ClariFix IM Study Statistical Analysis Plan

CT-0005-01 Rev A

Approved by:

Ellen O'Malley Pr. Clinical Research & Publications Specialist	Signature	11 Nov 2019	Date
Karen Peterson VP Clinical, Regulatory & Quality	Signature	11 Nov 2019	Date
April Slee Consultant Biostatistician	Signature	11 Nov 2019	Date

Revision History					
Rev	ECO	Release Date	Approval	Description	Training Requirements
A	N/A	11/11/2019	EMO	Initial Release	N/A

Stryker ENT

Statistical Analysis Plan

Feasibility Study of Multi-Treatment Posterior Nasal Nerve Modulation for Treatment of Chronic Rhinitis

(Clinical Investigational Plan CT-0005)

CT-0005-01

Revision A

November 2019

Stryker ENT

3600 Holly Lane N., Suite 40

Plymouth, MN 55447

Table of Contents

1 Abbreviations

AE	Adverse event
CGI-I	Clinician Global Impressions – Improvement
EDC	Electronic data collection
ENT	Ear, nose, throat
MCID	Minimal clinically important difference
Mini RQLQ	Mini Rhinoconjunctivitis Quality of Life Questionnaire
NOSE	Nasal Obstruction Symptom Evaluation
NPRS	Numeric pain rating scale
PRO	Patient-reported outcome
rTNSS	Reflective Total Nasal Symptom Score
SAE	Serious adverse event
SAP	Statistical analysis plan
SAS®	Statistical analysis software
SD	Standard deviation
SNOT-22	22-Item Sino-Nasal Outcome Test
VAS	Visual analog scale

2 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 study entitled, *"Feasibility Study of Multi-Treatment Posterior Nasal Nerve Modulation for Treatment of Chronic Rhinitis".* This is a prospective, multicenter, open-label, feasibility study (Arrinex Clinical Investigational Protocol #CT-0005). Participants will be consented and enrolled before study procedures. All participants who receive treatment will be assessed through the 90-day follow-up.

3 Study Design

3.1 Study objectives

This study will evaluate the feasibility of treatment to the posterior nasal nerve at both the middle and inferior meatus locations within the nasal cavity. The primary objective is to evaluate the safety of treatment to the inferior meatus location. Secondary objectives are to evaluate the effectiveness and tolerability of treatment at the inferior meatus location.

3.1.1 Study Endpoints

The primary endpoint is the incidence of procedure-related serious adverse events (SAEs) and/or serious adverse device effects (SADEs)

Secondary endpoints are:

- Incidence of procedure-related adverse events (AEs) and/or adverse device effects (ADEs)
- Tolerability of treatment as rated by participant verbal report of pain/discomfort on the 11-point Numeric Pain Rating Scale (NPRS)
- Change from baseline in nasal symptoms using the reflective Total Nasal Symptom Score (rTNSS) at 30 and 90 days post treatment
- Change from baseline in nasal congestion and turbinate hypertrophy as assessed by nasal endoscopy 30 and 90 days post treatment

Additional exploratory analyses are:

- Procedural characteristics
- Physician evaluation of ease of treatment delivery using study device for inferior meatus location
- Change from baseline in patient-reported outcome (PRO) assessments at 30 and 90 days post treatment:
    - Nasal Obstruction Symptom Evaluation (NOSE)
    - Sino-Nasal Outcomes Test (SNOT-22)
    - Nasal symptom visual analog scale (VAS)
    - Mini Rhinoconjunctivitis Quality of Life Questionnaire (mini RQLQ)

- Physician assessment of participant change from baseline as measured by the Clinical Global Impressions – Improvement

All statistical analyses will be performed using SAS (version 9.4), unless otherwise noted.

3.1.2 Statistical hypothesis

As a feasibility study, there is not a formal hypothesis, power calculation, or sample size estimates for the primary endpoint.

Statistical analysis of the secondary efficacy (rTNSS) endpoint is based on the following hypothesis:

H_0: Change from baseline = 0
H_a: Change from baseline ≠ 0

3.1.3 Sample size calculation and assumptions

A convenience sample of 30 treated participants was determined to be adequate for the study.

