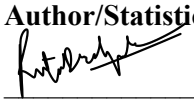
STATISTICAL ANALYSIS PLAN

PROTOCOL NAME:	A study of safety and efficacy of Transcutaneous Afferent Patterned Stimulation (TAPS) for reduction of action tremor in Parkinson's Disease (PD)
PROTOCOL NO:	PD-02
VERSION:	B
DATE:	25 March 2021
SPONSOR:	Cala Health Inc. 875 Mahler Rd, Suite #168 Burlingame, CA 94010

STATISTICAL ANALYSIS PLAN APPROVAL

The signatures below indicate the review and approval of the Statistical Analysis Plan for PD-02

Author/Statistician

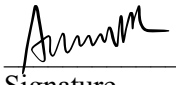


Signature — March 29, 2021 — Date

Ruta Deshpande — Name

Clinical Research Engineer — Title

Research Science

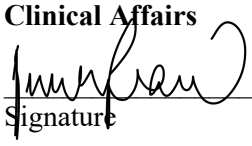


Signature — March 28 2021 — Date

Apoorva Rajagopal — Name

Sr. Research Scientist — Title

Clinical Affairs

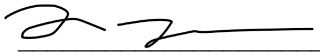


Signature — March 28 2021 — Date

Vivien Zraick — Name

Sr. Director, Clin Ops — Title

Data Manager



Signature — March 28, 2021 — Date

Felice Sun — Name

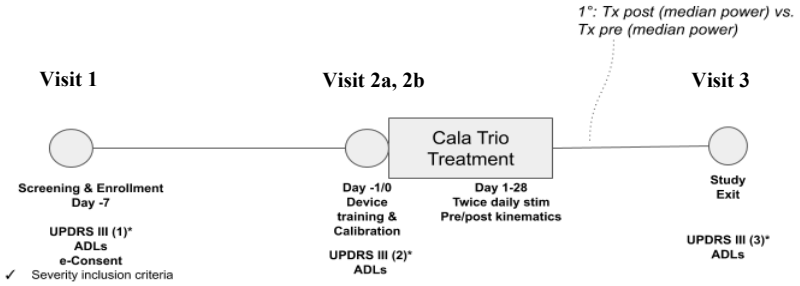
Data Manager — Title

1 TABLE OF CONTENTS

2 STUDY OVERVIEW

2.1 Design and Objective

This is a prospective, single-site, single-arm, non-significant risk study designed to evaluate the Cala Trio device for action tremor in PD. Subjects will be screened for eligibility and fitted with a Cala Trio device. Subjects will wear the device at home for a period of 4 weeks, during which they will be asked to stimulate their dominant hand twice daily. The stimulation amplitude will be based on each subject's stimulation threshold. Subjects will have in-clinic assessments at enrollment, as well as before and after treatment of 4 weeks. The study timeline, reproduced from the study protocol, is illustrated below. More detail on study data follows below.



All subjects will be informed of their right to withdraw from the clinical study at any time without penalty or loss of benefits. The Investigator may prematurely discontinue any subject from the study if the Investigator feels that the subject can no longer fully comply with the requirements of the study or if any of the study procedures are deemed potentially harmful to the subject.

The study objective is to evaluate symptomatic hand tremor relief in the treated hand following stimulation with the Cala Trio device in adults with action tremor in PD over a duration of 4 weeks.

a. Visit 1

Subjects will be reminded to be in the *off* state, i.e., to skip their last dose of PD medication, prior to their telemedicine Visit 1. The following assessments will occur:

- Screening activities including medication review and medical history. No medical records will be used for this study, so all eligibility criteria will be based upon self-reported information from each subject and verification by qualified study staff.
- Tremor assessments (while in the *off* state) including:
    - Modified MDS Unified Parkinson's Disease Rating Scale Part III (MDS-UPDRS) to confirm eligibility will be rated by the PI
    - Bain & Findley Activities of Daily Living (BF-ADL) upper limb subset score will be rated by the subject based on their recall of the previous week (no props will be used)
- Demographic information, Parkinson's clinical characteristics, and wrist size and handedness will be collected based on self-reported information from subjects.

Once a subject has been confirmed to be eligible for the study, a device kit containing a stimulator, base station and band will be shipped directly to the subject's home ahead of Visit 2.

b. Visit 2

Visit 2a

Once a subject has received the device kit, device training will occur via telemedicine, as follows:

Positioning and Fastening Cala Trio

Prior to device positioning, subjects should dampen the wrist of the prescribed hand to ensure good connection between the electrodes and wrist. Subjects should position the device on the prescribed hand such that the electrodes are positioned over the median and radial nerves, and securely tighten the band.

