structure will be assumed; if the model does not converge, a compound symmetry within-patient covariance structure will be assumed.

All remaining secondary endpoints assess durability of DMR response in DMR-randomized patients with improvement and do not require formal treatment comparisons. Results will be presented with appropriate descriptive statistics. There will be no imputation of missing data for these endpoints

10.2.1 Exploratory Endpoints

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

- Change from baseline in the following at 12 weeks in DMR vs. Sham
    - MMTT change from baseline (glucose AUC through 2 hours)

  - Change in MMTT measure of Insulin secretion
  - Change in MMTT measure of Insulin resistance
  - MR-LIC Liver Iron Content

- Change from baseline in the following by visit out to 24 weeks in DMR vs. Sham
  - FPI
  - Fasting C-peptide
  - Weight
  - ALT
  - AST
  - FIB-4
  - Ferritin
  - UACR
  - eGFR
  - Triglycerides
  - HDL
  - TG/HDL

- Change from baseline in the following at each of 4 weeks, 12 weeks, and 18 weeks, DMR vs. Sham:
  - HbA1c
  - FPG

- Proportion of randomized-DMR-treated subjects with an abnormal MR-LIC at baseline and MR-LIC improvement from baseline at 12 weeks that maintain an MR-LIC improvement at 48 weeks.

- In randomized-DMR-treated subjects with an abnormal MR-LIC at baseline and MR-LIC improvement from baseline at 12, average MR-LIC improvement at 48 weeks.

- Training Cohort only:
  - Change in SBP by ABPM from baseline in training cohort, at 12 weeks
  - Change in DBP by ABPM from baseline in training cohort, at 12 weeks

11 SAFETY ANALYSES

The safety analysis is performed separately on the safety population for the training cohort and the randomization cohort. The randomization cohort analysis is performed by treatment administered and will be presented for safety data collected during the randomized double-blind phase (24 weeks from index procedure). As a secondary analysis, analyses will also performed for all subjects undergoing Fractyl DMR after crossover.

11.1 PRIMARY SAFETY VARIABLE

The primary safety endpoint is the incidence rate of the device or procedure related Serious Adverse Events (SAEs), Unanticipated Adverse Device Effects (UADEs) and Adverse Events of Special Interest (AESIs) through the 24 weeks post treatment initiation. The safety endpoint summary will include the

number and percentage of subjects in each of the categories overall and by MedDRA System Organ Class (SOC) and Preferred Term (PT); for the AEs occurring during the randomized phase, these numbers and percentages will be presented within each treatment group. For the overall counts (i.e., for results not by SOC and PT), two-sided Clopper-Pearson 95% CI of the percentages will be presented. A formal hypothesis test comparing treatments is not planned. Further details on adverse event analyses are provided below.

11.2 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 summarized for each treatment group in the double-blind phase. 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: All AEs are coded using Medical Dictionary for Regulatory Activities (MedDRA). Treatment emergent AEs, adverse events of special interest (AESIs), SAEs and UADEs will be summarized for the training cohort and randomization cohort through 24 weeks for each treatment by number and percentage of patients with at least one adverse event overall and by primary System Organ Class (SOC) and Preferred Term (PT). AESIs are characterized as:

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

Specifically, irrespective of whether an AE is serious or non-serious, the following events are defined as ‘protocol-specified adverse events of special interest’ and have additional reporting requirements:

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

The summarizations by SOC and PT will bes repeated for (a) treatment emergent events occurring in the 24 to 48-week period (i.e., events occurring/worsening after Week 24 visit for patients originally receiving DMR, and events occurring/worsening at or after receiving DMR for the crossover patients), separated by original treatment received at start of the study (DMR and sham); and (b) treatment emergent events occurring in the 0-48 week period for patients receiving DMR at the start of the study. 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 is categorized as no, possibly, probably, and definitely. The highest level of association is reported for subjects with different relationships for the same AE. Details of AE analyses are provided below.

11.3 Adverse Events

All adverse events (AEs) will be coded using the standardized MedDRA central coding dictionary, version 19.1 or greater. Adverse event analyses will be performed on the Safety analysis population.

11.3.1 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 overall by SOC and PT. A TEAE is an event starting or worsening in severity at or after initiation of the index procedure for the randomized 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 randomized phase, results will be presented by treatment group. This analysis will be repeated for TEAEs of special interest (TEAESIs).