4 General Analysis Definitions

4.1 Study start and duration of follow-up

A participant is considered enrolled at the baseline visit once all enrollment criteria were confirmed to be met. Patients not meeting the enrollment criteria are considered screen failures and were not enrolled.

All participants will receive bilateral cryoablation treatment with the ClariFix device in accordance with the Instructions for Use (IFU). Each side of the nasal cavity is to be treated at both the middle meatus and the inferior meatus. Each location received approximately 30 seconds of cryoablation treatment. A second 30-second treatment was permitted at each location at the physician's discretion.

All participants will attend the following postprocedure study visits (window):

- 7-Day in-office visit (±3 days, AE and medication review only)
- 30-Day in-office visit (±7 days)
- 90-Day in-office visit (±14 days)

A listing of participants who discontinue before the 90-day visit will be provided in the clinical study report with the last visit completed and the reason for discontinuation. Possible reasons for early discontinuation include:

- Participant withdraws consent
- Participant is lost to follow-up
- Physician or sponsor withdraws participant for the health or welfare of the participant
- Participant death

4.2 Randomization

Not applicable.

4.3 Analysis sets

There are 2 analysis populations:

- The safety population includes all participants who receive ClariFix treatment
- The efficacy population includes all participants who receive ClariFix treatment and have least 1 valid follow-up assessment

4.4 Outcome variables

4.4.1 Efficacy outcome variables

The following outcome variables will be collected and analyzed as primary, secondary, or exploratory efficacy outcomes:

- Procedure pain scores (NPRS): during the procedure and immediately post procedure
- rTNSS: collected at baseline and at 30-day and 90-day follow-up visits
- Quantitative nasal endoscopy scores: collected at baseline, procedure, and at 30-day and 90-day follow-up visits
- NOSE: collected at baseline and at 30-day and 90-day follow-up visits
- SNOT-22: collected at baseline and at 30-day and 90-day follow-up visits
- Nasal symptom VAS: collected at baseline and at 30-day and 90-day follow-up visits
- Mini RQLQ: collected at baseline and at 30-day and 90-day follow-up visits

4.4.2 Safety outcome variables

- Serious adverse events and serious adverse device effects
- Nonserious adverse events and adverse device effects

4.4.3 Other variables

Other variables for analysis include procedure characteristics (See **Section 6**) and physician ease of device use.

5 Demographics and Baseline Characteristics

Demographics will include:

- Date of birth (age will be auto-calculated from the date of birth)
- Sex (male, female, not specified)
- Race (American Indian or Alaskan native, Asian, black or African American, native Hawaiian or other Pacific Islander, white or Caucasian, and other, specify)
- Ethnicity (Hispanic or Latino/not Hispanic or Latino)

Medical history will include allergy status, duration of rhinitis, and previous response to treatment with ipratropium bromide (IB). History of other medical conditions will include sinusitis, chronic nosebleeds/epistaxis, eye symptoms, facial pain, temporomandibular joint disorders (TMJ), hearing complaints, migraines, asthma, and gastroesophageal reflux disorder (GERD).

Previous ENT surgical procedures (eg, turbinate reduction, sinus surgery, septoplasty) will be noted. Concomitant medications will be collected.

Other baseline measures include the quantitative nasal endoscopy examination and PRO questionnaires (rTNSS, NOSE, SNOT-22, VAS, and mini RQLQ).

5.1 Participant disposition

A Study Exit form will be completed for all participants upon completion of all planned study requirements or early exit, as applicable. For those participants exiting before completing all the study requirements, the reason for the study exit will be noted. For follow-up visit compliance, once a participant has exited the study, they are no longer included in the denominator for future visits.

6 Procedure Data

Procedure data will include the date of treatment, treating physician, anesthesia medications and doses, and procedure times. For each treatment location (right and left, inferior and middle meatus), the freeze duration (seconds) and number of freezes will be collected. Up to 2 treatments are permitted per location. The number of devices used (and lot numbers) will be collected as will any reports of device malfunction.

7 Efficacy Outcomes

7.1 Efficacy analysis cohort

The efficacy analysis cohort is all participants treated with the ClariFix device who also have at least 1 valid follow-up assessment.