Device Calibration

To calibrate the Cala Trio device, subjects will perform a series of postural hold tremor tasks, in which the device measures the subject's tremor frequency and sets the stimulation burst frequency based on the measured tremor frequency.

For the purposes of this study, only the outstretched postural hold tremor task will be used.

Setting Stimulation Intensity

The subject will set the stimulation intensity to an amount they can comfortably tolerate while doing light activities. Paresthesia should be perceived in the parts of the hand and/or fingers innervated by the median and radial nerves. The stimulation intensity selected during the initial amplitude calibration is saved on the device for subsequent sessions. The subject can decrease or turn off the stimulation at any point in time and may adjust their stimulation intensity from session to session to maintain the minimum paresthesia threshold.

Device Training

During the course of the visit, study personnel will instruct the subject on proper use of the device. Subjects must demonstrate competence to operate the device and comprehension of the instructions for use prior to completing the training.

Visit 2b

Once a subject is trained on the device, the subject will receive a single 40-minute therapy session with their Cala Trio. The subject will be asked to not stimulate at least 8 hours prior to this visit. The following assessments will be made prior to the therapy session and again immediately post the therapy session:

- Tremor assessments (while in the *off* state) including:
  - Modified MDS Unified Parkinson's Disease Rating Scale Part III (MDS-UPDRS) will be rated by the PI
  - Bain & Findley Activities of Daily Living (BF-ADL) upper limb subset score will be rated by the subject. The BF-ADL assessment will be performed using the subject's own household props

- CGI-S, I/PGI-S, I
    - Clinical/Patient Global Impression of Severity (CGI-S/PGI-S): The investigator or designee (CGI) and Patient (PGI) will rate the overall tremor severity prior to stimulation
    - Clinical/Patient Global Impression of Improvement (CGI-I/PGI-I): The investigator or designee (CGI) and Patient (PGI) will rate the overall tremor improvement following stimulation

Additionally, the Cala Trio will direct the subject to perform the outstretched postural hold tremor tasks before and after the stimulation session to measure tremor power. Finally, subjects will input their Patient Session Impression of Improvement (PSI-I) rating on the device after the stimulation session

Visits 2a and 2b may be combined into a single Visit 2.

c. Home-Usage

The home use portion of the study will last 4 weeks. It will include:

- 2 stimulation sessions/day preferably at the same time every day recommended as one in the morning and one in the evening.
- Motion data from the device before and after each stimulation session. Cala Trio will direct the subject to perform postural hold tremor tasks before and after the stimulation session to measure tremor power. Additionally, subjects will input their Patient Session Impression of Improvement (PSI-I) rating on the device after the stimulation session
- Weekly Bain & Findley Activities of Daily Living (BF-ADL) upper limb subset score will be rated by the subject based on their recall of the previous week (no props will be used)
- Device factory reset of the Cala Trio will occur on Day 14 of treatment via a call with study staff

d. Visit 3/Study exit

Subjects will be reminded to be in the *off* state, i.e., to skip their last dose of PD medication, prior to their telemedicine Visit 3/Study exit. The subject will be asked to not stimulate at least 8 hours prior to this visit. The subject will receive a single 40-minute therapy session with their Cala Trio. The following assessments will be made prior to the therapy session and again immediately post the therapy session:

- Tremor assessments (while in the *off* state) including:
    - Modified MDS Unified Parkinson's Disease Rating Scale Part III (MDS-UPDRS) be rated by the PI

    - Bain & Findley Activities of Daily Living (BF-ADL) upper limb subset score will be rated by the subject. The BF-ADL assessment will be performed using the subject's own household props

- CGI-S, I/PGI-S, I
    - Clinical/Patient Global Impression of Severity (CGI-S/PGI-S): The investigator or designee (CGI) and Patient (PGI) will rate the overall tremor severity prior to stimulation
    - Clinical/Patient Global Impression of Improvement (CGI-I/PGI-I): The investigator or designee (CGI) and Patient (PGI) will rate the overall tremor improvement following stimulation
- Product/preference survey
- Subjects will return all study materials with provided shipping packages

A factory device reset, and re-calibration may be required if the device does not prompt a tremor task measurement of postural hold prior to starting the stimulation session.