The proportion of patients with at least one 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 AE SOC and PT, 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.

11.3.2 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, by SOC and PT will be presented in a similar manner as discussed above (with the exception of the plot). A data listing of TEAEs leading to withdrawal will also be provided, displaying details of the event(s) captured on the CRF.

11.3.3 Serious Adverse Events

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

11.3.4 Device and Procedure Related Adverse Events

Summaries of device-related TEAEs and summaries of procedure-related TEAEs will be conducted in the same manner as for all TEAEs discussed above; patients with he occurrence of more than one TEAE within a given SOC or PT will be counted only under the maximum severity/relationship experienced for that SOC or PT, respectively.

11.3.5 Unanticipated Adverse Device Effects

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

11.3.6 Hypoglycemia

Special focus is given to adverse events of special interest of hypoglycemia, clinically significant hypoglycemia (glucose level <54 mg/dL or <3.0 mmol/L), and severe hypoglycemia (hypoglycemia associated with severe cognitive impairment requiring external assistance for recovery). The adverse event analysis discussed in section 11.2 will be carried out separately for these hypoglycemia events.

11.3.7 Deaths

If a death occurs during the course of the REVITA 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.

12 Other Planned Analyses

12.1 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 endpoint of the double-blind phase will be presented within each of the following subgroups (this will be performed on the mITT population using multiply imputed data, where efficacy endpoint values collected after rescue medication use are first set to missing):

A. Sex
B. Age (<Median, ≥Median)
C. Race
D. Ethnicity
E. Body Mass Index (BMI; (<Median, ≥Median)
F. Geographic Region
G. Baseline value of the outcome (<Median, ≥Median); e.g., for analyses on the outcome of change from baseline to Week 24 HbA1c, treatments will be compared within the set of subjects with baseline HbA1c <median and within the set of subjects with baseline HbA1c ≥ 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.

12.2 Training Cases

The training cases will be evaluated separately from the randomized subjects. Specifically, as discussed above, descriptive statistics of the change from baseline through 48 weeks will be presented for efficacy variables for the training cases. Incidence of TEAEs, discussed above in Section 11.3.1 – 11.3.6 for the double-blind phase, will be presented in a similar manner for the training cases.

13 Reporting Conventions

All reporting will meet the standards of BBA SOP BS002 and its associated work instructions.

14 Attachment 1. Planned Tables, Figures and Listings

Addendum to the Statistical Analysis Plan

Protocol Title (Number):

Evaluation of the effect of Duodenal Mucosal Resurfacing (DMR) using the Revita System in the Treatment of Type 2 diabetes (T2D)

(C-30000: REVITA-2)

Addendum 1 to Rev. A of the SAP

Prepared by:

BBA

01AUG2019

This addendum to the SAP has been written after the blind has been broken and after initial analysis revealed important statistical considerations not adequately addressed within SAP Rev A. The analyses described within this addendum address these statistical considerations by clarifying and adding detail for statistical tests already specified within the SAP, or in some cases, refining the SAP to be more consistent with standard statistical methodology.

Note: In the original protocol and the original SAP, MRI-PDFF should have been specified as MRI-PDFF, but was incorrectly specified as MR-PDFF. Herein this document, it will be specified correctly as MRI-PDFF. In all analyses and documents going forward, it will be specified as MRI-PDFF.

Note: In the original SAP, the post-rescue medication definition was unclear (the per protocol definition was specified in the study design section of the SAP and a separate definition was used specifying only an increase in oral-diabetic medications. As per SAP addendum, we will be defining the post-rescue medications as per protocol to keep both the protocol and SAP consistent.)

1 Analyses to Assess the Normality Assumption of Primary and Secondary Endpoints

Section 8.1 of the SAP indicates that "distributions of each continuous variable are assessed prior to analysis and examined for normality", but does not specify how normality would be assessed. The details on assessment of normality are now provided in the current section.

The original SAP also makes only very general statements about how to analyze data when it is found to not be normally distributed. Standard practice would be to then use nonparametric methods. Details are provided in Section 2 on the nonparametric methods that would be used if the data for a given endpoint is found to not be normally distributed.

