**A RANDOMIZED, MULTICENTRE, DOUBLE-BLIND, PARALLEL, SHAM-CONTROLLED STUDY OF THE GAMMACORE®-R, A NON-INVASIVE NEUROSTIMULATOR DEVICE, FOR THE PREVENTION OF EPISODIC MIGRAINE**

**STATISTICAL ANALYSIS PLAN**

| CIP Number | GM-11 |
|---|---|
| CIP Version (Date) | 090916 GM-11 Migraine Protocol V 2 CLEAN |
| SAP Version (Date) | Version 1.0 (05 September 2017) |

**Authors:**

| Title | Name | Signature | Date |
|---|---|---|---|
| Principal Medical Research Biostatistician, NAMSA | Candace K. McClure, PhD | | |
| Senior Principal Medical Research Biostatistician, NAMSA | Lisa Thackeray, MS | | |

**Approval:**

| Title | Name | Signature | Date |
|---|---|---|---|
| Vice President, Scientific, Medical and Governmental Affairs, Electrocore, LLC | Eric J. Liebler | | |

# TABLE OF CONTENTS

| Abbreviation/Term | Definition |
| --- | --- |
| AE | Adverse event |
| ADE | Adverse device effect |
| ANCOVA | Analysis of covariance |
| ATC | Anatomic-Therapeutic-Chemical |
| BI | Blinding index |
| CI | Confidence interval |
| eCRF | Electronic case report form |
| HIT-6 | Headache Impact Test-6 |
| ITT | Intention to treat |
| IWRS | Interactive Web Response system |
| MCMC | Monte Carlo Markov Chain |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MIDAS | Migraine Disability Assessment |
| MNAR | Missing not at random |
| nVNS | Non-invasive vagus nerve stimulation |
| OL | Open-label |
| PMM | Pattern mixture model |
| PPS | Per protocol set |
| SADE | Serious adverse device effect |
| SAE | Serious adverse event |
| SAP | Statistical analysis plan |
| SAS | Statistical analysis system |
| SD | Standard deviation |
| SOC | System organ class |
| VAS | Visual analog scale |
| WHO-DDE | World Health Organization Data Dictionary Extended |

## 1. INTRODUCTION

This Statistical Analysis Plan (SAP) outlines the statistical methods for the display, summary and analysis of data collected within the scope of protocol GM-11, dated 2015-10-13. The purpose of this plan is to provide general guidelines from which the analysis will proceed.

The SAP should be read in conjunction with the study protocol and the electronic Case Report Forms (eCRFs). This version of the SAP has been developed using the final version of the protocol mentioned above and the final version of the annotated eCRFs dated 22-APR-2015.

electroCore® has designed a non-invasive vagus nerve stimulation (nVNS) device called gammaCore-R®. gammaCore-R® is a hand-held, battery-powered unit that produces a proprietary electrical waveform in the vicinity of the vagus nerve in the neck. Each dose is relatively brief (120 seconds) and the user maintains control over the stimulation intensity.

The post-marketing clinical follow-up study will collect further clinical data related to the extended safety and efficacy of nVNS with gammaCore-R® for thepreventive treatment of episodic migraine.

## 2. STUDY OBJECTIVES

### 2.1 Primary Objective

The primary objective is the difference between the gammaCore®-R and the sham treatment groups in mean reduction in number of <u>migraine days</u> during the last four weeks in the twelve-week randomized period compared with the four week run-in period.

### 2.2 Secondary Objectives

The secondary objectives are:

- To evaluate the rate of responders for the gammaCore®-R group compared to the sham group. A responder is defined as recording at least 50% reduction in <u>migraine days</u> during the last four weeks in the twelve-week randomization period compared to the four week run-in period.

- To evaluate the difference between the gammaCore®-R and sham treatment groups in mean reduction in number of <u>headache days</u> during the last four weeks in the twelve-week randomized period compared to the four week run-in period.

- To evaluate the difference between the gammaCore®-R and the sham treatment groups in the mean reduction in acute headache medications taken during the last four weeks in the twelve-week randomized period compared to the four week run-in period.

- Compare improvement in headache disability using Headache Impact Test-6 (HIT-6)

- Compare improvement in Migraine Disability Assessment (MIDAS)

- Compare Quality of Life EQ-5D-5L

- Reduction of number of headache/migraine days in the open label period compared to baseline run-in period and the randomized period

- Adverse events

## 2.3 Other Objectives

- Blinding questions

- Subject satisfaction question; and

- Ease of use

## 3. STUDY DESIGN

### 3.1 Study Design

This is a prospective, randomized, sham-controlled, multi-center investigation designed for comparison of two parallel groups, gammaCore®-R (active treatment) and a sham (inactive) treatment. The study period will begin with a four week run-in period, during which there is no investigational treatment. The purpose of the run-in period will be to establish the baseline for comparisons. The run-in period will be followed by a 12 week randomized period when the subjects will be randomized (1:1) to either active treatment or sham (inactive) treatment. Figure 1 graphically displays the study design.

The randomized period will be followed by a 24 week open label period, where the subjects in the sham treatment group will switch treatment assignment and receive a gammaCore®-R device and the gammaCore®-R group will continue to receive an active treatment (See Protocol Section 7.1 for details).
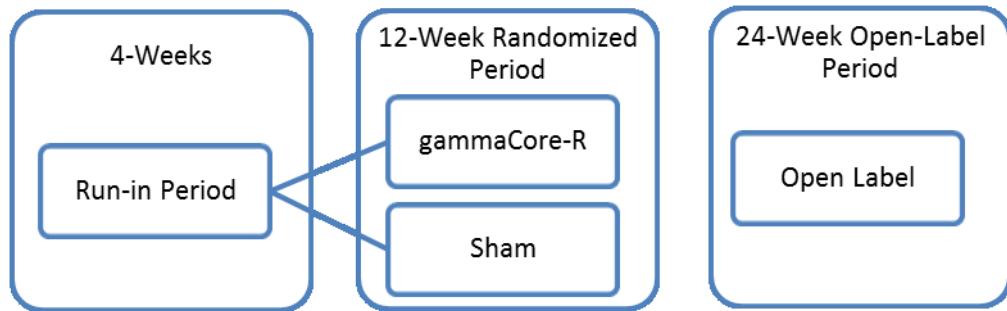


**Figure 1: Diagram of Study Design**

### 3.2 Randomization

Subjects will be randomized to either the active treatment arm or sham control arm (allocation 1:1) under a randomized variable block design, stratified by study site.