2.2 Analysis Set

Available data on all enrolled subjects will be summarized and reported, referred to in ICH E9 ("Statistical Principles for Clinical Trials") as the *full analysis set*.

2.3 Randomization

The study is non-randomized.

3 STATISTICAL METHODS

3.1 General Principles

- All hypothesis testing will be performed using a two-sided test at a 0.05 level of significance or a one-sided test at a 0.025 level of significance. Significance levels will be adjusted for multiple hypothesis testing where appropriate. P-values will be reported as raw uncorrected p-values and rounded to three decimal places. If a p-value is less than 0.001 it will be reported as "< 0.001."

- Summary statistics will be reported for all analyzed variables.

    - Continuous data will be summarized using descriptive statistics appropriate for the underlying data distribution, e.g., mean, standard deviation, median, and interquartile range (first and third quartiles).

    - Categorical, non-ordered variables will be summarized using frequency counts and percentages.

    - Ordinal variables, including those specific to study outcomes – e.g., the modified Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part III and the Bain & Findley Activities of Daily Living (BF-ADL) scale – will be summarized using measures of central tendency and variance as opposed to frequency counts and percentages; that is, these variables will be reported using the same metrics as defined above for continuous variables.

- For events which can occur more than once in a single subject (e.g., adverse events), the percentage will be based on number of subjects experiencing the event, and both subject and event counts will be reported.

- Baseline is defined as the last measurement for the outcome of interest obtained before the exposure to study treatment.

- For time-to-event analyses, complete dates for events of interest (e.g., index treatment, outcome events, censoring), will always be used. Incomplete dates will be incorporated by defining the event as occurring on the earliest date possible given the incomplete information. E.g., a date of UNK-FEB-2018 will be incorporated into the analysis as 01-FEB-2018.

- Statistical analyses will be performed in SAS (SAS Institute, Cary, N.C.) version 9.3 or later, R (R Foundation for Statistical Computing, Vienna, Austria) version 3.2 or later,

Matlab (The MathWorks, Inc., Natick, MA) version R2016B or later, Python (Anaconda, Inc., Austin, TX) version 3.6.6 or later, or in another validated statistical software package.

3.2 Analysis Populations

As stated in section 2.2, available data on all enrolled subjects will be summarized and reported, referred to in ICH E9 ("Statistical Principles for Clinical Trials") as the *full analysis set*.

Two analysis populations will be defined, one consisting of all available data evaluated under intent-to-treat (ITT) principles, and the other consisting of all available data evaluated under per-protocol (PP) principles. Both populations are defined in further detail below. For each analysis population, "available data" means that missing data will not be replaced or imputed, and that endpoints defined as change scores will require both baseline and follow-up data to be present to be included in analyses.

The ITT population consists of all enrolled subjects.

The PP population is defined as ITT subjects who furthermore:

- Met all inclusion/exclusion criteria (in particular, subjects enrolled but determined post-enrollment not to have met one or more eligibility criteria will be excluded from the PP population).
- Were dispensed a stimulator at the primary timepoint (visit 2b) that had been appropriately calibrated to the subject's tremor; subjects who were dispensed stimulators at the time of visit 2b that had not been calibrated to that subject's tremor will be excluded from the PP population.
- Remained on a stable dosage of tremor and antidepressant medications, if applicable, during the duration of the study.
- Were rated by a certified UPDRS rater at the primary timepoints (Visit 1, Visit 2b and Visit 3).

3.3 Primary Endpoint

The primary endpoint (definition below) of the study is a within-subject comparison of tremor power before stimulation to after stimulation, as captured by the Cala Trio device's on-board accelerometer. Unless otherwise specified, "tremor frequency", "tremor power" and "single-session kinematic response" are per-session metrics defined as follows:

- "Tremor frequency" is the frequency, in the 3 – 12 Hz range associated with Parkinson's tremor, where the power spectral density (PSD) of a session's pre-stimulation accelerometry data is largest.
- "Tremor power" is integral of the PSD of the tremor accelerometry data in the 3 Hz window centered on the session's tremor frequency.

- "Single-session kinematic response" is defined as the pre-stimulation tremor power divided by the post-stimulation tremor power.

The primary endpoint, "kinematic response", is defined as the per-patient median of single-session kinematic responses.

