The Vivaer® Procedure for Treatment of Nasal Airway Obstruction – A ProspecTive, Multicenter Randomized Controlled TriAl Comparing Vivaer to Sham Control (VATRAC)

Statistical Analysis Plan

Study: The Vivaer® Procedure for Treatment of Nasal Airway Obstruction – A Prospective, Multicenter Randomized Controlled Trial Comparing Vivaer to Sham Control (VATRAC)
(Protocol #CTP1006 / Rev C 15 September 2020)

Version: Rev A, 13 April 2022

Product: Vivaer® ARC Stylus

Sponsor: Aerin Medical, Inc.
232 E. Caribbean Drive
Sunnyvale, CA 94089 USA

Contact: Anais Laborde, Senior Director of Clinical Affairs
Phone: 650-518-9624
Email: alaborde@aerinmedical.com

Prepared by: Jeff F. Doerzbacher – Consulting Biostatistician

Plan Approvals

Name	Signature	Date
Desiree Hollemon, MSN, MPH Vice President Clinical Affairs Aerin Medical	Desiree Hollemon	13APR2022
Anais Laborde Senior Director, Clinical Affairs Aerin Medical	Anais Laborde	13APR2022
Jeff F. Doerzbacher Consultant Biostatistician	Jeff F. Doerzbacher	13APR2022

TABLE OF CONTENTS

1 ABBREVIATIONS

AE – Adverse event

CSR – Clinical study report

ESS – Epworth Sleepiness Scale

FAS – Full Analysis Set

NAO – Nasal airway obstruction

NOSE – Nasal Obstruction Symptom Evaluation

PPS– Per Protocol Set

PRN – as needed

RF - radiofrequency

SAP – Statistical analysis plan

SAE – Serious adverse event

VAS – visual analog scale

2 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to prespecify statistical analysis methods for supporting the completion of the publications and clinical study report (CSR) of CTP1006 (VATRAC) for the Vivaer product.

This SAP outlines the necessary data summaries, statistical techniques, and analyses required to meet the long-term follow-up study objectives. This plan is intended to supplement the analysis plan provided in the study protocol and provide additional detail regarding follow-up analyses beyond the primary endpoints at the 3-month evaluation, along with inclusion of control participants who received the Vivaer treatment after the 3-month primary endpoint (crossovers).

Results of analyses performed under this SAP may be included in regulatory submissions, clinical study reports and/or manuscripts. Exploratory analyses, which are not defined in this SAP, may be performed to support the clinical development program. Any post-hoc or unplanned analyses, which are performed, but not defined in this SAP, will be detailed in an amendment to this plan and will be documented in the CSR.

3 STUDY DESIGN

This was a prospective, multicenter, randomized, sham-controlled superiority trial to compare the Vivaer procedure for treatment of nasal airway obstruction (NAO) with a sham procedure that simulates the actual procedure as closely as possible absent the delivery of radiofrequency (RF) energy to the nasal tissue. A 2:1 site-stratified randomization was used to allocate participants with NAO to either the Vivaer procedure (active treatment) or a sham procedure (control). A one-way crossover component after the 3-month primary endpoint evaluation was available to participants randomized to the control arm

All participants were evaluated in-office prior to treatment and following treatment at week 4 (1 month) and week 13 (3 months). The 3-month evaluation conducted after the index VivAer and sham procedures were used for the primary endpoint analysis.

Participants randomized to the Vivaer procedure arm had extended follow-up evaluations conducted in-office at 6 months (26 weeks) and remotely at 12 months (52 weeks) and 24 months (104 weeks).

Participants randomized to the sham arm could elect to crossover to the active treatment arm (Vivaer procedure) within 30 days after the initial 3-month follow-up evaluation provided they still meet all eligibility criteria. Continued follow-up will be conducted at 1, 3, 6, 12, and 24 months after the Vivaer procedure to provide additional information on longer-term efficacy and duration of treatment effect. Participants that received the sham procedure and did not elect to cross over or no longer met all eligibility criteria were exited from the study following the 3-month evaluation.

Participants in either arm of the study that received a subsequent procedure or surgery after the Vivaer procedure were also exited from the study.

