

Study: FL001 - SAP ver3

Flow Neuroscience AB

Statistical analysis plan - FL001

Study Title: Transcranial Direct Current Stimulation In Major Depressive Disorder: A Double-Blind, Placebo-Controlled, Randomized, Superiority Trial

Clinicaltrials.gov number: NCT05202119

Document date: 13th January 2023

Sponsor: Flow Neuroscience AB

Statistical analysis plan - FL001

Version history:

1. 21Mar2022: ver1: Added details of MMRM and Missing data handling.
2. 07Dec2022: ver2: Re-organize document; added symbolic hypotheses; added interim analysis details. Clarified secondary endpoints.
3. 13Jan2023: Adjustments to align with new Adaptive design report.

1. Trial Overview	3
2. Study Populations	3
2.1. Intent-to-Treat Analysis Set (ITT)	3
2.2. As Treated Analysis Set (AT)	4
2.3. Modified Intent-to-Treat Analysis Set (mITT)	4
2.4. Per Protocol Analysis Set (PP)	4
3. Study Endpoints	4
3.1. Primary Effectiveness Endpoint	4
3.1.1. Testing of Primary Effectiveness Hypotheses	4
3.1.2. Multiple Imputation Analysis	5
3.1.3. Secondary Endpoints	5
3.1.4. Discussion on FDA Requested secondary endpoints and clarification of the endpoints controlled for multiplicity	5
3.1.5. Secondary Endpoint #1: Change in MADRS score	5
3.1.5.1. Endpoint Definition	6
3.1.5.2. Hypothesis Test	6
3.1.5.3. Analysis Set	6
3.1.5.4. Missing Data Analysis	6
3.1.6. Secondary Endpoint #2: Clinical Response	6
3.1.6.1. Endpoint Definition	6
3.1.6.2. Hypothesis Test	6
3.1.6.3. Analysis Set	7
3.1.6.4. Missing Data Analysis	7
3.1.7. Secondary Endpoint #3: Remission	7

3.1.7.1. Endpoint Definition	7
3.1.7.2. Hypothesis Test	7
3.1.7.3. Analysis Set	7
3.1.7.4. Missing Data Analysis	8
3.1.8. Secondary Endpoint #4: Quality of Life Improvement	8
3.1.8.1. Endpoint Definition	8
3.1.8.2. Hypothesis Test	8
3.1.8.3. Analysis Set	8
3.1.8.4. Missing Data Analysis	8
3.1.8.5. Additional analyses	8
3.1.9. Multiplicity Control for Secondary Endpoints	8
3.1.9.1. Hochberg Algorithm	9
3.1.10. Exploratory Endpoints	10
4. Details of the MMRM	11
5. Missing data	12
5.1. Procedures For Minimizing The Occurrence Of Missing Data	12
5.2 Data Collection For Subjects Who Discontinue From Randomized Treatment Or Discontinue From The Study	13
5.2.1. DISCONTINUATION DEFINITIONS AS PROVIDED IN THE PROTOCOL	13
5.2.2. DISCONTINUATION FROM STUDY BY SUBJECT	13
5.2.3. DISCONTINUATION FROM RANDOMIZED TREATMENT BY INVESTIGATOR	14
5.2.4. TEMPORARY DISCONTINUATION BY SUBJECT	14
5.2.5. DISCONTINUATION OF INADVERTENTLY ENROLLED SUBJECTS	14
5.2.6. LOST TO FOLLOW-UP (LTFU)	14
5.2.7. USE OF DATA FROM EARLY DISCONTINUATION CASES	14
5.3. SPECIFICATION OF THE ESTIMAND	15
5.3.1. PRIMARY ESTIMAND	15
5.4. Handling Of Missing Data	16
5.4.1. HANDLING DATA FOLLOWING SUBJECT WITHDRAWAL DUE TO LACK OF TREATMENT EFFECTIVENESS	16
5.4.2. HANDLING OF SUBJECTS REASONABLY MISSING AT RANDOM	16
5.4.3. HANDLING OF SUBJECTS WITH MISSINGNESS FOR ALL OTHER REASONS	16
5.4.4. MULTIPLE IMPUTATION - GENERAL	17

5.4.5. DEVELOPMENT OF MI MODEL	17
5.4.6. COVARIATES IN THE MI MODEL	17
5.4.7. CHOICE OF MI MODEL	17
5.4.8. BASELINE COVARIATES	18
5.4.9. LONGITUDINAL HDRS-17	18
5.5. Sensitivity Analysis	19
5.5.1. ALTERNATIVE MODELING APPROACHES	19
6. Interim Analysis	19
7. Sample Size Assumptions and Calculation	19
8. Additional Analyses	20
8.1. Evaluation of Blinding	20
8.2. Device Survival	20
8.3. Poolability Analysis	20
8.4. Safety Analyses	20

1. Trial Overview

This trial is a double-blind, placebo-controlled, randomized, superiority, remote trial for individuals with major depressive disorder. The primary objective of this trial is to show that active stimulation with the Flow FL-100 device is superior to sham stimulation for the treatment of major depressive disorder when used at-home. The study is expected to take approximately 10 months from first participant enrolled to the last open-label visit, which consists of a 5 months enrollment period (estimation) and 5 months treatment and open-label. The follow-up period during the clinical investigation shall permit the demonstration of performance over a period, sufficient to represent a realistic test of the performance of the investigational device and allow any risks associated with adverse device effects over that period to be identified and assessed. For each participant, the trial will be conducted in 2 phases: blinded and open label. During the treatment phase, the participants will be randomized into one of the two arms: active and sham. The participants will be divided equally between the arms and will not be informed of their assignment. Subjects will be followed for 10 weeks, at which point the primary endpoint will be assessed.

2. Study Populations

2.1. Intent-to-Treat Analysis Set (ITT)

All randomized subjects will be included in the "intent-to-treat" (ITT) analysis set and classified according to their intended treatment regardless of failure to complete the required follow-up examinations. Subjects excluded prior to randomization will be considered screening failures.

