STATISTICAL ANALYSIS PLAN

Protocol Title (Number):

High-FreQUEncy Nerve Block for PoST-Amputation Pain: QUEST IDE Trial (A Pivotal Study)
(003-0001)

Date: 09 February 2023

ClinicalTrials.gov Identifier: NCT02221934

Sponsor:

Neuros Medical, Inc.
35010 Chardon Road, Suite 210
Willoughby Hills, OH 44094

Prepared in collaboration with:

Avania US - Marlborough
100 Crowley Drive, Suite 216
Marlborough, MA 01752

TABLE OF CONTENTS

1 ABBREVIATIONS

Abbreviation	Definition
ADE	Adverse Device Event
ADL	Activities of Daily Living
AE	Adverse Event
AKA	Above Knee
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
AVANIA	Avania US – Marlborough, MA
BDRM	Biostatistics Data Review Meeting
BKA	Below Knee
BPI	Brief Pain Inventory
CRF	Case Report Forms
CSR	Clinical Study Report
DMC	Data Monitoring Committee
EQ-5D	EuroQol 5D
FAS	Full Analysis Set
FDA	United States Food and Drug Administration
HFNB	High Frequency Nerve Block
HR-QOL	Health-related quality of life
IPA	Independent Physician Adjudicator
IRB	Institutional Review Board
ITT	Intent-To-Treat Analysis Set
LTFU	Long-term Follow-up
MCS	Mental Component Score
MED	Morphine Equivalent Dose
N/A	Not Applicable
NRS	Numerical Rating Scale
PCS	Physical Component Score
PGIC	Patient Global Impression of Change
PP	Per-Protocol Analysis Set
SADE	Serious Adverse Device Event
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SF-12	Short Form Health Survey
UADE	Unanticipated Adverse Device Effect

2 CHANGES FROM STUDY PROTOCOL

This Statistical Analysis Plan (SAP) defines the statistical methods that will be used to analyze the data from the QUEST Pivotal Study. In developing the SAP, certain changes have been made compared to the statistical information that was provided in the Study Protocol. The following table provides a list of the changes from statistical sections of the Study Protocol to this SAP and the justification for each change.

Table 1. Changes from Study Protocol

SAP Section	Description	Justification
6.3 Full Analysis Set	The SAP replaces the protocol-specified term "Modified Intent-to-Treat" (mITT) Analysis set with "Full Analysis Set" (FAS).	The "Full Analysis Set" is the more commonly used term to describe this type of analysis set per Guidance for Industry E9 Statistical Principles for Clinical Trials, https://www.fda.gov/media/71336/download.
6.4 Per-Protocol Analysis Set	The protocol specified the Per-Protocol (PP) Analysis set will be restricted to study participants who fulfill the protocol in terms of eligibility, treatment and outcome assessment. The SAP adds the condition that participants only need to fulfil these criteria through 3 months of follow-up to be considered for the Per-Protocol analysis set. Add condition that PP analysis will only be analyzed if loss from FAS due to protocol deviations is >5%.	The 3-month post-implant phase (the randomized testing phase) is the main time period of interest for primary and key secondary efficacy endpoints and for primary safety, so it is primarily necessary for subjects to satisfy the Per-Protocol criteria only during this time period. There is no need for a PP analysis set if losses from the FAS due to protocol deviations are ≤5% because the difference in outcomes is unlikely to be meaningful.
7.3 Methods for Withdrawals and Missing Data and 10.1 Primary Effectiveness Variable	SAP adds detail and clarification about which treatment sessions will be used in the primary endpoint analysis and how missing data will be handled.	The protocol is vague in terms of how treatment sessions that appear in the device log but not in the eDiary, or vice versa, should be used, and how to handle missing pain scores at 30-minute follow-up. The SAP is more specific.
7.3 Methods for Withdrawals and Missing Data	SAP defines the covariates in multiple imputation and subgroup analysis (etiology of amputation, amputation location, pain type, baseline pain intensity and baseline pain duration).	Definition of covariates needed for programming.
7.3 Methods for Withdrawals and Missing Data	A tipping point analysis has been added to the SAP.	A tipping point analysis is a common missing-data sensitivity analysis, and it is felt it will be useful here.
7.6 Assessment of Homogeneity	SAP specifies that the Breslow-Day test will be used to assess homogeneity across sites.	The protocol is vague in terms of test for homogeneity stating an analysis will be performed. The SAP is more specific and outlines the analysis using the Breslow-Day test.
10. Effectiveness Analyses	Statistical testing in SAP changed from protocol-specified two-sided testing (at	The SAP was revised in this manner to clarify the direction of a successful rejection of the null hypothesis.