Subjects will be randomized using Merge eClinical OS Interactive Web Response system (IWRS) via computer key board data entry for a web based interface.

The site personnel (the investigator or his designee) will enter the required study and subject information and will then be provided with a subject randomization number and device serial number. Once subject numbers, serial numbers, and randomization numbers have been assigned, they cannot be reassigned.

The IWRS will provide the site personnel confirmation of the randomization number and device serial number assigned to each subject.

A sponsor designee will also have a copy of the randomization schemes for each study site and will be responsible for providing the un-blinded trainer with the randomization schemes.

## 3.3    Hypothesis Testing

Let $\mu_{Active}$ be the mean change (from the 4-week run-in period) in the number of migraine days for the active treatment group and $\mu_{Sham}$ be the mean change in the number of migraine days for the sham treatment group.  Then the effectiveness of the gammaCore®-R device can be measured by testing the difference in mean changes, $\mu_{Active} - \mu_{Sham}$. The null and alternative hypotheses of interest are:

$$H_{null}: \mu_{Active} - \mu_{Sham} = 0 \text{ versus } H_{alt}: \mu_{Active} - \mu_{sham} \neq 0$$

## 3.4    Interim Analysis

The analysis of the primary and secondary efficacy endpoints will occur after all subjects have completed the randomized/controlled period and their data is deemed clean; and, before the formal completion of the study (at the end of the open-label period). As this is the final analysis for the primary objective of the study, no adjustment of alpha will be performed.

## 3.5    Sample Size

Table 1 provides the required sample size per treatment group when the number of migraine days per 28-day period is used as the primary outcome measure. Sample size calculations are based on the clinically meaningful difference between active and sham treatment groups in change from baseline. Assuming a type I error of 5% (two-sided) and 90% power, Table 1 displays the required sample sizes for different values of the common standard deviation (2.0, 2.5, 3.0 and 3.5) and the presumed difference between treatment groups (1, 1.5 and 2 days). Selecting an improvement of 1 day and a common standard deviation of 2.5 requires 133 subjects per treatment group. With 15% drop-out during the randomized treatment period, a total of 160 subjects per treatment group (or 320 subjects total) would need to be randomized.

**Table 1: Required Sample Size per Treatment Group**
**Assuming Type I Error ($\alpha$) = 5% (two-sided) and 90% Power**

| Variability (Common SD*, $\sigma$) | Clinically Significant Improvement: Active – Sham = $\Delta$ | | |
|---|---|---|---|
| | **1-day** | **1.5-day** | **2-day** |
| 2.0 | 86 | -- | -- |
| 2.5 | *133* | 60 | 34 |
| 3.0 | 191 | 86 | 49 |
| 3.5 | 259 | 115 | 66 |

*Estimated standard deviation for change from baseline.  Standard deviations for baseline and post treatment scores are between 2.0 to 3.5.

## 3.6    Study Procedures

Procedures and outcome measures will be collected as described in the clinical investigation flow chart shown in Table 2.

**Table 2. Clinical Investigation Flow Chart**

| Period | Run-in | Randomized | | | | | Open label | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Visit number | V1 | V2 | V3* | V4 | V5 ** | V6 | V7 * | V8 | V9** | V10 | V11 | V12 |
| Week | -4 | 0 | 1 | 4 | 8 | 12 | 13 | 16 | 20 | 24 | 30 | 36 |
| Day | -28 | 1+/-3 | 7+/-3 | 28+/-3 | 56 +/- 5 | 84 +/-5 | 91 +/-3 | 112 +/- 3 | 140+/-5 | 168+/-5 | 210 +/-5 | 252+/-5 |
| Signed Consent Form | ✓ | | | | | | | | | | | |
| Randomization | | ✓ | | | | | | | | | | |
| Demographics | ✓ | | | | | | | | | | | |
| Inclusion/Exclusion criteria | ✓ | | | | | | | | | | | |
| History of migraine | ✓ | | | | | | | | | | | |
| Medical and surgical history | ✓ | | | | | | | | | | | |
| Physical examination | ✓ | | | | | | | | | | | |
| ECG | ✓ | | | | | | | | | | | |
| Urine pregnancy test[1] | ✓ | | | | | | | | | | | |
| Blood pressure and weight | ✓ | | | | | ✓ | | | | | | ✓ |
| HIT-6[3] | | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ |
| EQ-5D-5L[3] | | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ |
| MIDAS[3] | | ✓ | | | | ✓ | | | | ✓ | | ✓ |
| Subject satisfaction question | | | | | | ✓ | | | | | | ✓ |
| Easy to use question | | | | | | ✓ | | | | | | ✓ |
| Blinding question | | | ✓ | | | ✓ | | | | | | |
| Device training and placement | | ✓ | | ✓ | | ✓ | | | | | | |
| Collect device/ gammaCore®-R | | | | | | ✓ | | | | | | ✓ |
| Diary completion | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Concomitant medication | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Adverse Event[2] | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Study Termination | | | | | | | | | | | | ✓ |

[1] Fertile woman only.

[2] All new events during the run-in will be reported under adverse events.

