spirox

Statistical Analysis Plan

Spirox Latera™ implant support of lateral nasal wall cartilage (LATERAL-OR) study

Protocol SPI-CP-301

Revision Date: 03Jan2018

Approved By:

Vaishali Suraj
Director, Biometrics

Date

Elisa Hebb
Vice President, Clinical and Regulatory Affairs

04 Jan 2018
Date

spirox

Statistical Analysis Plan

Spirox Latera™ implant support of lateral nasal wall cartilage (LATERAL-OR) study

Protocol SPI-CP-301

Revision Date: 03Jan2018

Approved By:

Vaishali Suraj
Director, Biometrics

03Jan2018
Date

Elisa Hebb
Vice President, Clinical and Regulatory Affairs

Date

TABLE OF CONTENTS

ABBREVIATIONS

AAO-HNS	American Academy of Otolaryngology—Head and Neck Surgery
ADE	Adverse device effect
AE	Adverse Event
ANCOVA	Analysis Of Covariance
BMI	Body Mass Index
CI	Confidence Interval
CPAP	Continuous positive airway pressure
GCP	Good Clinical Practice
ICH	International Conference on Harmonisation
KM	Kaplan-Meier
LOCF	Last Observation Carried Forward
LS mean	Least Squares Mean
MedDRA	Medical Dictionary for Regulatory Activities
MH	Medical History
MID	Minimally important difference
mITT	Modified Intent-To-Treat
MMRM	Mixed Model for Repeated Measures
NOSE	Nasal Obstruction Symptom Evaluation
NVC	Nasal Valve Collapse
PP	Per Protocol
PRO	Patient-Reported Outcome
SADE	Serious adverse device effect
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS®	Statistical Analysis Software
SD	Standard Deviation
SE	Standard Error
SOC	System Organ Class
VAS	Visual Analog Scale

1. INTRODUCTION

This statistical analysis plan (SAP) contains the definitions of analysis sets, derived variables, and statistical methods for the analyses of efficacy and safety data from the Spirox Latera™ implant support of lateral nasal wall cartilage (LATERAL-OR) study, which is a multi-center, non-randomized, single arm study (Clinical Protocol SPI-CP-301). This study was designed to quantify changes in the symptoms of nasal airway obstruction associated with the use of the Spirox Latera™ implant. Subjects will be consented and enrolled prior to device implant, and assessed through 24 months post implant.

1.1. TRIAL OBJECTIVES

To obtain outcomes data in subjects with severe to extreme nasal obstruction undergoing placement of at least 1 Spirox Latera™ implant, with or without concurrent septoplasty and/or turbinate reduction procedures in operating room setting:

Primary Objective

- To assess the safety and efficacy of the Spirox Latera™ implant to 6 months post procedure

Major Secondary Objectives

- To assess the efficacy of the Spirox Latera™ implant at other time-points (1 month, 3 months, 12 months, 18 months, and 24 months)

1.2. TRIAL DESIGN

As described in the protocol, this is a prospective, multi-center, non-randomized study of the Spirox Latera™ implant in an intraoperative setting. These devices will be used to treat nasal valve collapse in approximately 170 subjects with extreme or severe symptoms of nasal obstruction. Follow-up assessments will occur at 1, 3, 6, 12, 18, and 24 months after device implant.

The primary performance measure of the Spirox Latera™ implant will be based on the Nasal Obstruction Symptom Evaluation (NOSE) instrument. This instrument measures the burden of nasal obstructive symptoms in increments of 5 points, and the total score ranges from 0 (asymptomatic) to 100. Categories for clinical severity are:

- Mild: Total score 0 – 25 points
- Moderate: Total score 30 – 50 points
- Severe: Total score 55 – 75 points

- Extreme: Total score 80 – 100 points

The primary endpoint will be the proportion of device responders at the 6 month follow-up visit. Subjects will be classified as responders if they meet at least 1 of the following criteria:

1. The 6 month NOSE score category has improved from the baseline category by at least 1 level
2. The 6 month change from baseline NOSE score is $\geq 20\%$

Subjects will be classified as non-responders if neither of these criteria is met.