SAP Section	Description	Justification
	a two-sided 0.05 level of significance) to one-sided testing (at a one-sided 0.025 level of significance).	
10.1 Primary Effectiveness Variable	Analysis method has changed from Cochran-Mantel-Haenszel (CMH) to logistic regression analysis controlling for etiology (dysvascular, trauma, other), amputation location (AKA, BKA), pain type (phantom, stump, both), baseline pain intensity (5-6, 7-10) and baseline pain duration (episodic, persistent).	The primary effectiveness endpoint is a binary success outcome to assess pain resolution. The change from CMH to logistic regression does not impact the definition, measurement or calculation of the primary effectiveness endpoint, only the method by which the statistical significance of the primary effectiveness endpoint in the two treatment groups will be analyzed. This is not a significant change in the context of determining study success. This change is being made prior to breaking the treatment blind. The CMH test was originally proposed with site as the only stratification factor in comparing treatment groups over the first three months. A problem with site as a stratification variable is that there are many possible ways to construct site groupings which introduces randomness into the primary effectiveness endpoint analysis. Site is by itself a complex covariate with many extraneous factors influencing prognosis. In contrast, it is well recognized that there are more prognostic outcome variables than site inclusive of etiology (dysvascular, trauma, other), amputation location (AKA, BKA), pain type (phantom, stump, both), baseline pain intensity (5-6, 7-10) and baseline pain duration (episodic, persistent). In addition, there are only 7 degrees of freedom lost in the above covariates whereas there are 11 possible site groups which consumes more degrees of freedom than from the five prognostic factors. Improved probability of success prognosis improves the overall goodness of fit. Taken together, for the QUEST Study, logistic regression allows the analysis to gain degrees of freedom in reducing the sources of variation (and to control for potential imbalances) in the model relative to the less efficient CMH test which just stratifies by site clusters.

SAP Section	Description	Justification
10.2 Secondary Effectiveness Variables Intended for a Labeling Claim	SAP specifies analysis of secondary endpoints intended for labeling will be adjusted for study center. Change all analyses in this section from two-sided with a significance level of 0.05, as specified, in the protocol to one-sided with a significance level of 0.025 in the SAP.	The protocol specifies these tests will be unadjusted; however, to be consistent with the analysis of the primary effectiveness endpoint the SAP indicates an adjusted analysis will be performed. The SAP was revised in this manner to clarify the direction of a successful rejection of the null hypothesis.
Covariate Analysis (protocol section 11.3.2)	Included in the protocol, removed from SAP	The covariate analysis using multivariable logistic regression was removed from the Additional Analysis section because the primary effectiveness endpoint analysis was changed to logistic regression.

3 SEQUENCE OF PLANNED ANALYSES

3.1 INTERIM ANALYSES FOR FUTILITY

Interim analysis will be conducted during the study to test for futility. The points of interim analysis will be: n=20 (12.3% of the planned final sample size), n=40 (24.7%) and n=80 (49.4%). After each analysis is conducted, the independent statistician will present the interim results to the Data Monitoring Committee (DMC). DMC recommendations at each interim analysis will proceed according to pre-specified decision rules for assessing futility. Specifically, the interim analyses will result in one of the following two DMC recommendations for the study:

A. Enrollment continues uninterrupted and reaches the planned maximum.
B. Enrollment is stopped and the study is terminated due to futility.

The group sequential design will utilize a beta spending function to distribute the probability of Type II error ($\beta = 0.10$) throughout the study. A beta-spending function was used to maintain the beta-spent from the first interim look already completed, while maintaining futility significance levels for the upcoming futility analyses that are approximately similar to the futility significance levels in the original protocol. The function uses the following cumulative beta-spent at each look: 0.006, 0.00615, 0.0102, 0.1, providing some chance for identifying futility early in the study. In this manner, study futility will be assessed at all interim analyses.

For each analysis point, the P-value boundaries to accept H_0 (claim futility) are provided in Table 2. These boundaries were calculated using PROC SEQDESIGN in SAS as follows:

```
proc seqdesign errspend bscale=pvalue;
      futility: design nstages=4 alpha=0.025 beta=0.10
                info=cum(0.123 0.247 0.494 1) alt=lower
                method=errspend(0.006 0.00615 0.0102 0.1) stop=accept;
run;
```

The probabilities of crossing a boundary (i.e., the probabilities of futility being reached) under the null hypothesis are listed within the column labeled “Boundary Crossing Probabilities”.

Table 2. Interim Analysis Boundaries

Information Fraction	Accrual	Cumulative β Spent	One-sided P-value Boundaries to accept H_0 futility	Boundary Crossing Probabilities Under H_0
0.123	20	0.006	≥ 0.91440	0.086
0.247	40	0.00615	≥ 0.96312	0.101
0.494	80	0.0102	≥ 0.62107	0.401
1.000	162	0.1	≥ 0.025	0.975

The DMC will review aggregate safety data on a regularly scheduled basis as identified in the DMC Charter. The independent physician adjudicator (IPA) and/or the DMC will alert the study sponsor if any safety concerns arise during the conduct of the study. Stopping rules for safety may be outlined in the DMC Charter. Information to be provided to the DMC will include treatment effect estimates and their confidence intervals (blinding the group assignment), whether a futility boundary has been crossed (i.e., whether futility can be claimed to have been met), the conditional power of crossing by the end of the study if none have yet been crossed, and safety assessments.

The DMC will make the recommendation on whether to continue or stop the study after each look. The planned statistical boundaries are "non-binding", meaning that the DMC is not required to stop the study if a boundary is crossed - all available information about the study will be used to make the recommendation. The DMC might also recommend stopping the study for low conditional power even though a boundary was not crossed. Note that if the study meets a futility boundary, this implies that, if the study proceeded to its planned end, it is unlikely the Test treatment would be significantly more beneficial than Control with respect to the primary endpoint responder rate. Examples of how this could occur include: (a) the Control group primary endpoint rate is as expected in the sample size calculations below but the Test treatment primary endpoint rate is similar to or less than (i.e., worse than) the Control group primary endpoint rate; or (b) the Test group primary endpoint rate is as large as expected in the sample size calculations below but the Control treatment primary endpoint rate is similar to or larger (better) than the Test group primary endpoint rate.