Call* - One week after the previous visit.

Call **- One month after the previous visit.

[3] At visit 2 and 6 these questionnaires will be completed by the subjects on paper forms during the visit.

## 4. DATA AND ANALYTICAL QUALITY ASSURANCE

The study endpoints and analytic approaches are both prospectively defined and documented in the protocol and in this SAP. The SAP will be finalized, and protocol violations will be identified and decisions for inclusion and exclusion of subjects from the per-protocol analysis population will be made prior to the database lock, unblinding, and data analysis.

## 5. ANALYSIS POPULATIONS

Subjects screened but not included in the study will not be presented in any listings or tables. Subjects that enter Run-in but do not meet criteria for randomized will be accounted for in tables, and their data will be provided in listings.

### 5.1 Intent-to-Treat (ITT) Population

The Intention to treat (ITT) population includes all randomized subjects with at least one verified treatment post-training during the 12 week randomized period. The ITT population will be analyzed using the subject's randomized treatment regardless of which treatment the subject actually received.

The ITT is considered the primary analysis dataset, and will be used for all primary and secondary efficacy analyses.

### 5.2 Per Protocol (PP) Population

The per protocol (PP) population includes all ITT subjects who received the correct treatment with no major protocol deviations which could affect the assessment of efficacy. Major protocol deviations include, but are not limited to, the following:

- The subject did not reach visit 6

- Less than 67% compliance with treatment per month in the randomized period (refer to section 6.3).

The final criteria for the PP population, regarding which protocol deviations warrant exclusions, will be determined when all data on protocol deviations are available and before breaking the blind.

If the difference in sample size between the ITT and PP populations is more than 5%, then the primary efficacy analyses will be repeated in the PP population.

### 5.3 Safety Population

The safety set will consist of all subjects who sign informed consent and are thus enrolled. The Safety population will be analyzed based on actual treatment received.

## 6. STUDY DAY AND VISIT WINDOW DEFINITION

### 6.1 Study Day

For each subject, the day of randomization will be considered study day 0. The day of the week 12 assessment will be considered study day 0 in the open label period. Each assessment will be assigned a study day. The calculation for study day is dependent on whether the actual date of assessment is before or after the date of randomization and is calculated as follows:

Run in period:

Study day = (date of enrollment – randomization date) (results in negative numbers)

Randomized period (Visits up to but not including the week 12 assessment):

Study day = (date of assessment – randomization date)

Open label period (Starting with week 12 assessment):

Study day (with reference to randomization date) = (date of assessment – randomization date - 1)

Study day (in open label period) = (Date of assessment – date of entry into OL period)

If the full date of the assessment is not known, then no study day will be assigned.

### 6.2 Baseline (Run-in Period)

Generally, a baseline measurement refers to the last non-missing assessment made before the first administration of study treatment (e.g., screening and baseline). For each subject, the run-in period will include the last 28 days of diary data prior to randomization.

### 6.3 Randomized Period

For each subject, the randomized period will be divided into 28-day intervals based on randomization date. An extra interval is included to allow for the visit windows specified in the protocol. These intervals may be used in exploratory analyses.

| Month in Randomized Period | Study Days |
|---|---|
| Month 1 | 1-28 |
| Month 2 | 29-56 |
| Month 3 | 57-84 |
| Month 4 | 85+ |

### 6.4 Open-Label Period

For each subject, the open-label period will be divided into 28-day intervals based on number of days in the open-label period, starting with issuance of the open-label period device. The date of issuance of the open-label period device is OL Day 0. An extra interval is included to allow for the visit windows specified in the protocol. These intervals may be used in exploratory analyses.

| Month in Open-Label Period | Days in OL Period |
|---|---|
| OL Month 1 | 1-28 |
| OL Month 2 | 29-56 |
| OL Month 3 | 57 - 84 |
| OL Month 4 | 85 - 112 |
| OL Month 5 | 113 - 140 |
| OL Month 6 | 141 - 168 |
| OL Month 7 | 169+ |

OL = Open-label

## 7.    SPECIFICATION OF ENDPOINTS AND VARIABLES

### 7.1    Demographic and Baseline Characteristics

Sex, age, and race can be used as collected on the eCRF.

### 7.2    Migraine History

Years since migraine diagnosis will be derived as:

(date of informed consent – date of diagnosis)/365.25

and truncated to 1 decimal point when a full date is available. Otherwise the following partial date rules will be followed:

1.  If month and day of diagnosis are missing: The year of diagnosis will be compared to the year in which study informed consent was obtained. If the two years are not the same, then July 1 will be used. If the two years are the same, then January 1 will be used.

2.  If day of diagnosis is missing: The year of diagnosis will be compared to the year in which study informed consent was obtained. If the two years are not the same, then 15 will be used for the day. If the two years are the same, then the month of diagnosis will be compared to the month in which study informed consent was obtained. If the months are not the same, then 15 will be used for the day.  If the two months are the same, then the first of the month will be used.

Age of onset, number of migraine days in the 4 weeks prior to enrollment, number of headaches (including migraine days) in the 4 weeks prior to enrollment, and number of days per month using acute medications can be used as collected on the eCRF. Whether the migraine is with or without aura will also be used as collected on the eCRF.

### 7.3    Medical and Surgical History

Medical and surgical history will be collected and coded using the current version of the Medical Dictionary for Regulatory Activities (MedDRA v18.1).

## 7.4 Efficacy

### 7.4.1 Primary Endpoint

The primary endpoint is the change from run-in to the last four weeks (Month 3) of the randomized period in the number of migraine days.

The diary data will collect for each day an indicator (1=Yes, 0=No, or missing) of whether the subject experienced a migraine. One 24 hour period per migraine will be designated as a migraine day.