1.3. Statistical Hypotheses for Trial Objectives

The primary study hypothesis is a superiority comparison to a target proportion:

Null Hypothesis H_0: the proportion of responders at 6 months is = 0.50, versus

Alternative Hypothesis H_A: the proportion of responders at 6 months is $\neq 0.50$

1.4. Sample Size Justification

The primary statistical hypothesis will be evaluated using an exact test. In order to have > 90% power to rule out a response rate $\leq 50\%$ at a 5% (2-sided) significance level, assuming a true response rate of 66%, 120 subjects with 6 months of follow-up are needed. Assuming a 20% drop-out rate, at least 150 subjects (and up to 170 subjects) will be enrolled. A 66% response rate is the minimum response rate that is consistent with a previous trial, where an 80% response rate at 6 months was observed.

Sample size calculations were performed using the package RCTDesign for R, version 3.2.3.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Visit Windows

The reference day is study day 1, the day of device implant. Study day will be calculated relative to the date of implant.

Baseline will be defined as the pre-implant measure closest to or including Day 1 (prior to implant procedure).

If a subject has two or more actual visits in one analysis visit window, the visit closest to the target day will be used as the study visit for that analysis visit window. If two study visits are the same number of days from the target day within the same visit window, the

later visit will be considered the study visit for that target day. Even though all visits will be allocated an analysis visit window, only planned protocol visits for each variable

2.2. Analysis Sets

Unless otherwise specified, all analyses will use the modified intent-to-treat analysis set (mITT), which is defined below.

2.2.1. Analysis Sets for Efficacy and Safety

The ITT analysis set will primarily be used for safety evaluations and will include all subjects who begin the implant procedure. Subjects who are consented but do not begin the implant procedure will not be included in the ITT analysis set. Because this is a single-arm, open-label study, the exclusion of subjects who do not begin the implant procedure is unlikely to result in biased response estimates. Subjects who begin the implant procedure but do not successfully receive an implant will be included in the ITT analysis set.

The mITT analysis set will consist of all subjects who successfully receive at least 1 Spirox Latera™ implant. Analyses supporting the primary and major secondary endpoints will, to the extent possible, be based on the mITT analysis set. Safety evaluations for events within 7 days of the implant procedure will be based on the ITT analysis set, and cumulative adverse events at 6 months will be based on the mITT analysis set. An important subgroup of the mITT analysis set is the 6 Month Evaluable (6ME) Set, which includes subjects with 6 month post-implant NOSE score assessments in clinic or by telephone.

The per-protocol (PP) analysis set will consist of all mITT subjects who have no major protocol deviations or events that might impact NOSE scores. These deviations may include changes to background nasal medications, prohibited medications and additional procedures prohibited by the protocol.

3. SUBJECT INFORMATION

3.1. Baseline Anthropometric and Demographics Characteristics

Baseline anthropometric and demographic characteristics will be summarized overall for the mITT and 6ME analysis sets. Descriptive statistics (N, mean, SD, median, and range) will be computed for continuous variables such as age, body weight, height, and BMI at baseline.

The distribution (as number and percentage of subjects) will be summarized for the following baseline demographic characteristics:

- Sex: Male, Female

- Race: Caucasian, African American, Asian, Other
- Ethnicity: Hispanic or Latino, not Hispanic or Latino
- Eyeglass wear: 1-5 Hours/dy, 6-10 Hours/dy, 11-24 Hours/dy

3.2. Baseline Disease Characteristics and Nose Demographics

3.2.1 Baseline Disease Characteristics

The number and percentage of subjects will be summarized for the following:

- History of Nasal Trauma
- History of Nasal Surgery
- Specific Surgical Types (e.g. Septoplasty, Rhinoplasty)
- Any Nasal Surgery with Unsatisfactory Results
- History of non-Surgical Nasal Treatment
- Sleep Apnea
- Any Nasal Obstruction Related Medication Use
- Presence of Nasal Allergies
- Frequency of Nasal Allergies – Number of Seasons: 1, 2, 3, 4 (year round)

3.2.2 Nose Demographics

Nose demographics will be reported in a similar manner for the ITT (or mITT) analysis sets. Continuous variables include days since endoscopy, nose length, nose bone length, nose height and nose width.