3.2 Primary Endpoint Analysis at Month 3

The unblinded analyses on the primary effectiveness and other endpoints will occur after the targeted number of subjects have reached their Month 3 primary endpoint. These unblinded primary analyses will be carried out by an independent statistician not otherwise involved in the conduct, analysis planning, and analysis of the study.

The Clinical Study Report for the Month 3 primary endpoints will include the primary effectiveness result by treatment group and the associated test statistic to test futility for each of the three prior interim analyses, as well as the results of other analyses performed at Month 3.

3.3 Final Analysis at Month 12

The final analyses of the study will occur after the targeted number of subjects have reached their Month 12 follow-up. Long-term (post-12 month) follow-up will be reported periodically until End-of-Trial declaration.

4 Study Objectives and Endpoints

4.1 Study Objective

4.1.1 Primary Objective

The primary objective is to evaluate the safety and effectiveness of the Altius® System High-Frequency Nerve Block (HFNB) treatment for the management of post-amputation pain.

4.1.2 Secondary Objectives

The secondary objective is to determine the impact of Altius HFNB treatment for post-amputation pain on health-related quality of life outcomes and medication use.

4.2 Study Endpoints

4.2.1 Primary Endpoint

The study's **primary effectiveness endpoint** is responder status. A subject will be considered a responder if they attain a significant pain reduction at the end of more than half of the treatment sessions during the parallel Randomized Testing phase of the study. Specifically, a responder must attain ≥50% pain reduction in ≥50% of the treatment sessions during the Randomized Testing phase of the study (through Month 3).

The study's **primary safety endpoint** is the incidence of all serious adverse events, including Serious Adverse Events (SAEs), Serious Adverse Device Events (SADEs), and Unanticipated (Serious) Adverse Device Events (UADE), from the time of injection through three months post-implant. The primary safety endpoint will be determined at the conclusion of the Randomized Testing phase of the study, after all active participants complete the Month 3 Visit.

4.2.2 Secondary Effectiveness Endpoints

Secondary Endpoints for Intended Labeling Claims:

- Change from baseline in Opioid Pain Medication Use (Morphine Equivalent Dose (MED)) at Month 3
- Change from baseline in Brief Pain Inventory (BPI) at Month 3
- Change from baseline in Short Form Health Survey (SF-12) Physical Component Summary (PCS) at Month 3
- Change from baseline in Short Form Health Survey (SF-12) Mental Component Summary (MCS) at Month 3
- Change from baseline in EuroQol (EQ-5D) at Month 3

Secondary Effectiveness Endpoints not Intended for Labeling Claims:

- Primary effectiveness beyond Month 3 through Month 12
- Pain Relief after 2 Hours
- Pain Days per Week
- Change from baseline in Non-Opioid Analgesic Pain Medication Use through Month 12
- Change from baseline in Opioid Pain Medication Use (Morphine Equivalent Dose (MED)) through Month 12
- Change from baseline in Brief Pain Inventory (BPI) through Month 12
- Change from baseline in EuroQol (EQ-5D) through Month 12
- Change from baseline in Short Form Health Survey (SF-12) Physical Component Summary (PCS) through Month 12
- Change from baseline in Short Form Health Survey (SF-12) Mental Component Summary (PCS) through Month 12
- Change from baseline in Patient Global Impression of Change (PGIC)
- Session Success Rate
- Composite Responder Rate (Reduction in pain AND absence of increase in medication usage)

Additional Outcome Assessments:

- Technical Success rate (Successful Device Placement and Activation)
- Prosthetic Use

4.2.3 Secondary Safety Endpoints

The secondary safety endpoint is the incidence of all adverse events including non-serious adverse event (AE), non-serious adverse device effects (ADE), serious adverse events (SAE), serious adverse device events (SADE), and unanticipated adverse device effects (UADE), from time of injection through the Month 12 visit. In addition, incidence of AEs and SAEs occurring post informed consent through initiation of implant procedure will be collected in both the Enrolled-not-Implanted and the Enrolled-and-Implanted populations.

5 Sample Size

The study is powered according to the primary effectiveness endpoint. Specifically, the primary null and alternative effectiveness hypotheses are:

$$H_0: P_T = P_C \text{ vs. } H_A: P_T > P_C,$$

where P_T and P_C are the proportion of responders (responder rate) for the Test and Control treatment, respectively.

Sample size is calculated under the following operational characteristics and assumptions:

- Probability of type I error, $\alpha = 0.025$ one-sided
- Probability of type II error, $\beta = 0.10$
- Power, $1-\beta = 0.90$
- Test, one-tailed
- Expected responder rate to Test treatment ($\mathbf{T}_t$), $P_T = 0.50$
- Expected responder rate to Control treatment ($\mathbf{C}_t$), $P_C = 0.25$
- Inflation factor (IF) for three interim analyses, IF = 1.09
- Percent attrition = 10%
- 1:1 randomization ratio into Test and Control

Total sample size based on the above parameters using a normal approximation multiplied by an inflation factor for group sequential design (PASS 15) is N=180.

6 ANALYSIS SETS

Safety endpoints will be evaluated for all subjects enrolled in the study who undergo surgery. The primary analysis of the primary effectiveness endpoint will be analyzed on a full analysis set basis (FAS). In addition, a per-protocol (PP) and intention-to-treat (ITT) analysis will be conducted. Secondary effectiveness endpoints will be evaluated utilizing the FAS, the ITT and the PP basis.

6.1 SAFETY ANALYSIS SET

The safety analysis set will include all subjects who undergo surgery. This will be the primary analysis set for the safety analyses. Subjects are analyzed under the treatment received.