The visit level value for a given subject is then the sum of the daily diary day indicators of whether the subject experienced a migraine which gives the number of days with a migraine. Calculations are as follows:

$$\text{Run-in Value} = \sum_{i=-z-28}^{-z} Day\ i, where\ Day\ i = 1\ (had\ a\ migraine)$$

where $z$ represents the subject's last day with diary data during the run-in period.

In the event that a subject does not 28 diary days in the run-in period the following rules will be used to calculate the run-in value:

1. If the number of diary days in the run-in period is less than 21, the run-in value will not be calculated, and will be considered missing.

2. If number of diary days in the run-in period is greater than or equal to 21, but less than 28, the run-in value would be calculated as:

$$\text{Run-in Value} = \sum_{i=-z-x}^{-z} Day\ i, where\ Day\ i = 1\ (had\ a\ migraine)$$

where $z$ represents the subject's last day with diary data during the run-in period, and x is the number of days with diary data during the run-in period.

To ensure consistency of the run-in value for all subjects, the value will be adjusted to reflect an estimated number of days with migraine per 28 day interval. For example, if a subject records data for x days out of 28 days and have y days with migraine, then the adjusted number of days with migraine in 28 days is (28/x)*y, with a maximum value of 28.

$$\text{Month 3 Value} = \sum_{i=57}^{84} Day\ i, where\ Day\ i = 1\ (had\ a\ migraine)$$

In the event that a subject does not have at least 84 study days in the DB period the following rules will be used to calculate the Month 3 value:

3. If the number of study days in the DB period is less than 70, the month 3 value will not be calculated, but rather imputed to the run-in value (and thus resulting in a Change$_{Run-in}$=0, implying no change between periods).

$$\text{Month 3 Value} = \text{Run-in Value}$$

4. If number of study days in the DB period is greater than or equal to 70, but less than 84, the month 3 value would be calculated as:

$$\text{Month 3 Value} = \sum_{i=z-28}^{z} Day\ i, where\ Day\ i = 1\ (had\ a\ migraine)$$

where $z$ represents the subject's last day of participation with diary data present.

To ensure consistency of this value for all subjects, the value will be adjusted to reflect an estimated number of days with migraine per 28 day interval. For example, if a subject records data for x days out of 28 days and have y days with migraine, then the adjusted number of days with migraine in 28 days is (28/x)*y, with a maximum value of 28.

Then the change from Run-in is calculated using the adjusted values as:

$$\text{Change}_{Run-in} = \text{Month 3 value} - \text{Run-In value}$$

### 7.4.2 Secondary Endpoints

*Percentage Reduction in Migraine Days and Responder Rate*

The percentage reduction in migraine days from the run-in period and achievement of 50% reduction in migraine days during the last 4 weeks of the randomized period are two secondary endpoints. Percentage reduction in migraine days is calculated as:

Percentage Reduction = -100 x (Change$_{Run-in}$ / Run-in Value)

Using this value, an indicator of achievement of 50% reduction during the last 4 weeks of the randomized period can be defined as:

Response = 1 (Yes) if Percentage Reduction $\geq$50%
Response = 0 (No) if Percentage Reduction is non-missing and < 50%

The number of migraine days, the change from run-in, and the percentage change from run-in to the open-label period will also be calculated. This will be completed for Month 3 and Month 6 of the OL period. For

$$\text{Month 3 OL Value} = \sum_{i=57}^{84} Day\ i, where\ Day\ i = 1\ (had\ a\ migraine)$$

In the event that a subject does not have at least 70 days in the OL period the following rules will be used to calculate the Month 3 OL value:

1. If the number of study days in the OL period is less than 70, the month 3 OL value will not be calculated, but rather imputed to the DB value (and thus resulting in a $Change_{OL3}= Change_{Run-in}$, implying no change between the 3 month DB and 3 month OL periods).

    Month 3 OL Value =Month 3 (DB) Value

2. If number of study days in the OL period is greater than or equal to 70, but less than 84, the month 3 OL value would be calculated as:

$$\text{Month 3 OL Value} = \sum_{i = z - 28}^{z} Day\ i, where\ Day\ i = 1\ (had\ a\ migraine)$$

where $z$ represents the subject's last day (within 84 days) of participation in the OL period with diary data present.

To ensure consistency of this value for all subjects, the value will be adjusted to reflect an estimated number of days with migraine per 28 day interval. For example, if a subject records data for x days out of 28 days and have y days with migraine, then the adjusted number of days with migraine in 28 days is (28/x)*y, with a maximum value of 28.

Then the change from Run-in is calculated using the adjusted values as:

$Change_{OL3}$ = Month 3 OL value – Run-In value

$$\text{Month 6 OL Value} = \sum_{i = 141}^{168} Day\ i, where\ Day\ i = 1\ (had\ a\ migraine)$$

In the event that a subject does not have at least 154 days in the OL period the following rules will be used to calculate the Month 6 value:

1. If the number of study days in the OL period is less than 154, the month 6 OL value will not be calculated, but rather imputed to the Month 3 OL value (and thus resulting in a $Change_{OL6}= Change_{OL3}$, implying no change between the 6 month OL and 3 month OL periods).

    Month 6 OL Value =Month 3 OL Value

2. If number of study days in the OL period is greater than or equal to 154, but less than 168, the month 6 OL value would be calculated as:

$$\text{Month 6 OL Value} = \sum_{i = z - 28}^{z} Day\ i, where\ Day\ i = 1\ (had\ a\ migraine)$$

where *z* represents the subject's last day of participation (within 168 days) in the OL period with diary data present.

To ensure consistency of this value for all subjects, the value will be adjusted to reflect an estimated number of days with migraine per 28 day interval. For example, if a subject records data for x days out of 28 days and have y days with migraine, then the adjusted number of days with migraine in 28 days is (28/x)*y, with a maximum value of 28.