Note: If there is genuine belief that an outlier is an error in the dataset because it is a clinically implausible value, then this will be further investigated, and if not able to be corrected by information obtained from the site then it will be set to missing in all analysis datasets because retaining such values could undermine the accuracy of the statistical analyses. Setting such values to missing will be carried out prior to conducting any analyses (including prior to the assessment of the normality assumption). However, any such cases will be summarized in an appendix to the clinical study report, with the full rationale provided on why such data values are regarded as clinically implausible.

The following analyses will be performed for the mITT population in order to assess the normality assumption for the primary and secondary efficacy endpoints:

1. Q-Q plot and histogram of residuals from an ANCOVA model (with no imputation) on HbA1c change from baseline to Week 24 (excluding patients taking rescue medication

prior to Week 24, and excluding patients with missing HbA1c Week 24), including baseline HbA1c, change from screening to baseline HbA1c, region (as an unordered categorical/class variable, defined as specified in Section 8.6 of the SAP), and treatment group in the model. The Shapiro-Wilk test of normality will also be performed on the residuals from this model.

2. Q-Q plot and histogram of residuals from an ANCOVA model (with no imputation) on MRI-PDFF change from baseline to Week 12 (excluding patients with missing MRI-PDFF at baseline or at Week 12, excluding patients with baseline MRI-PDFF ≤5%, and excluding patients taking rescue medication prior to Week 12), with baseline MRI-PDFF, region (as an unordered categorical/class variable, defined as specified in Section 8.6 of the SAP), and treatment group in the model. The Shapiro-Wilk test of normality will also be performed on the residuals from this model.

3. Q-Q plot and histogram of residuals from an ANCOVA model (with no imputation) on FPG change from baseline to Week 24 (excluding patients with rescue medication and excluding patients with missing FPG Week 24), including baseline FPG, change from screening to baseline FPG, region (as an unordered categorical/class variable, defined as specified in Section 8.6 of the SAP), and treatment group in the model. The Shapiro-Wilk test of normality will also be performed on the residuals from this model.

4. If #1 demonstrates non-normality, then we will not carry out similar tests of normality for HbA1c change from baseline to earlier time points. If #3 demonstrates non-normality, then we will not carry out similar tests of normality for FPG change from baseline to earlier time points. Assessment of normality for the other secondary endpoints (i.e. Relative MRI-PDFF change from baseline to Week 12, Weight change from baseline to Week 24, and HOMA-IR change from baseline to Week 24) will be performed similar to #3 above, but for relative MRI-PDFF CFB to Week 12 patients to be included will be as described in #2. For relative MRI-PDFF CFB to Week 12, change from screening to baseline will not be included in the model, similar to #2 above.

Note: The models within #1-#4 above are identical to the models specified in the original SAP as the primary analysis for each of these endpoints, except that testing of normality does not include imputation.

For each primary and secondary endpoint assessed for normality, the following criteria will be used in determining whether the normality assumption is met:

- If the p-value from the Shapiro-Wilk test is < 0.05, then we would reject the normality assumption and we would regard the rank-based method (described in Section 2 below) as the primary method to analyze the endpoint.

- If the Shapiro-Wilk test p-value ≥ 0.05, but the Q-Q plots and/or histograms show clear evidence of non-normality, then we would reject the normality assumption and regard the rank-based method (described in Section 2 below) as the primary method to analyze the endpoint.
- Otherwise, if the Shapiro-Wilk test and plots (Q-Q plots and/or histograms) do not provide evidence of non-normality, then we would assume the normality assumption is met and regard the analysis with the response variable on the original scale as the primary analysis for the endpoint.

Note: For each primary or secondary endpoint in which the normality assumption is not met, all of its analyses will be carried out on the rank scale in the manner described in Section 2.

Note: If the normality assumption for a particular primary or secondary endpoint is met, then the response variable will be analyzed on the original scale. However, we would assess whether the raw baseline values or difference between raw screening and baseline values (for endpoints with a screening assessment) are non-normal. These data will be assessed for normality using Shapiro-Wilk test, Q-Q plots, and histograms. If baseline and/or difference between screening and baseline were found to not be normally distributed based on the criteria described above, we would then use rank of baseline and rank of change from screening to baseline (for endpoints with a screening assessment) as covariates for all analyses of this endpoint even though the change from baseline response variable will still be in its original scale.