Patients will be included in the primary endpoint analysis only if they completed at least 10 therapeutic sessions that met the following data quality criteria:

- Session was started at least 2 hours after the completion of a previous session.
- Participant completed both pre- and post-stimulation motion recordings, that are each
    - sufficiently free of recording artifact, including accelerometer saturation
    - sufficiently free of motion artifact, as detected by excessive power outside the 3 – 12 Hz tremor range in the PSD.

Statistical testing of the primary endpoint will be based on the cohort of all enrolled subjects for whom outcome data are available. The statistical objective related to the primary endpoint is to demonstrate that action hand tremor severity in PD is reduced with the Cala Trio device. Formally, the null and alternative hypotheses to be tested for the primary endpoint is as follows:

H_0: $M_T \geq 1$

H_A: $M_T < 1$

where M_T is the median kinematic response captured over the 4-week study.

The primary endpoint is a continuous variable and will be summarized descriptively using log-transformed mean and standard deviation, median, and interquartile range (first and third quartiles). Hypothesis testing of the change score M_T will be performed using Wilcoxon's signed-rank test.

The primary endpoint is met if the two-sided p-value is < 0.05, such that the null hypothesis is rejected in favor of the alternative hypothesis. The study is successful if the primary endpoint is met.

3.4 Secondary Endpoints

The study's secondary endpoints reflect within-subject acute and longitudinal improvements of MDS-UPDRS score from tasks specific to action tremor and BF-ADL score from pre-stimulation to post-stimulation at Visits 2b and 3, captured as defined in the protocol. The statistical objective of the secondary endpoint is to demonstrate, using clinician-rated and patient rated measures, that action hand tremor severity in PD is reduced after treatment with the Cala Trio device.

For purposes of the secondary endpoint:

- ***MDS-UPDRS score*** is defined as the ***average of UPDRS ratings from the two assessed action tremor tasks (postural hold and kinetic) relevant to the stimulated upper limb***. Observations where either of these two assessments are missing will not be included in the analysis set.

- ***BF-ADL score*** is defined as the ***average of available BF-ADL ratings from the eight assessed tasks (use a spoon to drink soup, hold a cup of tea, pour milk from a bottle or carton, dial a telephone, pick up your change in a shop, insert an electric plug into a socket, unlock your front door with the key, write a letter) relevant to the stimulated upper limb***. For observations where ratings for a subset of these tasks are missing, the BF-ADL ratings from the remaining tasks will be averaged and the observation will still be eligible for inclusion in the analysis set. This was done because missing data are assumed to result from participants not having the necessary props readily available (i.e., data are "missing at random") and not because participants were systematically unable to perform the respective ADL task.

The secondary endpoints are:

1a. Improvement in the MDS-UPDRS score from pre-stimulation to post-stimulation at Visit 2b.
1b. Improvement in the MDS-UPDRS score from pre-stimulation to post-stimulation at Visit 3.
1c. Improvement in BF-ADL score from pre-stimulation to post-stimulation at Visit 2b.
1d. Improvement in BF-ADL score from pre-stimulation to post-stimulation at Visit 3.
2a. Improvement in the MDS-UPDRS score from Visit 2b pre-stimulation to Visit 3 post-stimulation.
2b. Improvement in BF-ADL score from Visit 2b pre-stimulation to Visit 3 post-stimulation.
3a. Improvement in the MDS-UPDRS score from Visit 2b pre-stimulation to Visit 3 pre-stimulation.
3b. Improvement in BF-ADL score from Visit 2b pre-stimulation to Visit 3 pre-stimulation.

The secondary endpoints will be summarized descriptively using mean, standard deviation, median, first and third quartiles, and interquartile range (between first and third quartiles). Hypothesis testing of the mean change score (μ_T) will be performed using one-sample t-tests (equivalent to a two-sample paired test). Formally, the null and alternative hypotheses to be tested for each of the secondary endpoints are as follows:

H_0: $\mu_T = 0$
H_A: $\mu_T \neq 0$.

Secondary endpoints will only be tested for statistical significance if the primary endpoint is met. Secondary endpoints will be tested in a hierarchical stepwise manner, using a fixed sequence procedure: 1a -1d first, 2a and 2b second, and 3a and 3b third. Statistical significance level will be set at $p < 0.05$ for each of the three endpoint groups, with group 2 (2a, 2b) only tested if all tests in group 1 (1a – 1d) meet significance, and group 3 (3a, 3b) only tested if group 2 meets significance. Because all endpoints within a step (e.g., endpoints 1a – 1d in endpoint group 1) must meet the two-sided $p < 0.05$ in order to proceed to the next step, no adjustment for multiplicity is required.