6.2 INTENT TO TREAT ANALYSIS SET (ITT)

The Intention-to-Treat (ITT) Analysis Set includes all subjects who were randomized. This analysis estimates the causal effect of being assigned to the treatment vs. control. Subjects who were enrolled-not-implanted, or enrolled-and-implanted but were terminated from the study prior to randomization will not be included in the ITT analysis set. Subjects who were enrolled-and-implanted and randomized but never used the device to deliver treatment will be included with missing data imputed per the procedures described in Section 7.3. The ITT analysis set will be utilized for a supportive analysis of both the primary and key secondary study endpoints.

6.3 FULL ANALYSIS SET (FAS)

The full analysis set (FAS) includes all subjects who were randomized and had documented treatment use. Subjects who were enrolled-not-implanted, or enrolled-and-implanted but were terminated from the study prior to randomization, or enrolled-and-implanted and randomized but never used the device to deliver treatment will not be included in the FAS. The FAS population is the primary analysis set for effectiveness. Subjects are analyzed under the treatment to which they were randomized. This analysis dataset was previously referred to as the Modified Intent-to-Treat (mITT) dataset in the protocol.

6.4 Per-Protocol Analysis Set (PP)

The Per-Protocol (PP) analysis set will be restricted to study participants who fulfill the protocol through 3 months of follow-up in the terms of the eligibility, treatment, and outcome assessment; any subject with a major protocol deviation will be removed from the PP analysis set as noted below. This analysis attempts to estimate the treatment effect when complying with the protocol but is typically subject to selection bias and thus leads to a biased treatment effect estimate. Subjects will be excluded from this PP analysis set for reasons including: programmed to a treatment opposite their group assignment, missing study required visits, use of non-contract pain medications, and device non-compliance including inadequate occurrences of device use (duration of sessions, number of sessions). If necessary, inclusion of a subject or session in the PP analysis dataset may be assessed by the IPA in a blinded manner. The PP analysis set will be omitted if losses relative to the FAS due to protocol deviations are small (≤5%).

6.5 Enrolled-Not-Implanted and Enrolled-Implanted-Not Randomized

Enrolled subjects in whom the device is not implanted will be summarized separately; summary will include reason for screen failure. Similarly, subjects who are enrolled-and-implanted but were terminated from the study prior to randomization will be summarized separately; summary will include reason for termination.

7 General Issues for Statistical Analysis

The study will use a frequentist approach to statistical analysis. Descriptive statistics to be presented include mean, standard deviation, median, quartiles, minimum and maximum for continuous variables, and number and percentage of subjects in each category for categorical variables. P-values will be one-sided and considered significant at the 0.025 level, or two-sided and considered significant at the 0.05 level of significance.

7.1 Analysis Software

Analysis data sets, statistical analyses and associated output generated by Avania will be generated using SAS® Software version 9.4 or later and/or R version 3.3.2 or later.

7.2 Disposition of Subjects and Withdrawals

The number and percent of subjects in the ITT, FAS, PP and Safety populations will be presented by treatment group. Percentages will be based on the number of ITT subjects in the given randomized treatment group. In addition, the number of subjects in the Enrolled-not-Implanted and in the Enrolled-and-Implanted “but not randomized or device never used” populations will be provided.

The number and percentage of ITT subjects within each Status at Exit category (Randomized but Device Not Used, Treated and Exited Prior to Month 3, Treated and Exited at Month 3, Treated and Exited Prior to Month 12, Treated and Exited at Month 12 or During Long-Term Follow-up (LTFU), or continuing in LTFU) and within each Reason for Study Exit, as collected on the Study Exit Case Report Form (CRF) page, will be presented overall and by randomized treatment group.

The reasons for Study Exit for the ITT Analysis Set are:

- Voluntary Subject Withdrawal for any Reason
- Voluntary Subject Withdrawal due to an AE
- Voluntary Subject Withdrawal NOT due to an AE
- Investigator has requested that the subject be withdrawn from the study
- Subject completed device placement, and chose not to continue to long-term follow-up
- Lost to Follow-up
- Subject has died
- End of Trial
- Other

Within each Status at Exit Category, the number and percentage of subjects with each reason for Study Exit will be presented overall and by randomized treatment group.

The number and percentage of ITT enrolled-and-implanted subjects who had the device explanted will be presented overall and by randomized treatment group.

7.3 Methods for Withdrawals and Missing Data

For effectiveness, the best indicator of a subject's overall response to treatment, Test ($\mathbf{T}_t$) or Control ($\mathbf{C}_t$), will be the subject's available responses to treatment. As such, subjects who are randomized into the Test ($\mathbf{T}_t$) or Control ($\mathbf{C}_t$) group to receive treatment but who terminate prior to their scheduled Month 3 Visit (Day 91 + 14 days post-implantation) will be determined to be a responder or non-responder based on their available data prior to termination. This is the primary analysis for effectiveness. For this primary analysis, missing observations for pain score at 30 minutes for a particular complete treatment session will be considered a failure for that session. For this primary analysis, treatment sessions that are interrupted with rescue (p.r.n.) pain medications will utilize the assessment of pain at the time of the interruption when available; if such pain assessment is not available at the time of rescue medication, missing observations will be considered a failure for that session. See Table 3 in Section 10.1 for a description of how missing data scenarios will be handled in the primary analysis.