The following categorical variables will be summarized (number and percentage):

- Nose Skin Thickness: Thin, Normal, Thick
- Reason Endoscopy Not Recorded
- Presence of Deviated Septum
- Presence of Inferior Turbinate Hypertrophy
- Modified Cottle Results (by side)

3.3. Disposition Information

Subject disposition will be summarized with the following

- Subjects who are consented and enrolled but the implant procedure was not started
- Subjects in the ITT analysis set
- Subjects in the mITT analysis set
- Subjects in the mITT analysis set who discontinue before the 6-month assessment
- Subjects in the PP analysis set

- Subjects who complete the study (24 months of follow-up)

3.4 Latera™ Implant

3.4.1 Procedure Information

Procedure information will be summarized for the ITT Analysis Set by whether the subject underwent a unilateral or bilateral implant placement for subject-level variables. Continuous variables including anesthesia duration, procedure duration and Latera™ implant duration will be summarized with descriptive statistics (N, mean, SD, median, and range). Categorical variables summarized by number and percentage include:

- Procedure Type: Latera™ Alone, Latera™ + Septo, Latera™ + Turbs, Latera™ + Septo + Turbs
- Turbinate Reduction Method
- Any Adverse Events During Procedure

3.4.2 Device Performance

Device delivery data will be summarized by nostril. For descriptive tables, the within-subject correlation will be ignored. The following categorical variables will be summarized (number and percentage):

- Implant Successfully Loaded into Delivery Tool
- Implant Successfully Delivered
- Implant Exposed
- Implant Retrieved
- Delivery Device Malfunction
- Latera™ Implant Malfunction

3.4.3 Post Procedure

Categorical variables will be summarized (number and percentage):

- Subject Success (whether all planned implants were successfully delivered on first attempt)
- Number of Implants Delivered
- Any Adverse Events Prior to Discharge

3.5. Duration of Follow-Up

Follow-up duration is defined as the number of days from implant until the first of early study discontinuation, completion of the 24 months of follow-up visit, study termination, or database lock. The cumulative distribution of time under follow-up will be calculated and summarized using Kaplan Meier methods.

For time-dependent safety analyses of adverse events, the duration of follow-up will be defined as the number of days from implant until the last date when the presence or absence of the event in question could be assessed. For example, for Adverse Events (AE), the duration of follow-up would be the number of days from implant to the onset of the first AE for subjects reporting any event, and the number of days between implant and the last date when AEs were assessed for those who did not experience any event.

3.6. Protocol Deviations

The following list of protocol deviations may affect the interpretation of the primary and major secondary efficacy endpoints, and therefore, subjects with any of these protocol deviations will be excluded from the PP analysis set. The complete list of subjects with these protocol deviations will be identified prior to database lock, and a summary of reasons for exclusion will be tabulated.

- Subject did not meet eligibility criteria
- Follow-up evaluation performed outside protocol-required window
- Other protocol deviation that may affect the interpretation of efficacy endpoints, with justification

4. EFFICACY

The efficacy analyses will be performed based on the mITT analysis set. The analyses will be repeated for the PP set to determine the impact of protocol deviations on the primary results.

4.1. Analysis Specifications

4.1.1. Level of Significance

Unless otherwise specified, all statistical tests will be interpreted at a 2-sided significance level of 5% and all confidence intervals at a 2-sided confidence level of 95%. The study type 1 error rate of 5% will all be allocated to the primary endpoint hypothesis, so p-values beyond the primary hypothesis should be considered descriptive.

4.1.2. Data Handling Rules

4.2. Primary Efficacy Endpoint

4.2.1. Definition

The primary endpoint will be the proportion of treatment responders at 6 months post procedure. Subjects will be classified as responders if they meet at least 1 of the following criteria:

1. The 6 month NOSE score category has improved from the baseline category by at least 1 level

2. The 6 month change from the pre-implant average NOSE score is ≥ 20%

Subjects will be classified as non-responders if neither of these criteria is met. The baseline category is defined based on the pre-study average within-subject NOSE score.

4.2.2. Analysis Methods

The primary efficacy endpoint of 6 month response proportion will be assessed using a 2-sided exact test to determine whether the proportion of responders is significantly different than 0.5. Only subjects in the 6ME set will be included in the primary endpoint analysis. The corresponding exact (Clopper-Pearson) 2-sided 95% confidence interval will be calculated. The null hypothesis that the proportion of responders is = 0.5 will be rejected if the p-value is < 0.05.