3.5 Planned Analyses

The analysis of study outcomes will be conducted once all enrolled subjects have been exited from the study. At this time, analyses of the above hypotheses for the primary and secondary endpoints will be performed, and subgroups of interest will also be investigated, tabulated and reported.

3.6 Sample Size and Power

The study sample size (N) is predicated on primary endpoint as detailed above. For the analysis, primary endpoint as well as the secondary endpoints will be tested on the ITT & PP populations. Estimates of effect size for the various outcomes are shown below. Assuming the same effect sizes observed in PROSPECT (clinicaltrials.gov NCT03597100), at p = 0.01 we have the following results:

Metric	p value	N
Kinematic measurements for study duration of 2 weeks	0.01	40
Kinematic measurements for study duration of 4 weeks	0.01	30

Under these effect sizes, assuming up to 10% data attrition and with a paired difference test at 0.01 alpha against the null hypothesis of no change from baseline, the 34 to 50 subject analysis provides 90% power for testing of the primary endpoint.

3.7 Subject Disposition and Accountability

Subject disposition will be presented in a table and flow diagram depicting the disposition of all enrolled subjects, including the number enrolled, the number who completed the study and the number and reasons of those who discontinued.

3.8 Demographics and Baseline Characteristics

Demographic and baseline characteristics, including MDS-UPDRS Part III and BF-ADL upper limb subset scores at enrollment (Visit 1) and MDS-UPDRS Part III and BF-ADL upper limb subset scores at Visit 2b pre-stimulation, will be summarized and reported.

3.9 Effectiveness Outcomes

The effectiveness of the Cala Trio device in reducing action hand tremor symptoms in PD will be evaluated by the following:

Primary Effectiveness Endpoints

- Tremor power (kinematic response), as collected with the device during postural holds (outstretched or wing beating), change from pre-stimulation to post-stimulation across sessions. (See section 3.3 for detailed description).

This endpoint will be tested statistically using hypotheses as defined above.

Secondary Effectiveness Endpoint

- Improvements in MDS-UPDRS action tremor and BF-ADL upper limb subset scores, relevant to the stimulated upper limb, from pre- to post-stimulation (i.e., acute improvement from stimulation)
- Improvements in MDS-UPDRS action tremor and BF-ADL upper limb subset scores, relevant to the stimulated upper limb, from baseline (pre-stimulation Visit 2b) to study exit (post-stimulation Visit 3) (i.e., accumulated baseline improvement from one month therapy + acute improvement from stimulation)
- Improvements in pre-stimulation MDS-UPDRS action tremor and BF-ADL upper limb subset scores, relevant to the stimulated upper limb, from baseline (pre-stimulation Visit 2b) to last session (pre-stimulation Visit 3)
  (i.e., accumulated baseline improvement from one month therapy)

These endpoints will be tested statistically using hypotheses as defined above (See section 3.4 for details).

Additional Exploratory Endpoints, Clinical Data

- Raw change from pre-stimulation to post-stimulation, assessed separately for Visit 2b and Visit 3, for each of the following:
  - Per-task MDS-UPDRS Part III score
  - Per-task BF-ADL upper limb subset score
- % Subjects that had a ≥ 0.5-point average per-task improvement in MDS-UPDRS Part III total score (six tasks, relevant to the stimulated upper limb) from pre-stimulation to post-stimulation, assessed separately for Visit 2b and Visit 3
- % Subjects that had a ≥ 0.5-point average per-task improvement in BF-ADL upper limb subset score (nominally eight tasks, relevant to the stimulated upper limb) from pre-stimulation to post-stimulation, assessed separately for Visit 2b and Visit 3
- Per-task improvements in MDS-UPDRS Part III and BF-ADL from pre-stimulation to post-stimulation, assessed separately for Visit 2b and Visit 3 as follows:
  - Number of patients who had a pre-stimulation score ≥ 2 for that task
  - Raw change in per-task MDS-UPDRS Part III score, for the patients that had a pre-stimulation score ≥ 2 in that task
  - Raw change in per-task BF-ADL upper limb subset score, for the patients that had a pre-stimulation score ≥ 2 in that task