4 STUDY OBJECTIVES

The purpose of this trial is to compare the Vivaer procedure for treatment of NAO with a sham procedure that simulates the actual procedure as closely as possible without the delivery of RF energy to the nasal tissue. The primary objective is to assess the performance of the Vivaer procedure compared to a sham control procedure with respect to individual participant success rates when used as a treatment for NAO. Additional objectives include assessment and comparison of secondary and informational outcome measures between the groups receiving the Vivaer procedure and the sham procedure.

5 STUDY ENDPOINTS

The primary objective was assessed through evaluation of the primary endpoint defined as the responder rate at 3 months for randomized patients receiving the VivAer and sham procedures. Individual participant success ("responder") is defined as at least 1 Nasal Obstruction Symptom Evaluation (NOSE) Scale class improvement or an improvement (decrease) in NOSE Scale score of 20% or more from baseline at the 3-month evaluation.

Secondary endpoints are:

- Mean change in NOSE Scale scores from baseline to 3 months after procedure.
- Frequency of device-related and procedure-related serious adverse events through the 3-month evaluation.

Other Endpoints include:

- Adverse events - Incidence (type, severity, relatedness) of adverse events overall and by follow-up interval.
- Nasal Assessment – Visual assessment of the target nasal valve area within each nostril ocurring at baseline, just prior to procedure (if screening and procedure occur on different days), immediately after procedure, at 1 month, 3 months and 6 months after procedure.
- NOSE Scale score:
    - Mean and mean change from baseline at the 3-, 6-, 12-, and 24-month follow-up evaluations.
    - Distribution of NOSE Scale score severity categories (mild, moderate, severe, extreme) at the 3-, 6-, 12-, and 24-month follow-up evaluations.
    - Mean, change from baseline in mean, and response distribution of the 5 components of the NOSE Scale score (nasal congestion, nasal blockage, trouble breathing, trouble sleeping, and getting enough air during exercise) at the 3-, 6-, 12-, and 24-month follow-up evaluations.
    - Proportion of responders based on improvement in NOSE Scale score at the 3-, 6-, 12-, and 24-month follow-up evaluations.

- Visual analog scale (VAS) for pain - assessed immediately after procedure, and at 1-month and 3-month follow-up evaluations.
- Visual analog scale (VAS) for ease of breathing - assessed at baseline, 1-month, 3-month and 6-month follow-up evaluations.
- Epworth Sleepiness Scale (ESS) - mean and change from baseline administered at 1-, 3-, 6-, 12-, and 24-month follow-up evaluations.
- Participant Satisfaction Assessment - assessed at the 3-, 6-, 12-, and 24-month follow-up evaluations.
- Change in amount of "as needed" (PRN) medication/device use for nasal obstruction symptoms - Self-reported assessment of an increase, no change, or decrease in as needed medications and/or devices being used for treatment of nasal symptoms at each evaluation compared to use prior to the procedure administered at the 1-, 3-, 6-, 12-, and 24-month follow-up evaluations.
- Medications - Medications associated with relief or treatment of nasal airway obstruction symptoms will be documented at baseline and updated as necessary at each evaluation. In addition, medications associated with treatment of adverse events will be documented.

6 SAFETY AND RISK PROFILE

The safety and risk profile of the Vivaer procedure relative to the sham procedure will be evaluated with respect to overall incidence of adverse events and procedure/device-related adverse events.

7 SAMPLE SIZE CALCULATION AND ASSUMPTIONS

The study will enroll up to 120 participants. The power analysis was performed such that there is adequate power to reject the null hypothesis of no difference between treatments. Sample size estimation was based on the test for differences between two proportions[1] using an Exact test and on the following assumptions:

- Significance level $\alpha = 0.05$ (two-sided)
- Power = 80%
- $\pi_C = 0.50$ (allowance for a 50% control responders)
- $\pi_T = 0.80$ (assumed 80% responders in the treated group)
- Treatment allocation is 2:1
- Estimated 10% dropout/non-evaluable patients