Note: For each primary or secondary endpoint, an additional secondary analysis, as described in Section 7, will be provided on the opposite scale to the scale used for the response variable in its primary analysis.

2 Analyses for Primary Endpoints When Normality Assumption is Not Met

This section of the addendum is meant to clarify how primary endpoints are analyzed when the normality assumption is not met.

ANCOVA on Ranks using Multiple Imputation (MI) for HbA1c change from baseline at Week 24

If non-normality is demonstrated (based on the criteria described in Section 1 of this addendum) for the first primary efficacy endpoint of change from baseline (CFB) at week 24 in HbA1c, then ANCOVA on ranks with multiple imputation will be used and regarded as the primary method for analyzing the endpoint. This analysis would be the rank-based equivalent to the primary analysis that was specified for HbA1c in SAP Section 10.1. This analysis will be performed in

the mITT and PP analysis population where data post-rescue medication is first set to missing. For this analysis, the following will be derived: (i) rank of HbA1c CFB to week 24; (ii) rank of baseline HbA1c; (iii) rank HbA1c screening minus baseline value; (iv) rank values for all additional continuous variables included in the imputation model (baseline FPG, baseline fasting plasma insulin [FPI], change from screening to baseline FPG, change from screening to baseline FPI, duration of Type 2 diabetes, BMI, age, and change from baseline FPI, FPG, and HbA1c at all scheduled visits prior to Week 24). Due to the nature of rank-based analyses, in accordance with standard statistical methodology, the imputation model will derive CFB HbA1c to Week 24. Therefore, the imputation model will include rank of CFB FPI, CFB FPG, and CFB HbA1c variables for all weeks prior to week 24 rather than rank-based FPI, FPG, and HbA1c. This is an appropriate method for rank based analyses because rank (CFB HbA1c to Week 24) ≠ rank (Week 24 HbA1c) - rank (Baseline HbA1c). Each rank variable is derived on the modified ridit scale. Then, the multiple imputation procedure as described in SAP Section 8.3 will be carried out to derive change from baseline (CFB) at 24 weeks in HbA1c but with all continuous variables now based on the rank scale. ANCOVA will then be carried out on each of the 50 imputed datasets as described in the SAP Section 10.1 but with (i) as the response variable and with the model including terms for (ii), (iii), treatment, and region (as a class/unordered categorical variable). Then, the estimates from this analysis will be combined across the datasets using SAS PROC MIANALYZE as described in the SAP Section 8.3.

<u>ANCOVA on Ranks for MRI-PDFF change from baseline at Week 12</u>

If non-normality is demonstrated (based on the criteria described in Section 1 of this addendum) for the second primary efficacy endpoint of change from baseline (CFB) at week 12 in MRI-PDFF, then ANCOVA on ranks with multiple imputation will be used and regarded as the primary method for analyzing the endpoint. This analysis would be the rank-based equivalent to the primary analysis that was specified for MRI-PDFF in SAP Section 10.1. This analysis will be performed in the mITT and PP analysis population where data post-rescue medication is first set to missing, and subjects that have baseline MRI-PDFF <=5% are excluded. The original SAP indicated that for the MRI-PDFF CFB to Week 12 primary endpoint in patients with baseline MRI-PDFF >5%, "the imputation approach will be similarly carried out" to the approach used for the first primary endpoint (CFB at week 24 in HbA1c). However, the imputation model used for HbA1c includes predictors that are not likely to be predictors of CFB MRI-PDFF at Week 12. Weight variables will be included in the imputation model as well because the published literature has clearly established that weight is strongly correlated with MRI-PDFF CFB to Week 12. For this analysis, the following will be derived: (i) rank of MRI-PDFF CFB to week 12; (ii) rank of baseline MRI-PDFF ; and (iii) rank values for all additional continuous variables included in the imputation model (duration of Type 2 diabetes, age, change from screening to baseline HbA1c, baseline HbA1c, baseline BMI, baseline weight, baseline MRI-PDFF, CFB to Week 4 HbA1c, weight change from baseline to week 4, and weight change from baseline to week 12). Each rank variable is derived on the modified ridit scale. Then, the multiple imputation procedure as described in SAP Section 8.3 will be carried out to derive change from baseline (CFB) at 12 weeks in MRI-PDFF but with all continuous variables now based on the

rank scale. The same reason provided for deriving rank-based imputation above for CFB HbA1c to Week 24 applies to rank-based imputation for CFB to week 12 MRI-PDFF, where CFB variables will be included in the imputation model for variables prior to Week 12. ANCOVA will then be carried out on each of the 50 imputed datasets as described in the SAP Section 10.1 but with (i) as the response variable and with the model including terms for (ii), treatment, and region (as a class/unordered categorical variable). Then, the estimates from this analysis will be combined across the datasets using SAS PROC MIANALYZE as described in the SAP Section 8.3.