2.2. As Treated Analysis Set (AT)

The As Treated Analysis Set (AT) will consist of ITT subjects who receive at least 1 treatment of study device (Active or Sham). Subjects will be analyzed according to the study treatment received. The As Treated Analysis Set will be used for all summaries of safety and tolerability data.

2.3. Modified Intent-to-Treat Analysis Set (mITT)

The modified intent-to-treat (mITT) analysis set will include ITT subjects who receive at least 1 treatment of study device (Active or Sham) and will exclude patients who were randomized in error. The number of exclusions from the mITT analysis set is expected to be small and will only be based on information that was known or could have been known prior to the first treatment. Subjects in the mITT population will be analyzed in the group to which they were randomized.

2.4. Per Protocol Analysis Set (PP)

The Per Protocol (PP) analysis set will be comprised of the following:

- Subjects in the mITT analysis set
- Subjects with device failure within the 10-week follow-up period
- Subjects with deviation from the clinical investigation plan is caused by the investigational device or by problems with respect to tolerability

Subjects who will be excluded from the PP analysis set are as follows:

- Subjects with major protocol violations that would be expected to confound clinical assessment during follow-up as determined by the PIs.
- Subjects who took new pharmaceutical substances or treatments during the clinical investigation which are listed as an exclusion criteria;
- Subjects who do not meet the inclusion criteria or exclusion criteria.
- Users who have performed less than 10 sessions (300 minutes) during the first 3 weeks will be excluded from the analysis.

Subjects in the PP population will be analyzed in the group to which they were randomized.

3. Study Endpoints

3.1. Primary Effectiveness Endpoint

3.1.1. Testing of Primary Effectiveness Hypotheses

The primary effectiveness endpoint is the change in adjusted mean group difference in the HDRS-17 scores between 10 weeks and baseline between subjects randomized to the active study device relative to those subjects randomized to sham. The formal statistical hypothesis is as follows:

$$H_0: d_{\text{flow}} - d_{\text{sham}} \leq 0$$

$$H_a: d_{\text{flow}} - d_{\text{sham}} > 0$$

Where d_{flow} and d_{sham} are the adjusted mean group difference in HDRS-17 scores in subjects randomized to Active and Sham groups, respectively.

The mITT population will be the analysis population used to analyze the primary effectiveness endpoint. The endpoint will be analyzed when all randomized subjects have completed their 10-week visit, been withdrawn from the trial or passed the end of their 10-week follow-up visit window. The difference in the group mean differences will be assessed using a mixed model for repeated measures. If the p-value of the difference in the group mean differences is less than one-sided $p = 0.025$, then the endpoint declared a success.

3.1.2. Multiple Imputation Analysis

The mixed model for repeated measures (MMRM) allows for inclusion of patients with missing Week 10 values. If all subjects have at least one follow-up evaluation, the following MMRM will be utilized for the primary superiority test. If there are subjects without at least one post baseline assessment, the multiple imputation analysis summarized below will be the primary analysis method. Least squares estimated values from the MMRM will be reported for HDRS-17 at each visit and as such these estimates include implicit imputations of missing values.

3.1.3. Secondary Endpoints

There are eight secondary endpoints. **Three of the eight secondary endpoints will be secondary endpoints controlled for multiplicity, the remaining five secondary endpoints will be provided descriptively.** Endpoints #2a, 3a, and 4 will be controlled for multiplicity:

1. The mean change in points for MADRS score from baseline to Week 10.
2. Clinical Response will be analyzed as responder endpoints. The 3 Clinical Response rates are defined as
 - a. a >50% reduction from the baseline HDRS-17 at Week 10.
 - b. a >50% reduction in MADRS score at Week 10.
 - c. a >50% reduction from the baseline MADRS-s at Week 10.
3. Remission will be analyzed as a responder endpoint where remission is defined as:
 - a. HDRS-17 score ≤ 7 at Week 10
 - b. MADRS score ≤ 10 at Week 10
 - c. MADRS-s score ≤ 12 at Week 10
4. Quality of life improvement as measured by EQ-5D-3L in the active compared to sham arm at week 10 compared to baseline.

3.1.4. Discussion on FDA Requested secondary endpoints and clarification of the endpoints controlled for multiplicity

The eight (8) endpoints listed above are consistent with the original IDE submission except for some re-organization for clarity (i.e., reordering and using numbers instead of bullets). Endpoint #5 (EQ-5D) was listed as an exploratory endpoint in the IDE submission. As per FDA's study design consideration (G210328/S001), EQ-5D has been moved to the secondary endpoint section and is now a multiplicity-controlled secondary endpoint.

3.1.5. Secondary Endpoint #1: Change in MADRS score

This endpoint is described within this section but is not considered one of the three endpoints multiplicity controlled.

3.1.5.1. Endpoint Definition

Secondary endpoint #1 is the adjusted mean change in group difference in the MADRAS scores between 10 weeks and baseline between subjects randomized to the active study device relative to those subjects randomized to sham.

3.1.5.2. Hypothesis Test

The formal statistical hypothesis is as follows:

$$H_0: d_{\text{flow}} - d_{\text{sham}} \leq 0$$

$$H_a: d_{\text{flow}} - d_{\text{sham}} > 0$$

Where d_{flow} and d_{sham} are the mean change in group difference in the MADRAS scores in subjects randomized to Active and Sham groups, respectively.

3.1.5.3. Analysis Set

The mITT population will be the analysis population used to analyze secondary endpoint #1. The endpoint will be analyzed when all randomized subjects have completed their 10-week visit, been withdrawn from the trial or passed the end of their 10-week follow-up visit window. The mean change in group difference in the MADRAS scores will be assessed using a mixed model for repeated measures (MMRM). The significance level for this endpoint will be determined using the Hochberg procedure, which is described in Section 3.2.5.1. Details on the statistical model for MMRM are also described in Section 4.

3.1.5.4. Missing Data Analysis

The mixed model for repeated measures (MMRM) allows for inclusion of patients with missing Week 10 values. If all subjects have at least one follow-up evaluation, the following MMRM will be utilized for the primary superiority test. If there are subjects without at least one post baseline assessment, the multiple imputation analysis summarized below will be the primary analysis method. Least squares estimated values from the MMRM will be reported for each MADRAS score and as such these estimates include implicit imputations of missing values.