7.2 Efficacy outcomes

7.2.1 Procedure pain/discomfort

A secondary endpoint is the tolerability of the procedure. Current intensity of pain/discomfort is verbally provided by the participant during the procedure for each treatment location and overall immediately post procedure. Pain/discomfort is assessed using the 11-point NPRS where 0 indicates no pain/discomfort and 10 indicates the worst pain/discomfort imaginable.[1]

The count and percentage of participants indicating pain and the mean ± SD pain score will be calculated. The termination of any procedure due to pain will also be noted.

7.2.2 Reflective Total Nose Symptom Score (rTNSS)

A secondary efficacy endpoint is the mean change from baseline in the rTNSS at the 30-day and 90-day follow-up periods. The rTNSS is a PRO used to describe symptoms of rhinitis.[2] The assessment consists of 4 nasal symptom domains (runny nose [rhinorrhea], itchy nose, sneezing, and stuffiness [nasal congestion]). Each item is rated from 0 (absent) to 3 (severe). The 4 domains are added together to provide an overall score ranging from 0 to 12.

The mean ± SD for the change from baseline to the 30-day and 90- day follow-ups in the rTNSS will be calculated within subjects (paired observations). Significance of the change from baseline will be assessed using a paired *t*-test with (2-sided) 0.05 alpha level indicating significance. If the data are not normally distributed, a Wilcoxon signed rank test will be conducted in place of the *t*-test and the median and interquartile range will be presented instead of the mean ± SD. Normality will be assessed by the Shapiro-Wilk test, where a *p*-value <0.05 indicates the non-normality of the data.

Additional exploratory calculations may include:

- Count and percentage of participants with any improvement (≥1-point)
- Count and percentage of participants with rTNSS improvements of ≥25%, ≥30%, and ≥50%.
- Evaluation of individual rTNSS domain scores (rhinorrhea, congestion, itching, and sneezing)
- A subscale derived as the sum of the rhinorrhea and nasal congestion components

It is anticipated that changes in the nasal congestion and the runny nose scores of the rTNSS will be more impacted than the itching and sneezing scores and, therefore, will be of more interest for additional analysis.

7.2.3 Quantitative nasal endoscopy scores

Endoscopic nasal examinations (before and after decongestant administration) will be performed at baseline and at the 30-day and 90-day follow-up visits. Photos will be taken of the middle and inferior turbinates (before and after decongestant). After deidentification and randomization, an independent physician, blinded to the participant and study visit, will review and grade each photo for 3 findings (inferior turbinate hypertrophy, middle turbinate hypertrophy, and nasal congestion). Each finding will be rated on a scale from 0 (none) to 3 (severe).[3] Predecongestant and postdecongestant scores for each finding will be summarized using counts and percentages at baseline and the at the 30-day and 90-day visits.

7.3 Exploratory efficacy outcomes

The following measures will be collected and analyzed as exploratory efficacy outcomes.

7.3.1 Nasal Obstruction Symptom Evaluation

The NOSE scale is a validated PRO instrument that allows participants to quantify their perception of nasal obstruction.[4] The tool measures the severity of 5 symptoms (nasal congestion

or stuffiness, nasal blockage or obstruction, trouble breathing through the nose, trouble sleeping, and difficulty breathing through the nose during exercise or exertion) on a Likert scale from 0 (no problem) to 4 (severe problem). The sum of the 5 symptom scores is multiplied by 5 to give a score range from 0 to 100.

Lipan and Most developed a clinically relevant classification system for NOSE scores: mild (5-25 points), moderate (30-50 points), severe (55-75 points), or extreme (80-100 points).[5] This system has been used to define treatment responders. In previous clinical studies, responders were defined as participants who had at least 1 NOSE class improvement or a NOSE score reduction of at least 20% from baseline.

The mean ± SD for the change from baseline to the 30-day and 90- day follow-ups in the total NOSE score will be calculated for matched pairs. Significance of the change from baseline will be assessed using a paired *t*-test with (2-sided) 0.05 alpha level indicating significance. If the data are not normally distributed, a Wilcoxon signed rank test will be conducted in place of the *t*-test and the median and interquartile range will be presented instead of the mean ± SD. Normality will be assessed by the Shapiro-Wilk test, where a *p*-value <0.05 indicates the non-normality of the data.