Three separate sensitivity analyses will be performed to evaluate the potential impact of analyzing available data from subjects who terminate prior to the primary endpoint. The following three scenarios will each be implemented, unless otherwise specified, to explore the impact on the primary effectiveness outcome:

1. Assume dropouts are missing completely at random. Sensitivity analysis will remove all subjects from the FAS analysis set who did not reach the Month 3 primary endpoint. In this way the primary effectiveness endpoint would be analyzed using only subjects who completed the primary endpoint assessment through Month 3. To further explore the effect of random dropouts, sensitivity analysis will consider subjects within analysis set who did not reach the primary endpoint as "responder" or "non-responder" in the same proportion as

subjects who did complete the primary endpoint. For example, if 54% of the subjects in the Test group who completed the primary endpoint are responders, of the subjects in the Test group who did not reach the primary endpoint, 54% would be randomly assigned a value of "responder" and 46% the value "non-responder". However, it may not be a valid assumption that missing data are missing-completely-at-random and are not related to a subject's unobserved outcomes, and therefore additional sensitivity analyses will be performed, as follows.

2. Assume informative dropouts. Sensitivity analysis will consider all FAS subjects who did not reach the Month 3 primary endpoint as non-responders. It is plausible that subjects withdraw due to a lack of effectiveness from this "as needed" treatment, and it would be appropriate to consider these subjects non-responders. Thus, it will be assumed that Test treatment ($\mathbf{T}_t$) or Control treatment ($\mathbf{C}_t$) had no effect and a subject who did not reach the primary endpoint would be assigned a value of "non-responder", independent of group assignment.
3. Assume treatment policy strategy. Sensitivity analysis will include all ITT subjects who were implanted, but not included in the FAS, as non-responders. Subjects enrolled-and-implanted but never used the device to deliver treatment, will be included in the analysis as non-responders. All dropouts will be analyzed as non-responders. A data set including all subjects who received an implant will represent an ITT analysis.

Multiple Imputation. For FAS, subjects with incomplete follow-up through Month 3, multiple imputation logistic regression with fully conditional specification will be used to impute responder status (yes/no) at 3 months. Included in the imputation model will be baseline variables of:

- Etiology (dysvascular, trauma, other)
- Amputation location (AKA, BKA)
- Pain Type (phantom, stump, both)
- Baseline Pain Intensity (5-6, 7-10)[1]
- Baseline Pain Duration (episodic, persistent)

Any missing covariates will be imputed once using FCS logistic regression approach in SAS PROC MI where all covariates listed above are included in the VAR statement. Once a complete dataset of covariates is generated, 50 imputed datasets with complete primary outcome data will be generated from it using logistic regression with the above covariates as the imputation model. For each dataset, the logistic regression test-statistic testing the primary null hypothesis will be generated. The test-statistic will be combined across the 50 imputed datasets using PROC MIANALYZE in SAS to create one overall summary statistic to test the primary null hypothesis across the imputed datasets.

[1] Defined as the average of the end-of-day worst pain scores from the subject's e-diary compliant eligibility window, consistent with the study eligibility criteria.

Tipping Point. For the ITT population, a tipping point analysis will be used to impute missing data. Specifically, each Test and Control subject with incomplete data through Month 3 will first be considered a success (responder). The logistic regression model analysis on the primary endpoint will then be carried out. If the null hypothesis is rejected, one randomly selected Test subject with missing data will be imputed as a failure (non-responder) and all Control subjects with missing data will still be imputed as successes, and the logistic regression model analysis on the primary endpoint will again be carried out. If the null hypothesis is rejected, this process will be repeated, where two Test subjects will be randomly selected and imputed as a failure and the null hypothesis will again be tested. This algorithm will be repeated, imputing three randomly selected Test subjects as failures, then four, etc., until the "tipping point" is reached (i.e., until the null hypothesis is no longer rejected).

This entire process in the previous paragraph will then be repeated where, at first, one randomly selected Control subject with missing data will be imputed as a failure for the primary endpoint. It will then be repeated where two Control subjects with missing data will be randomly imputed as failures, etc.

There will be no imputation of missing data for secondary effectiveness endpoints or for safety endpoints.

7.4 Protocol Deviations

All site-specific protocol deviations will be reviewed and, if necessary, site corrective actions will be implemented to mitigate future deviations. eDiary non-compliance will not be reported as a protocol deviation but will be aggregated and presented at time of final study analysis. A Biostatistics Data Review Meeting (BDRM) will be held prior to database lock to determine which protocol deviations are major; the major deviations will be removed from the PP population.

Protocol deviations will be summarized in the CSR. This summary will include the number and percent of FAS, ITT and PP subjects (overall and by site) with each deviation type within each randomized treatment group.

7.5 Multiple Comparisons and Multiplicity

There is one primary effectiveness endpoint, and it will be compared between treatment groups at a one-sided 0.025 level of significance. There are 5 secondary endpoints intended for a labeling claim. In order to control type I error, the 5 secondary endpoints intended for labeling claim will be tested in a hierarchical, gatekeeping manner in the order specified below, each at a one-sided 0.025 level of significance, and only after the primary study effectiveness endpoint has been achieved in favor of the test treatment. Further, each secondary endpoint numbered 2 through 5 below will be tested for significance between treatments only if the previous secondary endpoint in the list has been shown to be significantly beneficial for the Test treatment.

This hierarchical closed test procedure is used to account for multiple testing with the goal of controlling the "maximum overall Type I error rate", which is the maximum probability that has been

prespecified that one or more null hypotheses are rejected incorrectly, at a one-sided 0.025 level of significance.