3.1.6. Secondary Endpoint #2: Clinical Response

3.1.6.1. Endpoint Definition

Secondary endpoint #2 is the clinical response between subjects randomized to the active study device relative to those subjects randomized to sham. The three Clinical Response rates are defined as a >50% reduction from the baseline HDRS-17 and MADRS and MADRS-s score at Week 10.

The clinical response per HDRS-17 is controlled for multiplicity (see below). This endpoint is listed as #2a in the list of endpoints above. This endpoint based on HDRS-17 is the endpoint referred to as "Endpoint #2" in the Adaptive Design Report. The clinical response based on MADRS and MADRS-S are listed as endpoints #2b and #2c in the list of endpoints above. Those two endpoints will be provided with descriptive statistics.

3.1.6.2. Hypothesis Test

The formal statistical hypothesis is as follows:

$$H_0: p_{\text{flow}} - p_{\text{sham}} \leq 0$$

$$H_a: p_{\text{flow}} - p_{\text{sham}} > 0$$

Where p_{flow} and p_{sham} are the clinical responses in subjects randomized to Active and Sham groups, respectively.

3.1.6.3. Analysis Set

The mITT population will be the analysis population used to analyze secondary endpoint #2. The endpoint will be analyzed when all randomized subjects have completed their 10-week visit, been withdrawn from the trial or passed the end of their 10-week follow-up visit window. The endpoint will be tested using a two-group large-sample normal approximation test of proportions. The significance level for this endpoint will be determined using the Hochberg procedure, which is described in Section 3.2.5.1.

3.1.6.4. Missing Data Analysis

Multiple Imputation will be used through SAS PROC MI with age and sex as covariates along with longitudinal responses from all other timepoints. The 20 imputations will be combined using Rubin's Rules through SAS PROC MIANALYZE.

3.1.7. Secondary Endpoint #3: Remission

3.1.7.1. Endpoint Definition

Secondary endpoint #3 is the difference in remission proportion between subjects randomized to the active study device relative to those subjects randomized to sham. The three remission rates will be analyzed as responder endpoints where remission is defined as an HDRS-17 score ≤ 7 , a MADRS score ≤ 10 , or MADRS-s score of ≤ 12 .

Remission per HDRS-17 is controlled for multiplicity (see below). This endpoint is listed as #3a in the list of endpoints above. Remission based on MADRS and MADRS-S are listed as endpoints #3b and #3c in the list of endpoints above. Those two endpoints will be provided with descriptive statistics.

3.1.7.2. Hypothesis Test

The formal statistical hypothesis is as follows:

$$H_0: p_{\text{flow}} - p_{\text{sham}} \leq 0$$

$$H_a: p_{\text{flow}} - p_{\text{sham}} > 0$$

Where p_{flow} and p_{sham} are the remission proportion in subjects randomized to Active and Sham groups, respectively.

3.1.7.3. Analysis Set

The mITT population will be the analysis population used to analyze secondary endpoint #3. The endpoint will be analyzed when all randomized subjects have completed their 10-week visit, been withdrawn from the trial or passed the end of their 10-week follow-up visit window. The endpoint will be tested using a two-group large-sample normal approximation test of proportions. The significance level for this endpoint will be determined using the Hochberg procedure, which is described in Section 3.2.5.1.

3.1.7.4. Missing Data Analysis

Multiple Imputation will be used through SAS PROC MI with age and sex as covariates along with longitudinal responses from all other timepoints. The 20 imputations will be combined using Rubin's Rules through SAS PROC MIANALYZE.

3.1.8. Secondary Endpoint #4: Quality of Life Improvement

3.1.8.1. Endpoint Definition

Secondary endpoint #5 is the quality of life improvement as measured by EQ-5D-3L in the active compared to sham arm at week 10. This is one of the three second endpoints controlled for multiplicity.

3.1.8.2. Hypothesis Test

The formal statistical hypothesis is as follows:

$$H_0: d_{\text{flow}} - d_{\text{sham}} \leq 0$$

$$H_a: d_{\text{flow}} - d_{\text{sham}} > 0$$

Where d_{flow} and d_{sham} are the mean change in group difference in the EQ-5D-3L scores in subjects randomized to Active and Sham groups, respectively.

3.1.8.3. Analysis Set

The mITT population will be the analysis population used to analyze secondary endpoint #4. The endpoint will be analyzed when all randomized subjects have completed their 10-week visit, been withdrawn from the trial or passed the end of their 10-week follow-up visit window. The mean change in group difference in the EQ-5D-3L scores will be assessed using a mixed model for repeated measures (MMRM). The significance level for this endpoint will be determined using the Hochberg procedure, which is described in section 3.2.5.1.

3.1.8.4. Missing Data Analysis

The mixed model for repeated measures (MMRM) allows for inclusion of patients with missing Week 10 values. If all subjects have at least one follow-up evaluation, the following MMRM will be utilized for the primary superiority test. If there are subjects without at least one post baseline assessment, the multiple imputation analysis summarized below will be the primary analysis method. Least squares estimated values from the MMRM will be reported for each EQ-5D-3L score and as such these estimates include implicit imputations of missing values.

3.1.8.5. Additional analyses

EQ-5D will be provided using descriptive statistics including an analysis based on health states.

3.1.9. Multiplicity Control for Secondary Endpoints

Should the primary endpoint be achieved, the next endpoint that will be tested is the clinical response endpoint based on HDRS-17 50%. Should that endpoint be achieved, the remaining endpoints in this document will be tested using the methodology in this section.

The Sponsor understands that multiplicity correction and Type I error in general are not the most relevant or complete statistical quantities of interest in all statistical perspectives. As summarized by Vanderbilt University and frequent FDA Statistical Advisor, Dr. Frank Harrell.