The 80% responder rate assumed for the active Vivaer procedure group is a conservative estimate based on the prior clinical study of this procedure in which all 50 participants were responders at 12 weeks and 46 of 49 (94%) were responders at 26 weeks. The lower 95% confidence bound on the estimate of 94% is 83.5%, which supports the conservative

estimate of 80% responders assumed for the purpose of sample size calculation. The 50% responder rate assumed for the sham control procedure is based on literature for placebo and sham controls in therapeutic and device studies suggesting from 30% to 60% responder rates, with device studies tending to be at the higher end of the range[2]. A randomized study of a bioabsorbable implant treatment for nasal valve collapse reported a 54.7% responder rate in the sham procedure arm of the study.[3]

The minimum number of participants to achieve 80% power with a 2:1 active treatment to sham control allocation is 66 in the active treatment group and 33 in the sham. The sample size, allowing for 10% loss (non-evaluable) in each group and adjusting for a balanced distribution across 20 sites is 120 participants (80 active treatment (Vivaer procedure), 40 sham procedure control).

It is anticipated that participants will be enrolled at sites on a competitive basis; however, a reasonable balance of participants among sites may be maintained by potentially capping enrollment at individual sites based on the final number of participating sites.

8 RANDOMIZATION

Randomization occurred after the participant was enrolled, met all eligibility criteria, and prior to the treatment procedure allowing sufficient time to prepare the required treatment. The randomized assignment (active treatment arm or control arm) was determined using a web-based service (Medrio, San Francisco CA). Participants were assigned to the active treatment arm (Vivaer procedure) or the control arm (sham procedure) in a 2:1 allocation, using the computer-generated block randomization scheme stratified by site.

9 BLINDING

It was not possible to blind the investigator administering the treatment because of obvious differences in the active application of the RF energy versus the sham control. Patients receiving sham treatment were blindfolded, the RF stylus was applied in the nose in a manner identical to the actual procedure and sound mimicking treatment was played but no RF energy delivered. Efforts were made to keep the participant blinded as to the treatment received through the 3-month primary endpoint evaluation. Following the 3-month primary endpoint evaluation, participants were unblinded and offered the cross-over procedure if they met all eligibility criteria.

10 OUTCOME MEASURES

10.1 Nasal Assessment

The target nasal valve area within each nostril was visually assessed at baseline, just prior to procedure (if screening and procedure occur on different days), immediately after procedure, at 1 month, 3 months and 6 months after the procedure. The use of an endoscope for visual assessment was required. Observations were categorized as not present, mild, moderate, or severe. Representative video of each nasal passage was captured for each assessment.

Endoscopic video of participant holding their breath and during maximal inhalation was used to capture dynamic changes.

Assessments include:

- Saddle nose deformity
- Bruising around orbital area
- Soreness, pain
- Numbness.

Endoscope required for assessment of:

- Inflammation / generalized redness
- Swelling, edema
- Blanching (generalized whiteness)
- Bleeding at anesthetic injection site (not requiring physician intervention)
- Bleeding at treatment site (not requiring physician intervention)
- Nasal obstruction from tissue edema
- Disruption of mucosal flow / crusting.

10.2 NOSE Scale score

To evaluate the significance of a patient's nasal obstruction both before and after the procedure, this study used the well-known subjective patient-reported Nasal Obstruction Symptom Evaluation (NOSE) Scale for determining the primary endpoint of the study. The NOSE Scale is a validated disease-specific health status instrument used by clinicians to measure the outcome of patients treated for nasal obstruction.[4] The NOSE Scale consists of 5 items, each scored using a 5-point Likert scale to make a total score range of 0 through 100, where higher scores indicate worse obstruction. Severity of symptoms can be classified as mild (range, 5-25), moderate (range, 30-50), severe (range, 55-75), or extreme (range, 80-100) nasal obstruction, based on responses to the NOSE survey.[5]

Treatment responder based on NOSE Scale improvement

Individual participant success (responder) is defined as at least 1 Nasal Obstruction Symptom Evaluation (NOSE) Scale class improvement, (eg, going from a score in the severe range (55-75) at baseline to a score in the moderate range (30-50) at the follow-up evaluation), or an improvement (decrease) in NOSE Scale score of 20% or more from baseline to the follow-up time point.