1. Change from baseline in Opioid Pain Medication Use (Morphine Equivalent Dose (MED)) at Month 3
2. Change from baseline in Brief Pain Inventory (BPI) at Month 3
3. Change from baseline in Short Form Health Survey (SF-12) Physical Component Summary (PCS) at Month 3
4. Change from baseline in Short Form Health Survey (SF-12) Mental Component Summary (MCS) at Month 3
5. Change from baseline in EuroQol (EQ-5D) at Month 3

There will be no adjustment for multiple comparisons across any remaining effectiveness endpoints, nor will there be adjustment for multiple comparisons across safety endpoints.

7.6 Assessment of Homogeneity

The analysis on the primary effectiveness endpoint will be performed on data pooled across study sites. To assess homogeneity of treatment difference on the primary endpoint across sites, Breslow-Day tests will be used. Study sites with <5 subjects will be excluded from this analysis. A Breslow-Day result that: (i) is not significant at a 0.15 level of significance; or (ii) is significant but where the Test treatment has a higher response rate than the Control treatment within each site, supports pooling of results across sites. Otherwise, demographics and baseline characteristics will be inspected within each site to assess if difference in demographic and baseline characteristics are the cause of any lack of consistency across sites.

Given the large number of study sites with <5 subjects, a secondary poolability analysis is planned with study sites pooled by region. The following regions will be used: Southeast, South, Midwest, West. The sites will be grouped as follows:

Site	Location	Region	# Randomized Patients
01 – Center for Clinical Research	NC	Southeast	28
02 – University of Michigan	MI	Midwest	1
03 –St. Lukes Presbyterian	CO	West	1
04 – Kettering Medical Center	OH	Midwest	5
05 – Cleveland Clinic Foundation	OH	Midwest	16
06 – Louis Stokes Cleveland VA	OH	Midwest	2
08 – University of Alabama Birmingham	AL	South	1
10 – Arizona Pain Specialists	AZ	West	3
12 – University of Illinois at Chicago	IL	Midwest	2
15 – University of Arkansas for Medical Sciences	AR	South	3
17 – University of Washington	WA	West	4
18 – Drug Studies America	GA	Southeast	4
19 – Ochsner Clinic Foundation	LA	South	1

21 – Sarah Cannon Research Foundation	TN	South	5
22 – Henry Ford Hospital	MI	Midwest	2
24 – Advanced Surgical and Research Solutions	OK	South	35
25 – Baylor Scott and White Research Institute	TX	South	8
26 – Meta Medical	OH	Midwest	19
27 – Nona Medical	FL	Southeast	8
28 – Mayo Clinic	MN	Midwest	1
30 – Emory University Hospital/Grady Hospital	GA	Southeast	1
31 – University of Louisville	KY	South	15
32 – The Surgical Clinic	TN	South	7
33 – Legacy Brain & Spine	GA	Southeast	4
34 – Cardiovascular Surgery Clinic	TN	South	2

This provides a total of 5 sites (45 randomized patients) in the Southeast, 9 sites (77 randomized patients) in the South, 8 sites (48 randomized patients) in the Midwest and 3 sites (8 randomized patients) in the West. A similar Breslow-Day analysis as described above will be carried out to assess homogeneity of treatment effect across regions.

A similar Breslow-Day analysis will be carried out to assess homogeneity of treatment effect across type of post-amputation limb pain (stump [or residual] limb pain vs. phantom limb pain).

8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Descriptive statistics of demographics and other baseline characteristics will be presented for the ITT and FAS populations overall and by treatment group unless otherwise specified. Descriptive statistics include mean, standard deviation, median, quartiles, minimum and maximum for continuous variables, and number and percentage of patient in each category for categorical variables. There will be statistical comparisons (e.g., Fisher Exact, or t-test/ANOVA) between randomized treatment groups with respect to the distribution of these variables to test if the two treatment groups were similar at baseline. A two-sided 0.05 significance level will be used to determine significance unless otherwise stated.

8.1 DEMOGRAPHICS

The demographic variables to be collected and presented with descriptive statistics overall and by randomized treatment group are:

- Age (years)
- Sex at Birth (Male, Female)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not Reported, Unknown)
- Race (American Indian/Alaskan Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Other, Unknown).
- Height (cm)
- Weight (kg)
- Body Mass Index (BMI)

The demographics table will be repeated in the PP population.

8.2 Baseline Amputation Details

The following variables are collected regarding patient amputation; descriptive statistics will be presented by randomized treatment group:

- Side of Amputation (Right, Left)
- Level of Amputation (Above Knee [AKA], Below Knee [BKA])
- Cause of Amputation (Dysvascular, Trauma, Other)
    - Other causes of amputation will be provided in a listing
- Months from Amputation to Baseline Visit (calculated from Amputation Date and Baseline Date of Visit in which Amputation Details CRF page is completed)
- Hours/Day of Prosthetic Leg Use (N/A – no prosthetic leg, Zero, >0-2, 2-4, 4-6, 6-8, 8-12, 12-16, 16-20, 20-23, All Day)

The baseline amputation details table will be repeated in the PP population.