Statisticians have convinced regulators that long-run operating characteristics of a testing procedure should rule the day, e.g., if we did 1000 clinical trials where efficacy was always zero, we want no more than 50 of these trials to be judged as ""positive"." Never mind that this type I error operating characteristic does not refer to making a correct judgment for the clinical trial at hand. Still, there is a belief that type I error is the probability of 'regulator's regret (a false positive), i.e., that the treatment is not effective when the data indicate it is. In fact, clinical trialists have been sold a bill of goods by statisticians. No probability derived from an assumption that the treatment has zero effect can provide evidence about that effect. Nor does it measure the chance of the error actually in question. All probabilities are conditional on something, and to be useful they must condition on the right thing. This usually means that what is conditioned upon must be knowable.

The probability of 'regulator's regret is the probability that a treatment 'doesn't work given the data. So the probability we really seek is the probability that the treatment has no effect or that it has a backwards effect. This is precisely one minus the Bayesian posterior probability of efficacy.¹

The following strategy is specified to achieve two primary goals:

- Provide FDA pre-specified statistical quantities to enable a straightforward regulatory review of the PMA data.
- Provide comprehensive statistical summaries of the PMA data which exist beyond multiplicity adjusted p-values.

Multiplicity for the three secondary endpoints will be controlled using Hochberg correction implemented through the following algorithm:

3.1.9.1. Hochberg Algorithm

A Hochberg² "step-up" approach will be utilized to control the overall type I error at the desired 5% level across the three secondary efficacy analyses. To implement the Hochberg method, the p-values for the set of multiple null hypotheses are ordered from largest to smallest, and each p-value is compared to a sequentially decreasing alpha-level to determine whether the null hypothesis (and, potentially, subsequent hypotheses) should be rejected. Symbolically, for the set of p-values $\{p_1, \dots, p_k\}$ ordered from largest to smallest and testing the corresponding set of null hypotheses $\{H_{o1}, \dots, H_{ok}\}$, the Hochberg procedure is implemented as:

Step 1: Evaluate whether $p_1 < \alpha$. If yes, reject H_{o1} and all subsequent null hypotheses $\{H_{o2}, \dots, H_{ok}\}$. Else, do not reject H_{o1} and go to Step 2.

Step 2: Evaluate whether $p_2 < \alpha/2$. If yes, reject H_{o2} and all subsequent null hypotheses $\{H_{o3}, \dots, H_{ok}\}$. Else, do not reject H_{o2} and go to Step 3.

[...]

Step k: Evaluate whether $p_k < \alpha/k$. If yes, reject H_{ok} . Else, none of the null hypotheses $\{H_{o1}, \dots, H_{ok}\}$ are rejected and stop.

¹ Dr. Frank Harrell, p-values and Type I Errors are Not the Probabilities We Need, Last updated on 2020-09-15 and Accessed 2021-11-18, <https://www.fharrell.com/post/pvalprobs/>

² Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 1988; 75:800–802

As only the smallest p-value is compared to the traditional Bonferroni-corrected α -level, the Hochberg method is more statistically powerful for controlling Type I error in the context of testing multiple null hypotheses. The Hochberg approach to Type I error control has been shown to be uniformly more powerful than the Bonferroni and Holm procedures and have non-restrictive distributional assumptions that are often met. Therefore, it is always preferred when hypotheses are not ordered.³ Should the assumption that the endpoints are not negatively correlated not be met, a bootstrapping approach for multiplicity will be provided in addition to the Hochberg correction. The adjusted p-values from the Hochberg procedure will be considered the primary study results.

The specified multiplicity algorithm maintains overall Type I error at the desired α -level across a set of null hypotheses while also improving statistical power over other approaches, and importantly, allows for regulatory review of the secondary endpoints.

3.1.10. Exploratory Endpoints

The following list of secondary endpoints will be analyzed using the mITT population and will be summarized based on the type of data: Continuous (CONT), Binary (BIN) or Correlation (CORR).

In general, for all endpoints, the values over time (at each timepoint) and the change from baseline (at each timepoint) will be presented. The following list contains the specific endpoints that will be provided using descriptive tabular and graphical methods:

- Montgomery–Åsberg Depression Rating Scale (MADRS) response rate at all follow-up time points (CONT)
- Montgomery–Åsberg Depression Rating Scale - Self-report (MADRS-s) scores of the two arms (response and remission rates) at all time points (CONT)
- Hamilton Anxiety Rating Scale (HAM-A) at all time points (CONT)
- Young Mania Rating Scale (YMRS) at all time points (CONT)
- tDCS Adverse Events Questionnaire (AEQ) at 10 and 20 weeks (CONT)
- HDRS-17 (CONT), MADRS (CONT), MADRS-s mean decrease (CONT), remission (BIN), and response rate (BIN) at 6 weeks comparison between groups.
- HDRS-17 (CONT), MADRS (CONT), MADRS-s (CONT) at 20 weeks (after open-label)
- Correlation between MADRS and MADRS-s. (CORR)
- Correlation between the percent of sessions using the FLOW FL-100 and HDRS-17/MADRS decrease in the active group at 10 weeks. (CORR)
- The within-patient clinically meaningful improvement as defined as at least -3 points on the HDRS-17 scale. The percentage of subjects for each arm that reaches -3 points or more improvement.
- Cumulative Proportion of Responders Analysis- Cumulative Distribution Functions for the Change in HDRS-17 Score from Baseline to Primary end-point by arm will be provided as per FDA Guidance on reporting patient reported outcomes⁴

³ Dmitrienko, A., & D'Agostino, R. (2013). Traditional multiplicity adjustment methods in clinical trials. *Statistics in Medicine*, 32(29), 5172–5218.

⁴ PATIENT-FOCUSED DRUG DEVELOPMENT GUIDANCE PUBLIC WORKSHOP Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making; FDA Workshop- December 6, 2019

For continuous variables (CONT), results will be summarized with the numbers of observations, means, standard deviation, median, minimums and maximums. Differences between the two groups, where specified, will be summarized with the differences of the two means, and 95% confidence intervals for the difference between the means. These calculations will be done under the assumption that data for the two groups are independent and approximately normally distributed. The confidence interval for the difference of two means will be calculated under the assumption of unequal variances. If the asymptotic assumptions fail, then nonparametric summary statistics (medians, 25th and 75th percentile) may be displayed as an alternative.

For binary variables (BIN), results will be summarized with patient counts, percentages/rates. Differences between the two groups, where specified, will be summarized with the difference in percent and the Newcombe score 95% confidence interval for the difference of two percentages.