As described in Section 1, if baseline HbA1c or HbA1c screening minus baseline is not normally distributed, then even if the response variable (CFB HbA1c at 24 weeks) is on the original scale, both baseline HbA1c and HbA1c screening minus baseline would be on the rank scale. If baseline MRI-PDFF is not normally distributed then even if the response variable is on the original scale, baseline would then be on the rank scale.

Note: For all analyses, in which the response variable is rank-based, LS-Means will not be reported in tables, listings, or figures (TLFs) since LS-Means are not useful on the rank scale. For such analyses, the clinical study report (CSR) will instead focus on the p-values and medians.

3 Analyses for Secondary Endpoints When Normality Assumption is Not Met

For each secondary endpoint where the normality assumption is not met (based on the criteria described in Section 1 of this addendum) and which was pre-specified in the SAP Section 10.2 to be analyzed using ANCOVA, then the treatment effect will be assessed using an ANCOVA model based on ranks (using modified ridits) with rank for the endpoint (e.g. FPG change from baseline at Week 24) as the response variable and adjusting for region (as a class/unordered categorical variable), rank for baseline value of the outcome (e.g. rank for baseline FPG), and rank for the difference between screening value and baseline value of the outcome (for endpoints where the screening value is available) (e.g. rank for FPG baseline minus screening). As specified in the SAP Section 10.2, last-observation-carried forward (LOCF) will be carried out for missing data, where post-rescue values are first set to missing. LOCF will only be carried out for secondary endpoints for which there is a preceding post-baseline visit at which the endpoint is scheduled to be assessed. For MRI-PDFF relative change, even though in SAP Section 10.2, it was specified for the secondary endpoints, that LOCF will be carried out; this is not valid because there is no other observation for MRI-PDFF except for baseline. Therefore, for MRI-PDFF relative change, no LOCF imputation will be included.

For each secondary endpoint where the normality assumption is not met (based on the criteria described in Section 1) and was pre-specified in the SAP Section 10.2 to be analyzed using MMRM, then the treatment effect will be assessed using the MMRM model specified in SAP Section 10.2 but now based on the rank scale (using modified ridits) for all continuous variables

and where MMRM is used but with no LOCF, and post rescue values first set to missing. The rank scale will be used for change from baseline at all visits in the MMRM analyses.

Per SAP Section 10.2, secondary endpoints will be compared between treatments on the mITT and PP analysis populations.

4 Clarification Regarding the Use of MMRM or LOCF for Secondary Endpoints

Section 10.2 of the SAP states, "for secondary endpoints, the only analysis to be carried out will be last-observation-carried forward (LOCF) for missing data…" LOCF assumes data is missing completely at random (MCAR). However, mixed model for repeated measures (MMRM) is a method of dealing with missing data, which makes the more stringent assumption of missing at random (MAR), in the same way that multiple imputation assumes MAR. Therefore, for secondary endpoints that are pre-specified in the SAP to be analyzed using MMRM (e.g. HbA1c change from baseline through Week 24 simultaneously across visits), no additional LOCF imputation will be carried out.

5 Clarification of Multiple Imputation

Section 8.3 of SAP Rev. A specified that the linear regression multiple imputation (MI) model will include treatment group, study center (centers with fewer than 5 subjects will be pooled with another study center by geographic region to ensure at least 5 subjects per study center), baseline HbA1c, baseline FPG, baseline FPI, change from screening to baseline HbA1c, change from screening to baseline FPG, change from screening to baseline FPI, duration of Type 2 diabetes, age, sex, BMI, and FPI, FPG, and HbA1c measured at scheduled visits prior to week 24.