4.2.2.1. Supportive and Sensitivity Analyses

The primary endpoint analysis will be repeated using the PP analysis set. In order to quantify the impact of missing data on the primary analysis results, missing data may be imputed using 2 techniques:

1. Last observation carried forward (LOCF)
2. Subjects with missing data classified as non-responders

Since the efficacy of the device may deteriorate over time, LOCF represents the least conservative imputation technique, while classifying subjects with missing data as non-responders represents the most conservative technique. These two analyses are reasonable upper and lower bounds for the response result, accounting for the missing data.

4.3. Subgroup Analyses for the Primary Endpoint

Analyses of the primary efficacy endpoint in the 6ME analysis set will be performed to examine the consistency of implant effect across subgroups listed below (if there are no fewer than 5-10 subjects within each subgroup level). The proportion of responders in each group and corresponding exact 95% confidence interval for response rate will be presented.

- Investigator subgroups
- Type of procedure (SPX alone versus combinations with septoplasty and/or turbinate reductions)
- Turbinate reduction method: Radiofrequency versus Mechanical

4.4. Major Secondary Endpoints

4.4.1. DEFINITIONS

The secondary efficacy endpoints are:

- Proportion of treatment responders at 1, 3, 12, 18 and 24 months post procedure.
- Change in nasal airway obstruction from baseline to 1,3, 6, 12, 18 and 24 months as reported by subjects on the VAS scale.
- Subject satisfaction questionnaire at 6 months.
- Procedure and device-related adverse events though 24 months.

4.4.2. ANALYSIS METHODS

Primary endpoint will be analyzed using a mixed model for repeated measures (MMRM). The analysis will be based on observed data and will include the fixed, categorical effect of visit, as well as the continuous, fixed covariates of baseline VAS score and baseline-by-visit interaction (if significant). An unstructured covariance matrix will be used to model the within-subject errors. If this model does not converge, a compound symmetry covariance matrix will be used. Least-square mean change from baseline and corresponding 95% confidence intervals will be reported.

A t-test will be used to determine if baseline VAS scores differ by Afrin use. Afrin use, to visualize the turbinate contribution to overall obstruction, will be used considered at baseline only.

Results from the subject satisfaction questionnaire, including a cosmesis assessment, will be tabulated at 6 and 12 months. These items will be summarized individually:

- Change in Nose Appearance: Worse, No Change, Better (collapsed from 5 categories)
- Subject Would Recommend Procedure
- Subject Would Repeat Procedure

4.5 EXPLORATORY EFFICACY ENDPOINTS

Additional endpoints of interest include:

- Index procedure resource utilization: Anesthesia, procedure time and time to discharge.
- Follow up resource utilization: Nasal airway obstruction related return visits and medication utilization.
- Subject satisfaction questionnaire at 1, 3 and 12 months.
- Degree of nasal way obstruction as reported by subjects on the VAS scale at baseline with decongestant use.
- Endoscopic lateral wall insufficiency score per side21 at baseline and 6 months

- 3-D camera lateral wall motion assessment per side at baseline and 3, 6 and 12 months (at select sites).
- Cosmesis changes from baseline evaluated by Independent Photo Review 3 and 6 months (at select sites).
- Allergic rhinitis status at 1, 3, 6 and 12 months.
- Nasal geometry: length of nose, height and width of nose, skin thickness of lateral wall at baseline.
- Type of turbinate procedure.

4.6. Multiplicity Adjustment

All type-1 error will be allocated to the primary hypothesis, and there will be no adjustments for multiple testing.

5. SAFETY

All safety analyses and summaries will be based on the ITT and mITT analysis.

5.1. Adverse Events

Adverse events (AEs) will be coded. Adverse events through 7 days post procedure will be summarized for the ITT analysis set, and cumulative summaries will be produced for the mITT analysis set.

Categories of adverse events include:

- Device-related AEs
- Implant procedure-related AEs
- Serious AEs
- Adverse Device Effect (ADE)
- Device retrievals
- Moderate or Severe AEs

Summaries may also be produced by follow-up interval.

For each AE category, the percentage of subjects who experienced at least one occurrence of the given event will be calculated. Overall frequency by coded type will also be summarized. In addition, each AE will be listed for each subject. These listings will include onset interval, relationship to procedure and implant, severity, whether the event is an ADE, SADE or SAE, onset and resolution dates, action taken and outcome.

If the timing of events such as implant retrieval is of interest, this will be analyzed using Kaplan-Meier techniques.