For correlation analysis, Spearman correlation will be used to assess the association between two continuous variables. The correlation coefficient and the 95% confidence intervals will be reported.

4. Details of the MMRM

Assessment of change from baseline in HDRS-17 will be evaluated at Week 10. As discussed above, a mixed model for repeated measures (MMRM) allows for inclusion of patients with missing Week 10 values. For example, patients dropping out after Week 2 or after Week 8 will be included in the primary effectiveness analysis.

The MMRM used to test the primary effectiveness hypothesis will include the baseline value for HDRS-17, a treatment group factor (Investigational or Sham), a categorical factor for visit, and treatment group by visit. The MMRM approach is a direct likelihood approach⁵ requiring specialized statistical software. For this study, all MMRM parameters will be estimated using SAS PROC MIXED (SAS Institute, Cary NC Version 9.4 or higher). The MMRM model is notable for its ability to include all available outcome data from all subjects including drop-outs and does not require their exclusion as in complete case analysis or arbitrary assignment of some value as in last observation carried forward (LOCF). The MMRM generally includes a factor for group by visit interaction to allow treatment group differences in mean changes to vary over time. Inclusion of outcome data from earlier time points informs the implicit imputation of values missing at later time points (e.g., at Week 10) through the covariance matrix used to model the random effects.

In a matrix equation, the MMRM can be expressed as: $Y_i = X_i\beta + Z_iu + e_i$, where β is the vector of fixed-effect regression parameters (for the overall mean change, the treatment effect θ , a vector of post-baseline time effects τ , a vector of treatment-by-time interaction effects η , and a vector of covariate effects φ that includes baseline HDRS-17 and optionally, other a priori selected covariates. X is a design matrix for the fixed effects, Z is a design matrix used to account for other random effects u , if any were included. Key assumptions are about e , the random error vector. It is assumed that the expected values are zero, i.e., $E(e) = 0$. An unstructured covariance is assumed requiring estimation of variances at each visit and all pairwise covariances, i.e., $Var(e) = \sigma_e^2 V_{unstructured}$.

⁵ Verbeke G. and Molenberghs G. Linear Mixed Models for Longitudinal Data. New York: Springer 2000.

The parameter estimates from this model will be used to construct an estimate of the treatment group difference in mean improvements at Week 10 between Investigational and sham. The advantages of MMRM have been outlined in Siddiqui, Hung, and O'Neill (2009).⁶

“According to Rubin (1976) and Little and Rubin (2002), data are considered missing completely at random (MCAR) if, conditional upon the independent variables in the analytic model, missingness is independent of both unobserved and observed outcomes of the variable being analyzed; data are missing at random (MAR) if, conditional upon the independent variables in the analytic model, missingness depends on the observed data of the variable being analyzed but is independent of the unobserved outcomes of the variable being analyzed; data are missing not at random (MNAR) if, conditional upon the independent variables in the analytic model, the missingness depends on the unobserved outcomes of the variable being analyzed.”

“A missingness mechanism is called “ignorable” if a likelihood-based analysis [such as MMRM] provides valid inferences of the model parameters even when the missingness mechanism is ignored; otherwise, it is called “nonignorable” missingness mechanism. Laird (1988) shows that MCAR and MAR are ignorable missingness, and likelihood-based analyses that ignore the missing data mechanism remain valid. For nonignorable missing data, however, likelihood-based analyses that ignore the missing data mechanism potentially produce biased results.”

MMRM produces unbiased parameter estimates under MAR, while complete case analysis only produces unbiased estimates under the much more restrictive assumption of MCAR. That is, MMRM produces unbiased parameter estimates if conditional on the baseline HDRS-17 and the change scores up to the visit at which the patient drops out (and other baseline covariates, if any) the probability of missing is independent of the missing value itself. In contrast, MCAR requires that the probability of missing is unconditionally independent of the missing value itself, an assumption that is likely violated in practice. The MAR implicit imputation of missing values can be improved by including additional baseline or site effects that are associated with the probability of having a missing value. In contrast, “the LOCF approach is simple, but it makes two strong assumptions that: (i) missing data due to dropouts follow MCAR and (ii) the responses following a patient dropping out remain constant at the last observed value prior to drop out. Both of the assumptions are often unrealistic in clinical trials.”

The handling of missing data is described in the next section.

5. Missing data

5.1. Procedures For Minimizing The Occurrence Of Missing Data

The preferred and often only fully satisfactory approach to addressing missing data is maximizing prevention. This trial defines the following proactive approaches to be implemented to minimize the number of patients who are not assessed at Week 10.

Among these approaches are the following:

⁶ Siddiqui OS, Hung HMJ, and O'Neill R. MMRM vs LOCF: A comprehensive comparison based on simulation study and 25 NDA data sets. Journal of Biopharmaceutical Statistics, 2009, 19: 227-246.

- The study protocol properly distinguishes between reasons for nonadherence (that is, for not receiving randomized therapy and hence for being "off study treatment") versus non-retention (that is, for not obtaining outcome information and hence for being "off study").
- The study protocol recognizes valid reasons for subject discontinuation of follow-up: subject withdrawal of consent, death, achievement of all required efficacy and safety endpoint information, or loss to follow-up.
- The term 'withdrawal of consent' is used only when the patient no longer wishes to participate in the trial and no longer authorizes the investigators to make efforts to continue to obtain their outcome data.
- Patients are educated during the informed consent process about the continued scientific relevance of their data even if they discontinue treatment and the deleterious effect that missing data has on trial integrity and credibility.
- All patients will be followed until death or trial completion, even after discontinuing study treatment or initiating other treatment.
- It is recognized that protocol-specified increases in sample size to adjust for missing data address the influence of missingness on the precision of estimates, but not its influence on bias induced in those estimates.
- The protocol specifies performance standards to be met to achieve high quality of trial conduct, including specification of targeted levels of data capture.
- Creative and effective procedures are being implemented during enrollment and follow-up to enhance achieving pre-specified targeted retention levels.
- One of these procedures will be the engagement of investigators committed to following all patients until their death or capturing primary endpoint data, even if patients have discontinued randomized treatment or initiated other interventions.
- An oversight process by the sponsor is in place during trial conduct to ensure the achievement of performance standards, including targeted levels of data capture.