In Section 10.1, the SAP Rev. A states that treatment groups will be compared using an Analysis of Covariance (ANCOVA) model with Multiple Imputation that adjusts for study region, baseline value of the outcome, and difference between the screening and baseline value of the outcome for the HbA1c primary endpoint and ANCOVA with Multiple Imputation that adjusts for study region, and baseline value of the outcome for the MRI-PDFF primary endpoint. Per the original SAP, the pre-specified regions in the study were Belgium, Brazil, Italy, Netherlands, and United Kingdom.

Since we are adjusting for region in the analysis, we include region (instead of study center) in the MI model. Any covariates used in the analysis model should be in the MI model (however, not all covariates in the MI model need to be in the analysis model). Note that this change was made prior to running the analyses stated in the SAP and prior to unblinding.

For both primary endpoints in the analysis model as well as for any post-baseline values in the imputation model that are after the patient has taken rescue medication then such values will be first set to missing before carrying out the imputation process.

The exact ordering of the terms included in the monotone linear regression imputation model for imputing missing CFB to week 24 HbA1c will be: treatment group, region, duration of Type 2 diabetes, age, sex, change from screening to baseline FPI, change from screening to baseline FPG, change from screening to baseline HbA1c, baseline BMI, baseline FPI, baseline FPG, baseline HbA1c, CFB to Week 4 FPI, CFB to Week 4 FPG, CFB to Week 4 HbA1c, CFB to Week 12 FPI, CFB to Week 12 FPG, CFB to Week 12 HbA1c, CFB to Week 18 FPI, CFB to Week 18 FPG, and CFB to Week 18 HbA1c.

The exact ordering of the terms included in the monotone linear regression imputation model for imputing missing CFB to week 12 MRI-PDFF will be: treatment group, region, duration of Type 2 diabetes, age, sex, change from screening to baseline HbA1c, baseline HbA1c, baseline BMI, baseline weight, baseline MRI-PDFF, CFB to Week 4 HbA1c, weight change from baseline to week 4, and weight change from baseline to week 12.

6 ASSESSMENT OF HOMOGENEITY ACROSS REGIONS

Section 8.6 of the SAP states in order "to evaluate consistency in treatment effect across geographic regions (Belgium, Brazil, Italy, Netherlands, United Kingdom) for each of the two primary endpoints, summary tables of treatment differences by region will be presented for each primary endpoint using the multiply imputed data where post-rescue medication is set to missing prior to carrying out the imputation."

For endpoints where the normality assumption is not met, these analyses will be carried out on the rank (modified ridit) scale. All interaction analyses will be carried out based on the mITT analysis population.

6.1 ASSESSMENT OF HOMOGENEITY ACROSS REGIONS FOR PRIMARY ENDPOINTS WHEN NORMALITY ASSUMPTION IS MET

Per SAP Rev. A, the pre-specified regions in the study were Belgium, Brazil, Italy, Netherlands, and United Kingdom. The significance of treatment-by-region interaction on the first primary endpoint (HbA1c CFB to Week 24) was assessed per SAP using analysis of covariance (ANCOVA) with multiple imputation (where values post-rescue medication are set to missing before carrying out the imputation) and effects for treatment, region, baseline value of the outcome, the difference between the screening and baseline value of the outcome, and treatment-by-region interaction.

For the second primary endpoint (MRI-PDFF CFB to Week 12) in accordance with the SAP, the model contains terms for treatment, region, baseline value of the outcome, and treatment by region interaction. Values post-rescue medication will be set to missing before carrying out the imputation.

The SAP specified five regions and so the treatment by region interaction tests for the primary endpoints are based on 4 degrees of freedom, using a 0.10 level of significance.

Regardless of whether the 4 degree of freedom test for a given primary endpoint obtains statistical significance, additional 1 degree of freedom tests of the treatment by region interaction tests (Brazil vs. other countries combined, Belgium vs. other countries combined, Netherlands vs. other countries combined, Italy vs. other countries combined, and UK vs. other countries combined) will be carried out on the mITT analysis population using a 0.10 level of significance. These additional 1 degree of freedom interaction tests will be carried out because the use of a 4 degree of freedom interaction test (due to having five regions) has low power given the small sample size within each region[1, 2, 3]. These interaction analyses based on 1 degree of freedom are viewed as sensitivity analyses for the pre-specified treatment by region interaction tests for the primary endpoints. If the interaction p-value is < 0.10 for either primary endpoint, further exploration of the cause of this statistical significance will be conducted.