8.3 Baseline Amputation Pain Assessment

The following variables are collected at baseline and will be presented with descriptive statistics by randomized treatment group:

- Months from First Limb Pain to Baseline Visit (calculated from Date of First Limb Pain and Baseline Date of Visit in which Amputation Pain Assessment CRF page is completed)
- Worst limb pain on a usual day (scale of 0-10 by intervals of 1, will be presented as a continuous variable and by categories: No pain (0), Mild (1-3), Moderate (4-6), Severe (7-9), Worst Possible Pain (10)). Note this data will be taken from the eDiary and averaged across the baseline eDiary Eligibility "PASS" window (2 weeks) per patient.
- Least limb pain on a usual day (scale of 0-10 by intervals of 1, will be presented as a continuous variable and by categories: No pain (0), Mild (1-3), Moderate (4-6), Severe (7-9), Worst Possible Pain (10)) . Note this data will be taken from the eDiary and averaged across the baseline eDiary Eligibility "PASS" window (2 weeks) per patient.
- Average limb pain on a usual day (scale of 0-10 by intervals of 1, will be presented as a continuous variable and by categories: No pain (0), Mild (1-3), Moderate (4-6), Severe (7-9), Worst Possible Pain (10)) . Note this data will be taken from the eDiary and averaged across the baseline eDiary Eligibility "PASS" window (2 weeks) per patient.
- Limb pain right now (scale of 0-10 by intervals of 1, will be presented as a continuous variable and by categories: No pain (0), Mild (1-3), Moderate (4-6), Severe (7-9), Worst Possible Pain (10))
- Pain Duration Type (Episodic, Persistent)
- Limb Pain Ever Reach Zero (No, Yes)
- Number of Episodes Per Week of Pain Rated 5 or Above (this will be based on those subjects without Persistent Pain; also, this is collected as Episodes Per Week or Episodes Per Day; all will be converted to Episodes Per Week prior to presentation)

- Duration (hours) of Usual Episode of Pain at 5 or Above When Not Taking Rescue Meds (this will be based on those subjects without Persistent Pain; also, this is collected in minutes, hours, or days; all will be converted to Hours prior to presentation)
- Other Area Pain (None, Back Pain, Shoulder Pain, Other)
- Ever Felt Phantom Pain (No, Yes)
- Worst Phantom Pain (scale of 0-10 by intervals of 1; subjects who don't experience phantom pain will be assigned a zero; will be presented as a continuous variable and by categories: No pain (0), Mild (1-3), Moderate (4-6), Severe (7-9), Worst Possible Pain (10))
- Ever Felt Stump Pain (No, Yes)
- Worst Stump Pain (scale of 0-10 by intervals of 1; subjects who don't experience stump pain will be assigned a zero; will be presented as a continuous variable and by categories: No pain (0), Mild (1-3), Moderate (4-6), Severe (7-9), Worst Possible Pain (10))
- Limb Pain Type (Stump Pain, Phantom Only, Stump is Much Worse, Phantom is Much Worse, Both Stump and Phantom Pain are Bad)

The baseline amputation details table will be repeated in the PP population.

8.4 Additional Clinical Care

A listing of subjects requiring at least one additional (non-Altius) pain treatment from randomization through Month 3 will be presented, including analysis populations the subject belongs to, treatment group, and type of treatment (Injection Therapy, Injectable Nerve Block, Pharmacotherapy, Cognitive Behavioral Therapy, Guided Imagery, Mirror Therapy, Psychological Therapies, Massage Therapy, Acupuncture, Other Implanted Systems, Neuromodulation, Other).

A similar listing will be presented for additional pain treatment taken from Month 3 through Month 12, given the crossover to Test treatment that occurs at Month 3.

8.5 Prior and Routine Medications

The number and percentage of FAS and ITT subjects requiring at least one additional pain treatment from randomization through Month 3 will be presented by randomized treatment overall and by medication name (as coded in the Pain Meds Codelist CRF page) and/or type of medication. All percentages are based on total number of subjects in the respective analysis set. A subject taking a given pain medication more than once will be counted once within that type/medication.

Similar analyses will be presented for concurrent medications taken from Month 3 through Month 12, with results presented separately as well as for both treatment groups combined given the crossover to Test treatment that occurs at Month 3.

Similar analyses will be presented for pain medications taken prior to randomization, with results presented by randomized treatment group (for FAS population only) and both treatment groups combined (FAS and ITT).

The data for these analyses will be presented for pain medications reported as used in the eDiary, with results presented by rescue medication treatment or routine daily medications, and by

randomized treatment group and with both treatment groups combined. These tables will be repeated in the PP population.

8.6 Baseline Medical History

Whether a condition currently exists or existed in the past will be collected for each condition below and presented by treatment group (the number and percentage of subjects *currently* experiencing the condition will be presented for each condition; the number and percentage of subjects experiencing a condition in the *past* will also be presented for each condition):

Alcohol Abuse, Anxiety, Congestive Heart Failure, Contralateral or Ipsilateral Major Amputation, Chronic Kidney Disease, Chronic Obstructive Pulmonary Disease, Depression, Diabetes, Insulin Diabetes, Non-Insulin Diabetes, Myalgia/Fibromyalgia, Myocardial Infarction, Peripheral Neuropathy, Peripheral Vascular Disease, Pneumonia, Psychiatric History, Spasticity, Substance Abuse (non-alcohol), Other.

This table will be repeated in the PP population. Additionally, Smoking History (Never smoked, Current smoker, Former smoker, Unknown, Number of Pack Years for Current Smokers, Number of Pack Years for Former Smoker) will be summarized. Other baseline conditions will be included in the subject listings.

8.7 Injection Evaluation Visit

The following is collected at the Injection Evaluation Visit prior to implantation. This may be collected more than once for a given subject if the subject fails the injection outcome on previous attempts due to inadequate pain at the Visit or inadequate pain reduction. Only information from the last Injection Evaluation Visit will be presented descriptively by randomized treatment group, unless otherwise specified below.