Then the change from Run-in is calculated using the adjusted values as:

$$\text{Change}_{OL6} = \text{Month 6 OL value} - \text{Run-In value}$$

Results will be stratified by randomized treatment group.

In addition, the change from the double-blind period to 3 and 6 months of the open label period will be calculated or derived in the same manner for the comparison of the run-in to double-blind, randomized phase. Results will be stratified by randomized treatment group. For the calculation of percent change from the double-blind period to each of the open label periods, if there were zero migraine days at the double-blind period, percent reduction will be imputed as 100%.

$$\text{Change}_{OL3DB} = \text{Month 3 OL value} - \text{Month 3 value}$$

$$\text{Change}_{OL6DB} = \text{Month 6 OL value} - \text{Month 3 value}$$

*Headache Days and Headache Medications*

Change from run-in to the last 4 weeks of the randomized period in the number of headache days (as opposed to migraine days) is another secondary endpoint and will be calculated in the same manner as the primary endpoint.

The number of days in which acute headache medications were used; and, the change in the number of days from run-in to the last 4 weeks of the randomized period are also secondary endpoints. Similar to the primary endpoint, an indicator of acute headache medication usage is collected in the e-diary and the number of days of usage and the change in the number of days from run-in to the last 4 weeks of the randomized period will be calculated in the same manner as the primary endpoint. In addition, the number of days in which 1) simple analgesics (aspirin, paracetamol, ibuprofen, NSAIDs), 2) triptans and ergots, 3) combined analgesics, and 4) opioids were used, and the change in number of days from run-in to the last 4 weeks of the randomized period will be analyzed similarly.

*MIDAS*

The total score on the MIDAS is captured on the eCRF, but will be recalculated for analysis as the sum of the response to the questions. Should any question be blank or unanswered, no total score will be calculated. The MIDAS eCRF will be completed at V2 (the end of the run-in period), V6 (the end of the

randomized period), V10 (12 weeks following the start of the OL period), and at V12 (24 weeks following the start of the OL period). The change in the MIDAS total score from the run-in period to each assessment will be calculated. Change in MIDAS total score may also be calculated from the end of the DB period to each assessment in the OL period.

The level of disability on the MIDAS, is defined using the total score as:

| | |
|---|---|
| 0 to 5 | MIDAS Grade I, Little or no disability |
| 6 to 10 | MIDAS Grade II, Mild disability |
| 11 to 20 | MIDAS Grade III, Moderate disability |
| 21+ | MIDAS Grade IV, Severe disability |

*HIT-6*

The total score on the HIT-6 is captured on the eCRF, but will be recalculated for analysis as the sum of the response to the questions. Should any question be blank or unanswered, no total score will be calculated. The HIT-6 eCRF will be completed at V2 (the end of the run-in period), V4, V5, V6 (the end of the randomized period), V8, V9, V10 (12 weeks following the start of the OL period), and at V12 (24 weeks following the start of the OL period). The change in the HIT-6 total score from the run-in period to each assessment will be calculated. Change in HIT-6 total score may also be calculated from the end of the DB period to each assessment in the OL period.

*EQ-5D-5L*

The EQ-5D is a generic, preference-based health utility measure with questions about mobility, self-care, usual activities, pain/discomfort, and anxiety/depression that are used to build a composite of the subject's health status. Each question is scored as 1 for no problem, 2 as slight problems, 3 as moderate problems, 4 as severe problems, and 5 indicating extreme problem. Ambiguous values (eg. 2 boxes are ticked for a single dimension) will be treated as missing (neither value used).

The 5 scores will then be converted into a single summary index by using the EQ-5D-5L Index Value Calculator tool (http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html#c655, last accessed 18-Jan-2016) published by the EuroQol group. If any of the 5 scores are missing, the summary index will not be calculated and set to missing.

The second component of the EQ-5D is a visual analogue scale (VAS), asking subjects to rate their health from 0 to 100 where 0 represents the worst imaginable health state and 100 represents the best imaginable health state.

The EQ-5D will be completed at V2 (the end of the run-in period), V4, V5, V6 (the end of the randomized period), V8, V9, V10 (12 weeks following the start of the OL period), and at V12 (24 weeks following the start of the OL period). The change in the EQ-5D total score and change in the EQ-5D VAS from the run-in period to each assessment will be calculated. Change in EQ-5D total score and change in the EQ-5D VAS may also be calculated from the end of the DB period to each assessment in the OL period.

*Other Endpoints*

The study subject will complete blinding questionnaires at 1-week (V3) into the randomized period and at the completion of randomized period (V6) as part of the data collection requirements. The blinding questionnaire asks subjects which treatment they think they received ("Active Stimulation", "Sham Control", or "Don't know").

## 7.5 Safety

Common safety variables include the following:

- Extent of Device Usage
- Adverse events
- Concomitant medications/treatments
- Vital signs

### 7.5.1 Extent of Device Usage

Administration of the 2 stimulations in the morning, midday and evening is collected in the e-diary as well as the number of stimulations if different from the planned number. In addition, dispensing and return of dispensed devices is recorded in the eCRF.

Extent of device usage will be defined as the percentage of days a subject used the device. Extent of device usage will be presented for each four week interval.

Compliance with device usage will be computed as the percentage of expected stimulations actually administered. The number of expected stimulations is determined for **each four week interval** as:

28 days x 3 times per day x 2 stimulations each time = 168 expected stimulations per month

If there are fewer than 28 days in a time period then the number would be modified accordingly.

The actual number of stimulations actually administered is calculated as the sum of the actual administrations for the same time interval as for the expected stimulations.

Compliance is then calculated as the actual number administered / expected number administered and multiplied by 100 to obtain the percentage.

Compliance will be then categorized into groups according to full compliance: 100%, 67%-<100%, 33%-<67% and <33%.