If any of the interaction p-values are <0.10 for either primary endpoint, then demographics and other variables will be assessed per SAP Rev A. to determine the poolability of regions for both primary endpoints and secondary endpoints. As a point of clarification, if differences in demographic characteristics or other variables are the cause of any lack of consistency in treatment effect across regions, then regions may be analyzed separately. In this case, separately for the two categories of region for the given primary endpoints, the two treatment groups will be compared using ANCOVA with multiple imputation (as described above for HbA1c and MRI-PDFF). These analyses will be carried out on the mITT analysis population, and will include the same covariates as included for that endpoint's primary analysis, but with no term for region in the model. Due to the smaller sample sizes within individual regions, the imputation model for deriving CFB HbA1c at Week 24 when analyzing separately by region will then just use terms for treatment group, HbA1c change from screening to baseline, baseline HbA1c, CFB to Week 4 HbA1c, CFB to Week 12 HbA1c, and CFB to Week 18 HbA1c, in that specific order. Also due to the smaller sample size, the number of imputations will be increased to 1000. The imputation model for deriving CFB MRI-PDFF at Week 12 when analyzing separately by region will include the same terms as identified in Section 5 for MRI-PDFF, but the number of imputations will be increased to 1000.

6.2 Assessment of Homogeneity Across Regions For Primary Endpoints When Normality Assumption is Not Met

For each primary endpoint where the normality assumption is not met (based on the criteria described in Section 1 of this addendum), homogeneity across regions will be assessed analogously to the methods described in Section 6.1, but where in each case the response variable, and all continuous covariates will be on the rank scale (using modified ridits). For HbA1c CFB to Week 24 all continuous variables in the imputation model will also be based on the rank scale (using modified ridits).

6.3 Additional Analyses Regarding Assessment of SECONDARY Endpoints BY REGION

The model for analysis of all secondary endpoints for the double-blind period (change from baseline in relative MRI-PDFF at 12 weeks among those with baseline MRI-PDFF > 5%, change from baseline in FPG at 24 weeks, change from baseline in HOMA-IR at 24 weeks, and change from baseline in weight at 24 weeks) will be analyzed separately by region according to the conclusion made in the interaction tests for the primary endpoints in Section 6.1 (or 6.2) and as described in SAP Section 10.2 except that it will not include a term for region, and in accordance with the SAP, imputation will be carried out by LOCF (after setting to missing any data post-rescue) for those secondary endpoints for which there is a preceding post-baseline value. For relative change in MRI-PDFF, even though in SAP Section 10.2, it was specified for the secondary endpoints, that LOCF will be carried out, this is not valid because there is no other observation for MRI-PDFF except for baseline. Therefore, for relative change in MRI-PDFF, LOCF will not be used.

Whether or not the analyses for a given secondary endpoint will be on the rank scale using modified ridits or on the original scale (and in the latter case, also whether the covariates are on the original scale or on the rank scale) will be as determined using the procedures described in Section 1.

7 Sensitivity Analyses for Handling Missing Data

The SAP specifies that the primary analysis of the primary endpoints will use multiple imputation using linear regression to impute missing data (where data collected post-rescue medication use are first set to missing prior to carrying out the imputation). The SAP specified the following sensitivity analyses:

a) Available case analysis / post-rescue set to missing: missing data for the primary endpoints are not imputed and data collected post-rescue medication use are set to missing. Per SAP, available case analysis will be conducted using mixed models repeated measures (MMRM).
b) Multiple imputation / post-rescue NOT set to missing: multiple imputation method specified for the primary analysis will be repeated where data collected post-rescue medication are NOT set to missing prior to carrying out the analyses.
c) Available case analysis / post-rescue NOT set to missing: missing data for the primary endpoints are not imputed and data collected post-rescue medication use are NOT set to missing. Per SAP, available case analysis will be conducted using mixed models repeated measures (MMRM).

Please note that since MRI-PDFF has only one post-baseline visit, ANCOVA will be used instead of MMRM in the sensitivity analyses stated above (i.e., a and c) for the MRI-PDFF

primary endpoint. The sensitivity analyses stated above will be performed on the mITT and PP populations.

Unless otherwise specified explicitly, the scale (rank scale or original scale) for the sensitivity analyses of the primary and secondary endpoints, and the scale on which the covariates are analyzed will now be determined by the rules given in Section 1.