5.2 Data Collection For Subjects Who Discontinue From Randomized Treatment Or Discontinue From The Study

5.2.1. DISCONTINUATION DEFINITIONS AS PROVIDED IN THE PROTOCOL

In every instance where a subject does not complete the study, the Investigator will document the primary reason for discontinuation in the subject's records. In the case of early discontinuation of randomized treatment, the investigator will make all efforts to complete the final study assessments (unless contraindicated) before the subject's exit from the study.

5.2.2. DISCONTINUATION FROM STUDY BY SUBJECT

Subjects have the right to withdraw from the clinical investigation at any time and for any reason without prejudice to their future medical care by the investigation team or investigational site.

The Principal Investigator will ask the reason for a subject's withdrawal from the study and will record all information regarding the subject's discontinuation. If a subject expresses a desire to withdraw their consent for the study, the site should attempt to obtain written documentation for their study records.

Regardless of the reason for withdrawal, data available for the subject at the time of withdrawal, including the reason for withdrawal, will be entered on the Case Report Forms. A subject that has been withdrawn from the study after treatment will not be replaced.

5.2.3. DISCONTINUATION FROM RANDOMIZED TREATMENT BY INVESTIGATOR

A Subject may be withdrawn from randomized treatment by the Investigator for the following reasons:

- Any unanticipated adverse event that, in the Principal Investigator's opinion, is related to the treatment and will endanger the well-being of the subject if treatment is continued.
- Development of any intercurrent illness(es), infection or condition(s) that might interfere with the continuation of treatment.
- Non-compliance with the clinical investigation procedures deemed by the Principal Investigator to be sufficient to cause discontinuation of treatment.
- The investigator determines that continuing the randomized study treatment is not in the subject's best interest.
- Any other problem deemed by the Investigator to be sufficient to cause discontinuation from study treatment.

For subjects that discontinued from study treatment by the Investigator, all reasonable efforts should be made to ensure the assessment of the primary and key secondary endpoints. PROs will be collected in the event of an unscheduled visit; these will be included in the sensitivity analysis defined below.

5.2.4. TEMPORARY DISCONTINUATION BY SUBJECT

Not applicable in this study.

5.2.5. DISCONTINUATION OF INADVERTENTLY ENROLLED SUBJECTS

If the Sponsor or Investigator identifies a subject who did not meet enrollment criteria and was inadvertently treated, that subject will be followed. The subject will be asked to return for all protocol-specified visits and complete all assessments (unless contraindicated). Any eligibility violation would be captured as a protocol deviation in the database.

5.2.6. LOST TO FOLLOW-UP (LTFU)

For any nonresponsive subjects, site staff will perform and document a minimum of three attempts to contact them via phone and one attempt to reach them via certified mail. The staff will document the date and type of attempted communication.

5.2.7. USE OF DATA FROM EARLY DISCONTINUATION CASES

Study data collected previously for subjects who withdraw from the study or are LTFU will be included in the data analysis and Clinical Study Report.

5.3. SPECIFICATION OF THE ESTIMAND

Primary and key secondary efficacy data (e.g., HDRS-17 scores) will be captured using approaches that are creative and proper.

One important consideration is the handling of Discontinuation of randomized treatment due to AEs, lack of treatment effectiveness, or other reasons that may occur. These events have the potential to influence the study results. This section details how these will be handled and mitigated for this trial.

Such events are called "intercurrent events" in the FDA Draft Guidance E9(R1). "[I]ntercurrent events such as discontinuation or switching of treatment, or use of rescue medication, may in some circumstances render the later measurements of the variable irrelevant or difficult to interpret even when it can be collected." To provide specificity about the estimand, it is necessary to specify how data from subjects with intercurrent events will be handled.

E9(R1) discusses several approaches to handling data after an intercurrent event. A preferred approach to handling intercurrent events is to adopt the "Treatment Policy" estimand. This approach requires data collection after the intercurrent event. This approach enables preservation of the integrity of randomization and obtains a direct assessment of the clinically relevant comparison between the intention to deliver the experimental intervention vs. the intention to deliver the control intervention in the context of real-world supportive care.

Supportive analyses will be obtained using alternative approaches to handling intercurrent events. One of these is to incorporate the intercurrent event as a failure mode in a composite endpoint regarding clinically meaningful improvements, especially if this endpoint is assessed in a responder or time-to-event analysis.

Another approach to handling intercurrent events is the hypothetical approach to "envisage in which the intercurrent event would not occur." For example, the "treatment effect of interest might concern the outcome if the subsequent active treatment had not been administered." This is specified as the secondary estimand, see below.

This section describes the influence of the estimand of interest on the capture of primary outcome data. For this study, the estimand of interest is the improvement in HDRS-17 in subjects at Week 10 relative to baseline compared to Control.

5.3.1. PRIMARY ESTIMAND

The primary estimand of interest is the relative difference between investigational subjects and those who received the control treatment. All recorded scores in the database will be used regardless of additional therapy that may be sought during the course of the trial. All subjects who are missing Week 10 HDRS-17 will be included in the primary analysis through Multiple Imputation that attempts (defined below) to replicate what would have been captured had the patient been assessed on that date.

As described below, the subjects will be differentially imputed in a manner that considers what is known about the reason for the discontinuation from follow-up and what is known about the subject's outcome performance through the time of their discontinuation. The CRF will collect the specific reasons for discontinuing therapy to identify these subjects.

A sensitivity analysis will be provided for this estimand that uses LOCF prior to the discontinuation.

5.4. Handling Of Missing Data

Subjects lost to follow-up may produce bias in this study. Reasons for discontinuation from treatment and from follow-up will be evaluated and characterized by tabulating the reasons for missingness within the randomized treatment group overall and according to the study week. Any subject who is missing their primary Week 10 HDRS-17 evaluation will be included in the primary study assessment through Multiple Imputation (MI).