### 7.5.2 Adverse Events

Adverse events will be collected and coded using the current version of the Medical Dictionary for Regulatory Activities (MedDRA v18.1).

An adverse event is considered treatment-emergent if the date of onset is on or after the date of first stimulation in the randomized period, or worsened during the randomized or open-label periods (intensity and/or severity changed to a worse grade).

**Events with Irregular Start Dates:** All adverse events will be included in the tabulations regardless the completeness of the onset dates. Partial dates may be imputed when appropriate.

If a partial date is reported for the start of an adverse event, a complete date will be estimated by the following algorithm in order to determine treatment emergent:

1. If month and day are missing: The year of the event will be compared to the year in which study treatment began. If the two years are not the same, then January 1 will be used. If the two years are the same, then the earlier of the date of randomization and the date of first usage of the device will be used.

2. If day is missing: The month and year of the event will be compared to the date in which study treatment began. The onset date will be imputed as the date closest to (being either on or after) the earlier of the date of randomization and the first date of usage of the device. If the two did not occur in the same month and year, then 1 will be used for the day.

If a partial date is reported for the end of an adverse event and the adverse event is not continuing, a complete date will be estimated by the following algorithm:

1. If month and day are missing: The year of the event will be compared to the year of final contact with the subject. If the two years are not the same, then December 31 will be used. If the two years are the same, then the date of final contact with the subject will be used.

2. If day is missing: The year of the event will be compared to the year of final subject contact. If the two years are not the same, then the last day of the month will be used for the day. If the two years are the same, then the month of the event will be compared to the month of final subject contact. If the months are not the same, then the last day of the month will be used for the day. If the two months are the same, then the date of final subject contact will be used.

### 7.5.3 Concomitant Medications/Treatments

Verbatim terms will be coded by Sponsor and assigned an Anatomic-Therapeutic-Chemical (ATC) term using the World Health Organization Data Dictionary Extended (WHO-DDE), version 2015Sep.

## 8. STATISTICAL ANALYSIS

### 8.1 General Data Handling Rules and Definitions

If any randomized subject is found to not have valid documented informed consent, that subject's data will be excluded from the report, except as necessary to document the error.

Except where specified all continuous variables will be summarized with descriptive statistics (the number of non-missing values, mean, standard deviation, median, minimum and maximum) and all categorical variables will be summarized with frequency counts and percentages, by treatment group.

Missing data will be maintained as missing unless specified otherwise (see Section 7.4.1 (number of migraine/headache days and associated calculations)). For variables where missing data is imputed, the analysis dataset will include an indication that it is imputed.

### 8.2 Subject Disposition

The number and percentage of subjects enrolled, randomized, treated at least one time, completing/withdrawing from each period and from the study overall will be tabulated for each treatment

group and overall. Reasons for early withdrawal will be summarized by treatment group using numbers and percentages of subjects.

The number of subjects included and excluded from the analysis populations (safety, ITT, and PPS) will be summarized for each treatment group. In addition, the reason(s) for exclusion from each population will be summarized.

## 8.3  Demographic and Baseline Characteristics

Descriptive statistics of demographic variables (e.g. age, gender and race), and other baseline characteristics will be presented for the safety population by treatment group and overall. For quantitative variables, all summaries will include the number of non-missing observations, mean, median, standard deviation, minimum, and maximum. For qualitative variables, the summaries will include the number and percentage of subjects within each category for the outcome.

## 8.4  Efficacy Analyses

## 8.4.1  Primary Efficacy Analysis

The primary efficacy analyses will be performed on the ITT population and will be repeated in the PP population, if the sample size between the two groups differ by more than 5%.

Summary statistics for the primary variable, change from 4-week run-in period, will be presented by treatment group. Missing data will be imputed per section 7.4.1.

The number of migraine days will be summarized descriptively for each study phase (run-in and DB) and by treatment group. The summary statistics for each visit, change between the double-blind and 4-week run-in period and percentage change between the double blind and 4-week run-in period (a secondary endpoint) will be summarized by mean, standard deviation, median, minimum, and maximum.

The two-sided, 95% confidence interval (CI) of the difference between the two device groups (active minus sham) in change from 4-week run-in period will be produced and the p-value will be calculated based on a two-sample t-test.  If the p-value from the two-sample t-test is less than 0.05, then the null hypothesis will be rejected.

As sensitivity analyses, the two groups will be compared using analysis of covariance (ANCOVA) models with change from 4-week run-in period as the dependent variable and treatment group, and center as fixed effects with the number of migraine days in the 4-week run-in period as a covariate and again with treatment group, center, and aura as fixed effects with number of migraine days in the 4-week run-in period as a covariate.  The adjusted means, adjusted mean difference and 95% confidence interval for the adjusted mean difference based on the ANCOVA models will be reported. The p-values tests that the adjusted mean difference between groups at Month 3 equals zero will be reported.

Further sensitivity analyses will investigate the robustness of the analysis to missing data (missing data refers to data points imputed in the calculations presented in section 7.4.1) by implementing multiple imputation and the pattern-mixture model (PMM) with control-based pattern imputation as summarized by Ratitch and O'Kelly (2011).  This PMM sensitivity analysis addresses the possibility that the missing data is missing not at random (MNAR), meaning the missing data is not independent of the unobserved outcomes

even after conditioning on the observed data. Please see Appendix 1 for a more detailed description of the steps to implement the multiple imputation and PMM models.

### 8.4.2 Secondary Efficacy Analyses

*Responder Rate*

The number and percentage of subjects with at least a 50% reduction in migraine days will be summarized by device group, for each study phase (DB, 3 month OL, 6 month OL). The difference between the two device groups in will be compared using a two-sided, 95% CI (active minus sham). The p-value will be calculated using Fisher's exact test. If the p-value from the resulting exact test is <0.05, then the null hypothesis will be rejected.