For the primary endpoint of HbA1c change from baseline to Week 24, the following further sensitivity analysis of the multiple imputation process will be carried out for the mITT population:

- The primary analysis as specified in Section 8.3 of the SAP, but where the number of imputations will be set to 1000 rather than 50, where this analysis will be carried out on the scale (rank scale or original scale) using the decision process as described in Section 1, but using the following reduced imputation model: treatment group, region, duration of Type 2 diabetes, age, sex, change from screening to baseline HbA1c, baseline BMI, baseline HbA1c, CFB Week 4 HbA1c, CFB Week 12 HbA1c, and CFB Week 18 HbA1c.

For the primary endpoint of MRI-PDFF change from baseline to Week 12, the following further sensitivity analysis of the multiple imputation process will be carried out for the mITT population:

- The primary analysis as specified in Section 8.3 of the SAP, but where the number of imputations will be set to 1000 rather than 50, where this analysis will be carried out on the scale (rank scale or original scale) using the decision process as described in Section 1.

The following sensitivity analyses, which are completers analyses, but excluding patients that took prior rescue medication, will also be performed for the mITT and PP populations:

- ANCOVA (with no imputation) on HbA1c change from baseline to Week 24 (excluding patients with missing HbA1c Week 24, and data post-rescue set to missing), including baseline HbA1c, change from screening to baseline HbA1c, region (as an unordered categorical/class variable), and treatment group in the model.
- ANCOVA (with no imputation) on MRI-PDFF change from baseline to Week 12 (excluding patients with baseline MRI-PDFF <=5%, patients with missing MRI-PDFF Week 12, and data post-rescue set to missing), including baseline MRI-PDFF, region (as an unordered categorical/class variable), and treatment group in the model.

- The same analyses as described above for HbA1c CFB to Week 24 will be carried out for the following secondary endpoints: FPG CFB at Week 24, Relative MRI-PDFF CFB at Week 12, Weight CFB at Week 24, and HOMA-IR CFB at Week 24. Note: Change from

screening to baseline in MRI-PDFF is not available and will not be included in the model for the relative MRI-PDFF CFB at Week 12 endpoint.

Section 1 of the Addendum describes the assessment of normality and determines the scale (original scale or rank scale) which will be used for the response variable in the primary analysis for a given primary or secondary endpoint and the scale (original scale or rank scale) that will be used for the covariates in this primary analysis. In addition, sensitivity analyses are described below which are carried out on the opposite scale to that used for the primary analysis of the response variable:

- For primary endpoints where the normality assumption is met a secondary analysis will be carried out using the method described in Section 2 of the Addendum
- For secondary endpoints where the normality assumption is met a secondary analysis will be carried out using the method described in Section 3 of the Addendum
- For the first primary endpoint (HbA1c change from baseline to week 24) and for the secondary endpoints, when the assumption of normality is not met a secondary analysis will be carried out as described in the original SAP with the response variable on the original scale, but whether or not the covariates are based on the original scale or on the rank scale will be determined as described in Section 1 of the Addendum.
- For the second primary endpoint (MRI-PDFF change from baseline in patients with baseline MRI-PDFF > 5%) where the assumption of normality is not met a secondary analysis will be carried out with the response variable on the original scale using ANCOVA with terms for treatment, region, and baseline, data post-rescue set to missing, and excluding patients with missing Week 12 values. Whether or not baseline is based on the original scale or on the rank scale will be determined as described in Section 1 of the addendum.

References

1. Day, S.J. and Graham, D.F. (1989) Sample Size and Power for Comparing Two or More Treatment Groups in Clinical Trials. *BMJ: British Medical Journal.* 299 (6700): 663-665.
2. Schmidt, A.F., Groenwold, R.H.H., Knol, M.J., Hoes, A.W., Nielen, M., Roes, K.C.B, de Boer, A., and Klungel, O.H. (2014). Exploring interaction effects in small samples increases rates of false-positive and false-negative findings: results from a systematic review and simulation study. *Journal of Clinical Epidemiology.* 67: 821-829.
3. Sullivan, L.M., Weinberg, J., and Keaney, J.F. (2016). Common Statistical Pitfalls in Basic Science Research. *J Am Heart Assoc.* 5: e004.