10.3 Visual analog scale (VAS) for pain

Perception of pain associated with the procedure was measured using a horizontal 100 mm visual analog scale (VAS)[6] anchored on the left with the words "No Pain" (0) and on the right with the words "Worst Pain Imaginable" (10), was used to measure nasal pain associated with the procedure. Scores are obtained by measuring the distance in millimeters from the left origin of the line (0) to the point indicated with a vertical slash placed by the participant to indicate their current level of pain in and around the nose

10.4 Visual analog scale (VAS) for ease of breathing

The ability to breathe through the nose was measured using a horizontal 100 mm VAS anchored on the left with the words "No Difficulty Breathing" (0) and on the right with the words "Extreme Difficulty Breathing" (10), was used to assess the perception of the ability to breathe through the nose. Scores are obtained by measuring the distance in millimeters from the left origin of the line (0) to the point indicated with a vertical slash placed by the participant to indicate their assessment of ease of breathing.

10.5 Epworth Sleepiness Scale (ESS)

The ESS evaluates the self-reported likelihood of dozing or falling asleep in 8 daytime situations with likelihood rated No Chance (0), Slight (1), Moderate (2), or High Chance (3).[7] The total of the scores across the 8 questions ranges from 0 to 24 and can be categorized and interpreted as:

- 0-5 Lower normal daytime sleepiness
- 6-10 Higher normal daytime sleepiness
- 11-12 Mild excessive daytime sleepiness
- 13-15 Moderate excessive daytime sleepiness
- 16-24 Severe excessive daytime sleepiness.[8]

This measure is included in the study because results from at least 1 study suggest that chronic nasal obstruction impairs quality of life, at least partially, through excessive daytime sleepiness possibly caused by sleep-disordered breathing.**Error! Bookmark not defined.** Treatment effectiveness with the Vivaer procedure may be evident in improvement of the ESS score.

10.6 Participant Satisfaction Assessment

A 5-question self-reported survey of satisfaction using a 5-point scale was used to assess tolerability of the procedure, ease of recovery, breathing through nose, overall satisfaction with the procedure, and recommendation to others.

10.7 Change in amount of PRN medication/device use for NAO symptoms

Self-reported assessment of an increase, no change, or decrease from baseline in as needed medications and/or devices being used for treatment of nasal symptoms following the procedure. Medication/device use may be categorized for reporting purposes.

10.8 Medications

The current use of medication, devices or other therapies for symptoms of nasal obstruction, medication name, frequency, and dose were recorded at each evaluation visit. Medications may be categorized for reporting purposes.

10.9 Safety – Adverse Events

Incidence (type, severity, relatedness) of adverse events overall and by follow-up interval were recorded according to definitions provided in the study protocol.

11 STATISTICAL ANALYSIS

11.1 Analysis Populations

Full Analysis Set (FAS):
The FAS is comprised of all enrolled participants that have a baseline evaluation, were randomized, underwent the study procedure, and have the 3-month primary endpoint evaluation.

Per Protocol Analysis Set (PPS):
The PPS is comprised of all participants who met eligibility criteria, have at least a baseline and 3-month primary endpoint evaluation, were properly randomized and underwent the assigned study procedure and follow-up procedures with no exclusionary protocol deviations. Exclusionary protocol deviations will be adjudicated prior to data analysis. The effects of protocol deviations will only be examined when the validity of the study conclusion is in question.

Safety Analysis Set:
The safety analysis population is comprised of all enrolled patients that have received any study-specific procedure, including screening procedures (eg, nasal endoscopy) and any part of the Vivaer or sham study procedure, including anesthesia.

11.2 Treatment Groups

Index Active Treatment Group:
Participants randomized into the Active Treatment Arm

Index Sham Treatment Group:
Participants randomized into the Sham Arm

Crossover Active Treatment Group:
Participants originally randomized to the Sham Arm that completed the study to the 3-month primary endpoint that are requalified according to the original inclusion and exclusion criteria and receive active treatment with the Vivaer device.

Combined Active Treatment Group:
Participants who have received active treatment with the Vivaer device regardless of timing of the procedure. This will be the Index Active Treatment Group and the Crossover Active Treatment Group combined.