As a sensitivity analysis, for responder rate in the DB phase, the two treatment groups will be compared using logistic regression analyses with an indicator for achieving at least a 50% reduction in migraine days (0 indicates no, 1 indicates Yes [event]) as the dependent variable and treatment group and  center as fixed effects with number of migraine days in the 4-week run-in period as a covariate and again with treatment group, center, and aura as fixed effects with number of migraine days in the 4-week run-in period as a covariate.. The adjusted proportions, odds ratio, and 95% CIs f will be reported.

In addition, summary statistics and two-sided 95% CI will be produced for the following secondary effectiveness outcome measures and compared between the two treatment groups:

- The percentage reduction in migraine days during each time period (DB, 3 month OL, and 6 month OL) compared to the run-in period.

- The number of headache days during each time period (Run-in, DB, 3 month OL, and 6 month OL) as well as the change in number of headache days from the four week run-in period.

- The number of days in which acute headache medications were used during each time period (Run in, DB, 3 month OL, and 6 month OL); and, the change in the number of days from run-in.

- Scores at each visit and change in quality of life from run-in to each assessment as measured by the EQ-5D-5L,  HIT-6, and MIDAS.

In addition, similar summaries may be produced for change in outcomes reported from the end of the DB period to assessments in the OL period.

*Shift in MIDAS Categories*

In addition to the analyses described above, a shift analysis of the MIDAS level of disability categories from baseline to each later assessment will be presented for each treatment group. The proportion of subjects who shift from moderate or severe disability (Grade III or IV) during the run-in period to little or no disability (Grade I) and the proportion of subjects who shift from moderate or severe disability (Grade III and IV) to mild, little or no disability (Grade II and I) at each later assessment will be evaluated using a chi-square test.

### 8.4.3   Other Efficacy Analyses

### 8.4.3.1   Blinding Assessment

To assess whether subjects were blind to device assignment, the Bang blinding index will be used.  The number and percentage of subjects reporting each category will be reported. The Bang blinding index and the corresponding two-sided, 95% CI of the blinding index will be estimated for each treatment group as described in Appendix 2. If the two-sided, 95% confidence interval includes 0, then the null hypothesis of unsuccessful blinding will be rejected.

### 8.4.3.2   Subject satisfaction question

The number and percentage of subjects reporting each choice will be reported by device group.

### 8.4.3.3   Ease of use

The number and percentage of subjects reporting each choice will be reported by device group.

### 8.4.3.4   Subgroup Analyses

The primary efficacy analysis and key secondary analyses will be repeated in subgroups of subjects based on the presence or absence of aura. The subgroup analyses will be performed in the ITT population. The analyses to be repeated are:

- Number of migraine days: comparison of treatment groups

- Responder rate as measured by percentage reduction in migraine days

- Number of headache days: comparison of treatment groups

- Number of days with acute headache medication: comparison of treatment groups

Analyses will be performed for each subgroup separately.

Additional subgroup analyses stratified by sex and/or stratified by race may also be performed.

### 8.4.3.5   Complete Population Analysis

The primary efficacy analysis and key secondary analyses will be repeated in subjects who have non-missing/non-imputed data for the run-in period and the double-blind period. The analyses to be repeated are:

- Number of migraine days: comparison of treatment groups

- Responder rate as measured by percentage reduction in migraine days

- Number of headache days: comparison of treatment groups

- Number of days with acute headache medication: comparison of treatment groups

### 8.4.3.6   Pooling Analyses

Poolability of study sites will be assessed by comparing change in migraine days= trt*site interaction; if p<0.10, results will be stratified by site. Sites with less than five (5) subjects enrolled will be combined into a pseudo-site for purposes of analysis.  To protect against having an overly large pseudo-site, when one

pseudo-site exceeds ten subjects, a second pseudo-site will be formed. This process will continue as needed each time a pseudo-site exceeds ten subjects.

## 8.5    Safety Analyses

### 8.5.1   Device Exposure and Compliance

The number and percentage of subjects administering at least 1 stimulation for each time period during the randomized period, and separately during the open-label period will be summarized by device group.

Compliance with device usage will be summarized descriptively by device group for each time period.

### 8.5.2   Adverse Events/Adverse Device Effects

Analysis of AE/ADEs will be carried out on the safety population. All AE/ADEs will be included in the analyses, summaries, and individual subject data listings.

**AE/ADE Counting Rules:**

- A subject with more than one different adverse event in a particular system organ class (SOC) will be counted only once in the total of subjects experiencing adverse events in that particular SOC.

- A subject having experienced the same event (AE preferred term) more than once during the study will be counted only once in the number of subjects with that event.

- If an event changes in intensity or in seriousness during the study, a subject will be counted only once with the worst grade and seriousness respectively.

- If the causal relationship to the study device is assessed differently, a subject will be counted only once by considering the "Worst" documented degree of relationship.

Frequency and percentage of subjects experiencing a specific adverse event will be tabulated by system organ class, preferred term, and treatment group for each study phase (i.e., double-blind randomized period, open-label period). Summary tables and subject data listings will be provided for adverse events by maximum intensity, device-related adverse events, serious adverse events, adverse events causing discontinuation of the device or participation in the study, and deaths.

### 8.5.3   Concomitant Medications/Treatments

A listing of concomitant treatments and medications will be generated.

### 8.5.4   Vital Signs

Vital signs (blood pressure, pulse, height, weight and BMI) will be summarized using the number of non-missing observations, mean, median, standard deviation, minimum, and maximum. Summaries will be produced by device group for each assessment time.

### 8.5.5   Electrocardiogram

A listing of ECG results will be generated.

### 8.5.6   Physical Examination

The number and percentage of subjects with abnormal findings on the physical examination will be summarized for each device group by physical examination system.