    - Of the patients who had a pre-stimulation MDS-UPDRS score ≥ 2 for that task, % patients who had a ≥ 0.5-point or 1-point (based on resolution of scale for that task) improvement
    - Of the patients who had a pre-stimulation BF-ADL score ≥ 2 for that task, % patients who had a ≥ 0.5-point or 1-point (based on resolution of scale for that task) improvement
- Correlation between pre-stimulation MDS-UPDRS Part III score (average of six tasks relevant to stimulated upper limb) at Visit 2b vs Visit 3
- Correlation between pre-stimulation BF-ADL upper limb score (average of nominally eight tasks relevant to stimulated upper limb) at Visit 2b vs Visit 3
- Correlation between pre- to post-stimulation improvement in MDS-UPDRS Part III score at Visit 2b vs Visit 3.
- Correlation between pre- to post-stimulation improvement in BF-ADL upper limb subset score at Visit 2b vs Visit 3.
- Clinician/Patient Global Impression of Severity (C/PGI-S) summarized for each of Visit 2b and Visit 3
- Clinician/Patient Global Impression of Improvement (C/PGI-I) summarized for each of Visit 2b and Visit 3
- Across and within visit change in MDS-UPDRS – subset relevant to non-stimulated upper limb (raw change) (i.e., Visit 2b pre – post; Visit 3 pre – post; Visit 2b pre – Visit 3 post; Visit 2b pre – Visit 3 pre)
- Duration of effect as reported by participants (from product survey), including:
    - % of patients reporting duration of effect beyond end of stimulation
    - Descriptive statistics (mean, standard deviation, median, interquartile range) of duration of effect for the patients who reported having post-stimulation benefit)
- Summarized responses of each of the seven questions in the Product Survey

<u>Additional Exploratory Analyses At-home</u>

- Distribution of Patient Session Impression of Improvement (PSI-I) over the duration of the study
- Relationship between PSI-I and kinematic (tremor power) improvement
- Response at-home according to pre-stimulation tremor power severity (e.g., top 25% and 50% most severe sessions per subject), including:
    - Median kinematic improvement for each patient's high (≥ 75th percentile), medium (25th - 75th percentile) and low (≤ 25th percentile) pre-stimulation tremor
    - Of sessions that started with patient-specific "high" tremor, % sessions that had ≥ 0% and ≥ 50% reduction in tremor (i.e., rate of tremor "relief")
    - Of sessions that started with patient-specific "low" tremor, % sessions that ended in "low" tremor (i.e., rate of tremor "prevention")
- Average change in tremor power 1st, 2nd, 3rd and 4th weeks
- Correlation between log-transformed median pre-stimulation tremor power in first subset of days (e.g., 7, 14) compared to analogous data from final subset of days
- Correlation between log-transformed median kinematic response in first subset of days (e.g., 7, 14) compared to analogous data from final subset of days
- Summary of available BF-ADL scores based on recall at Visit 1, and 1st, 2nd, 3rd and 4th weeks
- Device usage metrics

These exploratory endpoints will be summarized and reported descriptively. Any hypothesis testing will be performed according to the general principles defined in this document.

3.10 Subgroup and Covariate Analysis

Primary and secondary endpoints will be investigated for each of the following subgroups:

- Gender (M/F)
- Age (< 65 vs. ≥ 65 years)
- Age onset of hand tremor in PD (by median split)
- Duration of PD from beginning of tremor symptom (by median split)
- Tremor Medications (On/Off)
- Antidepressant medications (On/Off)
- Preset Stimulation amplitude (by median split)
- Average Final Stimulation amplitude (by median/mean split)
- Stimulation calibration frequency (by median split)
- Baseline tremor severity as measured by UPDRS (by median split)
- Baseline tremor severity as measured by BF-ADL scale (by median split)
- Pre-stimulation tremor power, aggregated over first week (by median split)

Comparisons within subgroups will be performed using the same statistical tests as specified above for the overall cohort.

3.11 Safety Outcomes

The safety of the Cala Trio device will be evaluated by the incidence of device and therapy-related adverse events. Additionally, all adverse events documented during study conduct will be tabulated and reported.

3.12 Missing Data

The primary analysis cohort will be the full analysis set as defined above. Missing data will not routinely be replaced or otherwise imputed unless explicitly specified in an endpoint calculation. For any calculations that require data imputation, sensitivity analyses will be performed to characterize effect of best-case and worst-case outcomes.

Statistical Analysis Plan Revision History:

Effective Date	Version	Author	Description/Summary of Revisions
Feb 22 2021	A	Ruta Deshpande	Initial release of this Statistical Analysis Plan
Mar 26 2021	B	Apoorva Rajagopal	Explicit definitions of tasks for inclusion in secondary endpoints