The count and percentage of participants who meet the NOSE responder definition will be calculated using the previously used definition of NOSE responder (see above). Individual NOSE symptom scores may also be analyzed.

7.3.2 Sino-Nasal Outcomes Test (SNOT-22)

The SNOT-22 is a validated PRO consisting of 22 items, each item is scored using a 5-point Likert scale from 0 (no problem) to 5 (problem as bad as it can be) and is assess over the previous 2-week period. The total SNOT-22 score is the sum of the responses, producing a total possible score from 0 to 110, with higher scores indicating worse symptoms. The minimal clinically important difference (MCID) for the total SNOT-22 is –8.9 points.[6]

The mean ± SD for the change from baseline to the 30-day and 90- day follow-ups in the total SNOT-22 score will be calculated for matched pairs. Significance of the change from baseline will be assessed using a paired *t*-test with (2-sided) 0.05 alpha level indicating significance. If the data are not normally distributed, a Wilcoxon signed rank test will be conducted in place of the *t*-test and the median and interquartile range will be presented instead of the mean ± SD. Normality will be assessed by the Shapiro-Wilk test, where a *p*-value <0.05 indicates the non-normality of the data. Additionally, the count and percentage of participants who achieve the MCID (≥8.9 points) will be calculated.

7.3.3 Nasal symptom visual analog scale (VAS)

A 100-mm VAS is used to evaluate the symptoms rhinorrhea (runny nose), nasal congestion (stuffiness), and overall nasal symptoms. A response of 0-mm (left side) indicates no symptoms and a response of 100-mm (right side) indicates severe symptoms. At baseline and at the 30-day

and 90-day visits, participants marked on the continuum where they ranked their symptoms over the preceding week.

The mean ± SD for the change from baseline to the 30-day and 90- day follow-ups in the individual symptom VAS score and the overall VAS score will be calculated for matched pairs. Significance of the change from baseline will be assessed using a paired *t*-test with (2-sided) 0.05 alpha level indicating significance. If the data are not normally distributed, a Wilcoxon signed rank test will be conducted in place of the *t*-test and the median and interquartile range will be presented instead of the mean ± SD. Normality will be assessed by the Shapiro-Wilk test, where a *p*-value <0.05 indicates the non-normality of the data.

7.3.4 Mini Rhinoconjunctivitis Quality of Life Questionnaire – (mini RQLQ)

The mini RQLQ is a validated PRO that measures functional impairments due to allergic or non-allergic rhinoconjunctivitis.[7,] The assessment consists of 14 questions related to 5 domains of rhinoconjunctivitis: nasal symptoms, eye symptoms, other symptoms, practical problems, and activities. Each of the 14 items is scored from 0 (no impairment) to 6 (severely impaired). An overall score is calculated from the mean of the 14 item responses. Domain scores are the mean of the item scores within that domain. The MCID for the mini RQLQ has been established as 0.7.[8]

The mean change from baseline will be accessed for the mini RQLQ at the 30-day and 90-day follow-ups. The significance of the change will be assessed using a paired *t*-test with a (2-sided) 0.05 alpha level indicating significance. If the data are not normally distributed, a Wilcoxon signed rank test will be conducted in place of the *t*-test and the median and interquartile range will be presented instead of the mean ± SD. Normality will be assessed by the Shapiro-Wilk test, where a *p*-value <0.05 indicates the non-normality of the data. The count and percentage of participants achieving the MCID will be calculated on the total mini RQLQ score.

7.3.5 Clinician Global Impression of Improvement – (CGI-I)

The Clinical Global Impression--Improvement (CGI-I) is a clinician-completed assessment evaluating the clinician's impression of a patient's response to treatment based on their clinical experience.[9] The CGI-I is a 7-point Likert scale that ranges from 1 (very much improved) to 7 (very much worse). The CGI-I is completed by the physicians at the 30-day and 90- day follow-up visits to evaluate the physicians' impression of participant improvement over baseline.

The count and percentage of participants in each category will be calculated at the 30-day and 90-day follow-up visits. In addition, the count and percent of participants experiencing any improvement (very much improved, much improved, and minimally improved) at each visit will be calculated.