5.4.1. HANDLING DATA FOLLOWING SUBJECT WITHDRAWAL DUE TO LACK OF TREATMENT EFFECTIVENESS

Subjects who indicate, "Withdrawn due to lack of treatment effectiveness" and are missing the Week 10 evaluations will be imputed by replacing every subsequent follow-up visit with the baseline score. This methodology is understood to be a "Jump to Baseline." Provided that all treatments will be provided as medically necessary, there are expected to be very few subjects like this. Instead, subjects are expected to be recorded as having taken additional treatments but remain in the study.

5.4.2. HANDLING OF SUBJECTS REASONABLY MISSING AT RANDOM

Subjects who can be reasonably understood to be missing at random, that is, the reasons for missingness for subjects who appear pseudo-independent from their clinical status, for example, moved to a different state during the trial, will be imputed within their randomized treatment group.

5.4.3. HANDLING OF SUBJECTS WITH MISSINGNESS FOR ALL OTHER REASONS

This category likely would include most of the subjects with missing data for the primary endpoint and secondary endpoints at Week 10. With this approach, for each patient and each endpoint, the patients' actual trajectory for that endpoint will be used over the interval that it would be available, ideally from randomization through the time of their discontinuation from the trial. The remaining trajectory for that patient, for that endpoint, would be imputed using the trajectory for control patients over the window from their discontinuation from the trial (or from the last calendar time when it would be known for that patient) until the time of endpoint assessment, (i.e., Week 10). That control trajectory would be assumed to be piecewise linear between follow-up timepoints.

Given the expected high rate of follow-up achieved by the procedures for reducing missing data outlined above and given that the primary endpoint is evaluated at Week 10, very few subjects may be missing and eligible for this treatment of missing data. Pre-specified sensitivity analysis will provide additional insight into this handling of missing data; see below.

5.4.4. MULTIPLE IMPUTATION - GENERAL

If any patients are missing their primary Week 10 evaluation (or other earlier timepoint), multiple imputation (MI) (Rubin and Schenker 1991) will be used to create completed datasets appropriate for regulatory evaluation of the primary endpoint.

5.4.5. DEVELOPMENT OF MI MODEL

One of the key assumptions for a valid application of Multiple Imputation is that the probability of missingness is independent of the missing value itself, unconditionally or conditionally. If unconditionally, this situation is referred to as Missing Completely at Random or MCAR (Rubin 1976). If independence is conditional on observed data, the situation is referred to as Missing at Random or MAR. It is widely believed that MCAR is not a reasonable assumption, but MAR might be a reasonable assumption in many instances. If either MCAR or MAR holds, the missing mechanism is said to be ignorable (Little and Rubin 2002). Many conventional missing data strategies, including Mixed Models for Repeated Measures (MMRM) and Multiple Imputation (MI) require an ignorable missing value mechanism. When this assumption does not hold, the situation is referred to as Missing Not at Random, or MNAR. When MNAR holds, the application of MI is biased.

The situation is nearly as MNAR as any situation could get regarding subjects who discontinue treatment due to lack of efficacy. For this reason, subjects in both groups who discontinue due to lack of treatment effectiveness will be imputed using the Jump to Baseline methodology. For the subjects who are reasonably understood to be missing at random will be included in the MI model within their randomized treatment group. For the remaining subjects in both groups, missing values will be imputed from a model including only Control subjects. The covariates and model specifications are detailed in the following two sections.

5.4.6. COVARIATES IN THE MI MODEL

The covariates chosen for the MI model are:

- Sex
- Age
- Antidepressants use at baseline
- In Psychotherapy at baseline
- Baseline HDRS-17
- Longitudinal HDRS-17 scores

Should convergence issues arise during the modeling procedure, sex will be removed from the model, and only the continuous variables used.

5.4.7. CHOICE OF MI MODEL

The selected MI model will impute continuous primary endpoints over time. By imputing continuous HDRS-17 values over time, the secondary endpoints evaluating values over time and change from baseline may be analyzed using the same imputed dataset. This is expected to produce consistency and

coherence between the various inferential and descriptive statistics presented in the Clinical Study Report.

It is assumed that missing data may be non-monotonic. When missing data is non-monotonic, it is necessary to specify an appropriate modeling methodology to handle it. Further, because the specified covariates are both continuous and categorical, the choice of models additionally limited. For these reasons, the fully conditional specification (FCS) approach, as implemented in SAS PROC MI will be used to obtain 20 multiply imputed completed data sets. The FCS requires the specification of the predictive models for each variable.

Fully Conditional Specification (FCS) is an example of a "chained equations" approach. It is also known as "regression switching" approach. It is commonly referred to as MICE or Multivariate Imputation By Chained Equations. The FCS approach differs from other MI approaches in that it does not explicitly involve the joint distribution of the variables but focuses on the marginal distributions. It is like MCMC and Gibbs sampling in this regard, except that typically many fewer iterations are needed to determine the required distributions. The following describes the approach:

"The fully conditional approach to imputation is a more flexible method that does not rely on the assumption of multivariate normality (9, 11). Conditional distributions (regression models) are specified for each variable with missing values, conditional on all of the other variables in the imputation model. Imputations are generated by estimating each conditional distribution in turn, using observed cases for the variable being considered and imputed values for the other variables at that iteration and imputing missing values (again allowing for uncertainty in model parameters). The approach is appealing, since it does not restrict the conditional distributions to being normal, so that univariate regression models can be tailored appropriately—for example, using logistic regression for binary variables and ordered logistic regression for ordinal variables. However, it is possible for some of the conditional distributions to be incompatible with each other, potentially leading to unsound imputations (14). The FCS approach is also available in a number of statistical packages, including Stata (13), SAS (11), and R (21), and is increasingly appearing in practice (22, 23)." (Lee and Carlin 2010) See also Buuren SV. Flexible Imputation of Missing Data. New York: CRC Press 2012.

5.4.8. BASELINE COVARIATES

Subjects missing a baseline covariate (e.g., sex, age, or baseline HDRS-17) will have their baseline covariates within the FCS model.