11.3 General Considerations

The primary and secondary endpoint analyses were compariosns of the Index Active Treatment with the Index Sham Treatment Group using the FAS. Other endpoint analyses will also generally use the FAS. Additional analyses using the PPS will be noted. If a participant is discontinued from the study due to receiving an additional nasal procedure (e.g., septoplasty, turbinate reduction), data from follow-up visits completed prior to the additional procedure will be included in the analysis; however, data collected after an additional procedure but prior to discontinuation will not be included in the analysis.

Summary descriptive statistics, including mean, median, standard deviation, minimum, maximum, and interquartile range for continuous measures, and frequencies and percentages for categorical outcomes will be presented for all variables of interest. Outcome measures (primary, secondary, and informational) will be presented by treatment group and time. For most summary statistics, data will be analyzed and displayed by treatment groups in the following order: Index Active Treatment and Index Sham Treatment for the primary endpoint; Index Active Treatment, Crossover Active Treatment, and Combined Active Treatment for time points beyond the primary endpoint. Confidence intervals will be included for means of the various outcome measures.

Informational and exploratory outcomes analyses will use $P<.05$ as a measure of statistical significance. Where feasible, longitudinal analyses of individual outcome measures will incorporate *P*-value adjustment for comparisons of multiple time points. No adjustments will be made for due to analyzing multiple outcome measures.

Analyses using repeated measures linear mixed models with multiple comparions of time points are anticipated for outcome measures conventionally considered continuous data (NOSE, ESS). Components of the NOSE Scale score have also been treated as continuous but may also be anlayzed using generalized linear models as outcomes represented by ordered categories.

Results will be presented using least square (adjusted) means where data are not complete across all time points.

Nonparametric analysis will be employed where gross violations of assumptions for parametric analyses are demonstrated.

11.4 Data Pooling

All study data will be pooled across study sites to facilitate hypothesis testing in accordance with the sample size estimation and power analysis (Section 7). Comparability between study sites may be shown using summary statistics calculated by site.

11.5 Missing Values

All missing data in each treatment group will be imputed for the primary endpoint by assuming the missing outcomes are nonresponders. In addition, a post hoc sensitivity analyses may be performed, including a worst-case analysis (all missing primary outcomes in the Vivaer procedure group are considered nonresponders and all missing outcomes in the sham procedure control group are responders) and/or a change-point analysis, to assess the effect of missing data on the primary analysis. The results of the sensitivity analysis will not be used to adjust the conclusions drawn from the primary analysis.

Secondary outcome measures and additional observational measurements will be analyzed by using available data only

11.6 Data Collection and Quality Assurance

Data will be collected on electronic case report forms using the Medrio electronic data capture system and database. Qualified monitors will review the data electronically and through site monitoring visits to assure that the investigator and staff are compliant with this protocol and applicable regulatory requirements in addition to assuring timely and accurate data collection.

11.7 Statistical Software

SAS 9.4 (SAS Institute, Cary, NC) will be used to create datasets from SAS data exports from the Medrio system for reporting and analysis. Analyses will be performed using SAS 9.4 (SAS Institute, Cary, NC) and SigmaPlot 14.0 (Systat Software, Inc., San Jose, CA).

11.8 Demographics and Baseline Characteristics

Demography, baseline characteristics and the comparability of active treatment group and sham control group participants will be summarized using frequencies and percentages for categorical factors and mean, median, standard deviation, minimum and maximum for continuous factors. Demographic characteristics will be reported to describe the profile of samples.

11.9 Outcome Measures Analyses

Nasal Assessment - The visual physical and endoscopic assessment factors will be summarized to include frequency and percentage of responses in each category for each component of the nasal assessment by treatment group at baseline, just prior to procedure (if screening and procedure occur on different days), immediately after procedure, at 1 month, 3 months, and at 6 months after the procedure.

NOSE Scale score - Categorical responses and scores on the NOSE Scale and its individual components will be subject to multiple summary methods and analyses including the:

- Mean and mean change from baseline at the 3-, 6-, 12-, and 24-month follow-up evaluations.
- Distribution of NOSE Scale score severity categories (mild, moderate, severe, extreme) at the 3-, 6-, 12-, and 24-month follow-up evaluations for each group will be displayed as number and percent.
- Mean, change from baseline in mean, and response distribution of the 5 components of the NOSE Scale score (nasal congestion, nasal blockage, trouble breathing, trouble sleeping, and getting enough air during exercise) at the 3-, 6-, 12-, and 24-month follow-up evaluations for each group.
- Proportion of responders based on improvement in NOSE Scale score at the 3-, 6-, 12-, and 24-month follow-up evaluations for each group will be displayed as number and percent.