8 Safety Outcomes

8.1 Safety analysis cohort

Safety analysis cohort is all participants treated with the ClariFix device.

8.2 Safety outcomes

All AEs (serious and nonserious) will be adjudicated by the Medical Monitor for the following:

- AE term
- Relatedness (to study device and procedure)

The relatedness will be initially determined by the investigator; however, if there is a discrepancy between the investigator and the Medical Monitor, the final analysis will be based on the adjudication determination. Adjudication findings will be entered into the Medrio EDC.

8.2.1 Serious adverse events

The primary safety endpoint is the number of participants with 1 or more SAEs and/or SADEs. SADEs are SAEs that are adjudicated to be related (possibly, probably, or definitely) to the ClariFix device and/or procedure. The total number of SAEs will also be presented.

8.2.2 All device-related/procedure-related adverse events

The secondary safety endpoint is the number of participants with 1 or more AE and/or ADE. ADEs are AEs that are adjudicated to be related (possibly, probably, or definitely) to the ClariFix device and/or procedure. The total number of AEs and ADEs will also be presented.

A listing by participant of all SAEs, SADEs, and ADEs will be provided.

9 Protocol Deviations

Protocol deviations will be summarized by deviation type, and overall using counts and percentages. A listing of all protocol deviations by participant will also be provided.

10 Additional Statistical Analyses

10.1 Study termination criteria

There are no preset criteria for study termination of this study.

10.2 Pooling assessment

Due to the small number of investigative centers (≤5), no pooling analysis is planned.

10.3 Blinding assessment

Not applicable.

10.4 Sensitivity and missing data analyses

Due to the small sample size and short-term follow-up, no sensitivity or missing data analyses are planned.

10.5 Multiplicity adjustments

No multiplicity adjustments are planned.

10.6 Planned subgroup analyses

There are no planned subgroup analyses.

10.7 Interim analyses

No control of overall type 1 error was planned and no adjustments will be made for multiple outcomes.

11 Changes from Original Statistical Plan

This is the initial release of the statistical analysis plan, so there are no changes.

12 References

[1] Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. *Ann Rheum Dis*. 1978;37:378-381.

[2] Downie SR, Andersson M, Rimmer J, et al. Symptoms of persistant allergic rhinitis during a full calendar year in house dust mite-sensitive subjects. *Allergy*. 2004;59:406-414.

[3] Camacho M, Zaghi S, Certal V, et al. Inferior turbinate classification system, grades 1 to 4: development and validation study. *Laryngoscope*. 2015;125:296-302.

[4] Stewart MG, Witsell DL, Smith TL, Weaver EM, Yueh B, Hannley MT. Development and validation of the Nasal Obstruction Symptom Evaluation (NOSE) scale. *Otolaryngol Head Neck Surg*. 2004;130:157-163.

[5] Lipan MJ, Most SP. Development of a severity classification system for subjective nasal obstruction. *JAMA Facial Plast Surg*. 2013;15:358-361.

[6] Hopkins C, Gillett S, Slack R, Lund VJ, Browne JP. Psychometric validity of the 22-item Sinonasal Outcome Test. *Clin Otolaryngol.* 2009;34:447-454.

[7] Juniper EF, Thompson AK, Ferrie PJ, Roberts JN. Development and validation of the mini Rhinoconjunctivitis Quality of Life Questionnaire. *Clin Exp Allergy*. 2000;30:132-140.

[8] Pfaar O, Demoly P, van Wijk RG, et al. Recommendations for the standardization of clinical outcomes used in allergen immunotherapy trials for allergic rhinoconjunctivitis: an EAACI position paper. *Eur J Allergy Clin Immunol*. 2014;69:854-867.

[9] Guy W. Clinical Global Impressions. In: *ECDEU Assessment Manual for Psychopharmacology—Revised*. Rockville, MD: U.S. Department of Health, Education, and Welfare; Public Health Service, Alcohol; Drug Abuse, and Mental Health Administration; National Institute of Mental Health; Psychopharmacology Research Branch; Division of Extramural Research Programs. 1976:218–222. OCLC 2344751. DHEW Publ No ADM 76–338.