### 8.6   Exploratory Analyses

Additional, ad hoc exploratory analyses may also be conducted.

## 9.   STATISTICAL SOFTWARE

All tables, listings and figures will be primarily produced using SAS Version 9.3 (SAS Institute, Cary, NC) or a later version of SAS and R.

## 10.   REFERENCES

Bang, H., Flaherty, S.P., Kolahi, J., and Park, J.  Blinding assessment in clinical trials: A review of statistical methods and a proposal of blinding assessment protocol, *Clinical Research and Regulatory Affairs*, 2010; **27**(2): 42–51.

EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument. Version 2.1, April 2015. EuroQol Group. *http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-5L_UserGuide_2015.pdf (last accessed on 18 January 2016).*

van Hout B, Janssen MF, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health*, 2012 Jul-Aug;**15**(5):708-15. (*http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html#c655*, last accessed on 18-Jan-2016)

## Appendix 1: Sensitivity Analyses: Multiple Imputation and Control-based Pattern Mixture Model

Multiple imputation will be used to impute missing data. First, any non-monotone missing data patterns will be made monotone by imputing the intermittent missing data using the Monte Carlo Markov Chain (MCMC) approach. This imputation will be performed 100 times, resulting in a dataset with 100 copies of monotone data. If no non-monotone missing data patterns exist, this step is not necessary. Missing data will then be imputed using the monotone regression method. The following factors/covariates will be included in the model: treatment, center and number of migraines in the 4-week run-in period.

The data will be analyzed by t-test and ANCOVA, for each of the 100 imputations, and the results combined using standard multiple imputation methodology to obtain the appropriate estimates of variability.

The control-based pattern mixture model is a sensitivity analyses which addresses the possibility that the missing data is missing not at random (MNAR). The sham group will act as the control; and, the observed data from this group will be used to impute the missing data in both the sham and active groups.

The method is designed to impute data from monotone missing data patterns, thus the patterns of missing data are first examined and described. Any non-monotone missing data patterns will be made monotone first by imputing the intermittent missing data using the Monte Carlo Markov Chain (MCMC) approach. This imputation will be performed 100 times, resulting in a dataset with 100 copies of monotone data. If no non-monotone missing data patterns exist, this step is not necessary.

To implement the control-based pattern mixture model, the missing endpoint assessment for treatment and sham subjects is then imputed based on a linear regression model using observed data in the sham group, with the missing assessment as the dependent variable. The following factors/covariates will also be included in the model: treatment, center and number of migraines in the 4-week run-in period.

The data will be analyzed by t-test and ANCOVA, for each of the 100 imputations, and the results combined using standard multiple imputation methodology to obtain the appropriate estimates of variability.

An example of SAS code that may be used to implement the MCMC intermittent missing data imputation is:

```
proc mi data=indsn out=outmono nimpute=100 seed=startseed ;
    class trt center;
    var trt center base depvar;
    mcmc chain=multiple impute=monotone;
run;
```

An example of SAS code that may be used to implement the control-based PMM is:

```
proc mi data=missdsn out=outdsn nimpute=100 seed=startseed2 ;
```

```
    class trt center;
    var center base depvar;
    monotone reg(/details);
    mnar model(depvar / modelobs=(trt="sham"));
run;
```

## Appendix 2: Bang Blinding Index

To assess the blinding , the Bang blinding index and corresponding confidence interval will be estimated as follows. A 2x3 table will be constructed as shown:

| | | Guess | | |
|---|---|---|---|---|
| **Assignment** | Active | Sham | Do not Know (DK) | Total |
| Active | $n_{11}(p_{1|1})$ | $n_{12}(p_{2|1})$ | $n_{13}(p_{3|1})$ | $n_{1.}$ |
| Sham | $n_{21}(p_{1|2})$ | $n_{22}(p_{2|2})$ | $n_{23}(p_{3|2})$ | $n_{2.}$ |
| Total | $n_{.1}$ | $n_{.2}$ | $n_{.3}$ | N |

Where $p_{j|i}$ = P(guess j|assignment i) for i =1(Active), 2(Sham), and j=1(Active), 2(Sham), 3(DK), where DK denotes Don't Know. N is the total number of respondents.

If data is constructed in a 2x3 format as shown in the above table, let us define

$$r_{j|i} = p_{j|i}/(p_{1|i} + p_{2|i})$$

which is the proportion of correct guesses among participants who provided certain guesses other than DK in the i-th arm.  Bang's BI is defined as $(2 r_{j|I} - 1)*(p_{1|i} + p_{2|i})$ and can be estimated by:

$$\text{Bang's BI} = (2n_{ii}/(n_{i1} + n_{i2}) - 1)*((n_{i1} + n_{i2})/( n_{i1} + n_{i2} + n_{i3})$$

And its variance is given by

$$\text{Var(Bang's BI)} = (p_{1|i} (1 - p_{1|i}) + p_{2|i}(1 - p_{2|i}) + 2p_{1|i} p_{2|i})/n_{i.}$$

Bang's BI is between -1 and +1 inclusive and can be interpreted as the proportion of participants who answered correctly on the ith arm beyond chance level.

A rule of thumb explanation on how to interpret Bang's BI is provided as follows:

- If $r_{j|i}$=1 and $n_{i3}$=0 (means all responses are correct), then $BI_i$=1 (complete unblinding)

- If $r_{j|i}$=0 and $n_{i3}$=0 (means all responses are incorrect), then $BI_i$=-1 (opposite guessing)

- If $r_{j|i}$=0.5 (means 50% responses are correct), then $BI_i$=0 (random guessing)

So, it is desired to show $BI_i$ around 0.  To test successful blinding, the null and alternative hypotheses can be constructed as follows:

$H_0$: unsuccessful blinding vs $H_1$: successful blinding