Visual analog scale (VAS) for pain - Summary will include mean VAS pain scores assessed postprocedure, 1 month, and at 3 months for both groups.

Visual analog scale (VAS) for ease of breathing - Summary will include mean ease of breathing scores assessed at baseline, 1 month, 3 months, and at 6 months for both groups.

Epworth Sleepiness Scale (ESS) - mean and change from baseline in mean ESS score and categorical responses will be summarized for both groups at baseline and the 1-, 3-, 6-, 12-, and 24-month follow-up evaluations.

Participant Satisfaction Assessment - mean response for each of the five survey questions will be summarized by group at the 3-, 6-, 12-, and 24-month follow-up evaluations.

Change in amount of PRN medication/device use for nasal obstruction symptoms - Proportions of participants reporting increase, decrease, or no change in use for the categories of PRN medications or devices prior to procedure will be summarized by group (e.g., antihistamine, intranasal steroid, decongestant, anticholinergic, immunotherapy) the 1-, 3-, 6-, 12-, and 24-month follow-up evaluations.

Medications - Medications associated with relief or treatment of nasal airway obstruction symptoms or associated with the treatment of adverse events will be provided by group and documented at baseline and updated as necessary at each evaluation. Medication use by categories may also be presented as percentages of participants in each group.

11.10 Subgroup and Other Analyses and Summaries

Several subgroups of interest have been identified for further exploration in relation to the effect on NOSE Scale scores:

- valve collapse mechanism (dynamic, static, mixture)
- prior septoplasty or turbinate surgery
- presence/absence of septal deviation
- presence/absence of enlarged turbinates.

Other areas of interest include:

- Epworth Sleepiness Scale analysis for scores >10 and ≤ 10
- Characteristics of treatment nonresponders.

11.11 Safety Analysis

All adverse events will be analyzed for all participants. Participants will be included in the treatment group corresponding to the actual treatment received. Adverse events will be coded using a custom Aerin Medical dictionary so that adverse events may be categorized for analysis at an appropriate level of detail. Listings will be provided to detail individual events. The number of participants, number of AEs, and the proportion of participants reporting each AE will be summarized. Multiple occurrences of the same AE for a participant will be noted but only counted once for reporting of AE incidence. Seriousness and severity of AEs and their

relationship to the device and procedure will be summarized. A time course of adverse events will be presented. Any unanticipated adverse device experiences or adverse events that occur at an unexpectedly high incidence rate will receive detailed analyses. Narratives will be presented for all deaths, serious adverse events, unexpected adverse device experiences, and participants withdrawn due to an adverse event.

12 REFERENCES

1. Fleiss JL. *Statistical Methods for Rates and Proportions*, 2nd Edition. Wiley & Sons. 1991.
2. Kaptchuk TJ, Goldman P, Stone DA, Stason WB. Do medical devices have enhanced placebo effects? *J Clin Epidemiol* 2000;53(8):786-792.
3. Stolovitzky P, Senior B, Ow RA, et al. Assessment of bioabsorbable implant treatment for nasal valve collapse compared to a sham group: a randomized control trial. *Int Forum Allergy Rhinol* 2019;9(8):850-856.
4. Stewart MG, Witsell DL, Smith TL, et al. Development and validation of the Nasal Obstruction Symptom Evaluation (NOSE) scale. *Otolaryngol Head Neck Surg* 2004;130(2):157-163.
5. Lipan MJ, Most SP. Development of a severity classification system for subjective nasal obstruction. *JAMA Facial Plast Surg* 2013;15(5):358-361
6. Scott J, Huskisson EC. Graphic representation of pain. *Pain* 1976;2:175-184.
7. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991;14(6):540-545.
8. Johns MW. The Epworth Sleepiness Scale. https://epworthsleepinessscale.com/about-the-ess/. Accessed March 03, 2022.