5.4.9. LONGITUDINAL HDRS-17

An aim of this missing data modeling is to preserve the temporal nature of the longitudinal HDRS-17 values. To effectuate this:

1. For the model that imputes subjects presumed missing at random, HDRS-17 values at each timepoint will be included in the MI model. For each timepoint, the HDRS-17 values at all other timepoints, along with the baseline covariates will predict the missing values. This imputation is geared to incorporate subject-level clinical trajectories, which, as stated above, can theoretically provide more information and aid in the imputation.
2. For the model that impute the remaining subjects within the control group, the timepoint immediately prior to each visit along with covariates will predicted the missing value. That is,
 - Week 10 ~ Week 7 + covar_1 + covar_k
 - ...
 - Week 1 ~ Baseline + covar_1 + covar_k

5.5. Sensitivity Analysis

5.5.1. ALTERNATIVE MODELING APPROACHES

Sensitivity analyses for handling missing data are specified for the Primary Endpoint, which will produce specific and unique estimates of success. The sensitivity analysis will calculate Week 10 change scores after LOCF is applied prior to their missing visit. Subjects who discontinue will have an additional PRO evaluation recorded during an unscheduled visit. When this value exists, it will be subject to this LOCF procedure and presumably add additional information.

6. Interim Analysis

An interim analysis will be performed when 90 subjects have week 10 data, which will include both a futility assessment and sample size re-estimation. The futility assessment will be based on stochastic curtailing approach by Lachin et al (2005). The sample size re-estimation will be based on a Promising Zone methodology by Mehta and Pocock (2011) with a Fuzzy design approach introduced by Keenan and Maislin (2014). Separate sample size assessments will be performed for evaluating the secondary endpoint of clinical response. Adjustments to the trial for the secondary endpoint will only be performed if the primary endpoint meets criteria as dictated by the promising zone criteria. Sample size assessments for the primary and secondary are provided with extensive details including simulation analyses and methods for controlling operational bias are provided in the adaptive design report.

At a high level, the Interim analysis may modify the trial in two ways for the primary endpoint:

- Declare the trial futile and stop enrollment.
- Specify a number of subjects between 100 and 270 that is needed for powering the trial.

7. Sample Size Assumptions and Calculation

The sample size calculation is based on data from Brunoni et al (2017). The calculation is based on a two-sample t-test for mean difference with the following assumptions:

- Group difference in means = 3.2
- Standard deviation of Investigational and Sham groups = 7.1 and 7.9, respectively
- One-sided type I error = 0.025
- 80% power

The resulting sample size is 176 subjects. To obtain an increase in statistical power (87.6%), 216 subjects were chosen. Accounting for 20% attrition, the overall sample size will be 270 subjects.

Based on the results of the interim analysis, the trial size may be adjusted in the following ways:

- If the trial is declared futile, enrollment will stop
- If less than 270 subjects are needed for statistically powering the study, the trial will only enroll up to the number of subjects specified by the Promising Zone technology (see the section on the Interim Analysis and the Adaptive Design Report)

8. Additional Analyses

8.1. Evaluation of Blinding

An evaluation of blinding will be performed. Prior to subjects unblinding through the system they will respond to the question "Which treatment arm do you think you are in?" (answers: sham/active), followed by the question "How certain are you of this?", with the options 1 to 5 where 5 is very certain and 1 is very uncertain. The counts and percentages of the subjects responding correctly will be tabulated and reported.

8.2. Device Survival

Kaplan-Meier survival analysis will be used to characterize device failure time. Device failure in this study is defined as the device not operating as intended without unexpected failure. Unexpected failure is defined as a failure that is not related to device functionality. For example, the device accidentally falling out of a window would be categorized as an unexpected device failure. In contrast, the device failing to turn on would be considered an expected device failure. As the device is not expected to have any unexpected failures, unexpected failures will only be reported if they occur. Lifetables will be tabulated indicating the number of "failures" and the number of "at-risk" subjects over time for all failures, unexpected failures and expected failures. If there are more than one of each expected and unexpected failures, a log-rank test will be performed to compare the device failure time distribution.

8.3. Poolability Analysis

Poolability will be assessed through stratification of the primary endpoint by Country (OUS and US). Within country point estimates will be calculated along with nominal descriptive 95% Confidence Intervals. The focus of this analysis will be on the similarity of the country specific point estimates. Differences will be described in the Clinical Study Report.

Statistical evaluation of the primary endpoint will be evaluated by adding site by treatment group term to the primary MMRM model. A p-value of ≤ 0.10 will be considered significant for this poolability assessment. Should a significant site x treatment group interaction be present, the patient factors of age, gender, and baseline scores will be assessed to understand if differences in case-mix factors explain differences observed between sites.

8.4. Safety Analyses

Assessment of the safety of the investigational device will include an evaluation of the incidence and severity of complications and adverse reactions associated with the treatment.

Adverse event rates will be summarized by type of AE and for specific AEs in two ways:

1. per subject incidence of specific AEs and classes of AEs and
2. by event, summarizing event counts by visit interval over time and in accordance with FDA Guidance (CDRH 2004)⁷.

Device and procedure-related events will be summarized by severity and relatedness. Events listings that include details such as relatedness, severity, onset, and resolution status will be provided for all events and relevant subsets of events such as serious events and related events. Primary safety comparisons will be

⁷ CDRH. Guidance for Industry and FDA Staff: Clinical Data Presentations for Orthopedic Device Applications, December 2, 2004.

performed using the AT analysis set. The summary tables will show the adverse events (in coded terms), the total number of events, and the number and the percentage of subjects affected in each treatment group ("subject wise evaluation") and stratified by relation to device and severity. Data of dropouts will be presented separately, and possible bias will be discussed in the final report.

All AEs will be summarized in listings for the AT analysis set. The following AE listings will be constructed:

- AE Listing 1 All AEs Sorted by Group and Event Type
- AE Listing 2 All AEs Sorted by Group and Subject
- AE Listing 3 Serious AEs
- AE Listing 4 Severe AEs
- AE Listing 5 Device Related AEs
- AE Listing 6 Serious Device Related AEs
- AE Listing 7 Severe Device Related AEs
- AE Listing 8 Other AEs
- AE Listing 9 AEs among subjects discontinued early
- AE Listing 10 AEs among subjects who died