- Took rescue medication today (No, Yes)
- Overall Limb Pain Right Now (scale of 0-10 by intervals of 1, will be presented as a continuous variable and by categories: No pain (0), Mild (1-3), Moderate (4-6), Severe (7-9), Worst Possible Pain (10))
- Target of Saline Injection (Sciatic Nerve, Bifurcation of Tibial and Common Peroneal Nerves)
- Overall Limb Pain Before Saline Injection and at 15 Minutes Post-Injection (scale of 0-10 by intervals of 1, will be presented as a continuous variable and by categories: No pain (0), Mild (1-3), Moderate (4-6), Severe (7-9), Worst Possible Pain (10)). Percent change in pain from pre-injection to 15 minutes will also be summarized.
- Overall Limb Pain Before Lidocaine Injection and at 20 Minutes Post-Injection (scale of 0-10 by intervals of 1, will be presented as a continuous variable and by categories: No pain (0), Mild (1-3), Moderate (4-6), Severe (7-9), Worst Possible Pain (10)). Percent change in pain from pre-injection to 20 minutes will also be summarized.
- Outcome of Initial Evaluation Visit (Pass, Fail – Response to Saline, Fail – Inadequate Lidocaine Pain Reduction, Fail – Low Pain for 3 Injection Visits)

This table will be repeated in the PP population.

9 PROCEDURE

9.1 IMPLANT

Descriptive statistics of the variables specified below will be presented for the safety and FAS population subjects overall and by treatment group.

- Implanted (No, Yes)
- Months from amputation to implant
- Months from first limb pain to implant
- Procedure Time
- Number of electrodes implanted per patient (1 vs 2), overall and by amputation level (AKA, BKA)
- Electrode Nerve Location (Not Implanted, Sciatic, Tibial, CP), overall and by amputation level (AKA, BKA)
- Electrode Nerve Diameter (mm)
- Electrode Cuff Size (Not Implanted, 4-6 mm, 6-9 mm, 9-13 mm)
- Electrode Route (Not Implanted, AAL, Axillary Line [AL], PAL)
- Electrode Implanted Successfully (Yes, No)
- IPG Location (Not Implanted, Abdominal, Thigh, Buttocks)
- IPG Depth (Not Implanted, ≤0.4 cm, 0.5 – 0.7 cm, 0.8 – 1.0 cm)
- IPG Channel A Location (Not Implanted, Sciatic, Tibial, CP, N/A)
- IPG Channel B Location (Not Implanted, Sciatic, Tibial, CP, N/A)
- IPG Implanted Successfully (Yes, No)
- Any Procedural Complications or Interruptions (No, Yes)

9.2 EXPLANT

Descriptive statistics of the variables specified below will be presented for the FAS population subjects who were implanted, overall and by treatment group.

- Explanted (Yes, No)
- Electrode Nerve Location (Sciatic, Tibial, CP)
- Electrode Nerve Diameter (mm)
- Electrode Cuff Size (4-6 mm, 6-9 mm, 9-13 mm)
- Electrode Route (AAL, Axillary Line [AL], PAL)
- IPG Location (Abdominal, Thigh, Buttocks)
- IPG Depth (≤0.4 cm [≤1/8 inch], 0.5 cm, ≥0.6 mm [≥1/4 cm])
- Any Procedural Complications or Interruptions (No, Yes)
- Time to explant (0-3 Months, >3-12 Months, >12 Months)

9.3 REVISIONS

The number and percentage of FAS subjects who required at least one revision surgery will be presented. In addition, the number and percentage of FAS subjects who required at least one revision surgery due to an adverse event, due to a device deficiency, due to subject request, and due to other reason will be presented. A listing of such subjects will also be provided. Included in the listing will be treatment group, subject number, day of revision relative to implant (Day 0 is day of implant), time of first incision (HH:MM in 24-hour time), Reason for Revision Surgery, Electrode Nerve Location (Sciatic, Tibial, CP), Electrode Nerve Diameter (mm), Electrode Cuff Size (4-6 mm, 6-9 mm, 9-13 mm), Electrode Route (AAL, Axillary Line [AL], PAL), IPG Location (Abdominal, Thigh, Buttocks), IPG Depth (≤0.4 cm [≤1/8 inch], 0.5 cm, ≥0.6 mm [≥1/4 cm]), Time of Closure of Last Incision (HH:MM in 24-hour time), Explanted Parts, All Auxiliary Components, Any Procedural Complications or Interruptions (No, Yes), Narrative.

9.4 DEVICE USE

Descriptive statistics of overall device use will be presented, including at a minimum, the total number of treatment sessions administered by each FAS subject during the randomized testing phase (through Month 3) and from Month 3 to Month 12 per the IPG logs. Additional analysis of device use may be conducted, such as device use over time, number of treatment sessions per week, etc.

At each visit, the number and percentage of FAS subjects who did not use the device at all for pain management will be presented by randomized treatment group, overall and by reason of discontinuation (Sensation Too Painful or Too Strong, Does Not Provide Sufficient Pain Relief, Did Not Have Pain Requiring Treatment, Inconvenient, N/A – Not Activated, Other). Percentages will be based on number of FAS subjects. For each subject, the total number of days that the device was not used for pain management during the first 3 months and from Month 3 to Month 12 will be computed.

At each visit, the number and percentage of FAS subjects who interrupted treatment at least once with the device since last visit will be presented by treatment group, overall and by reason of discontinuation (Sensation Too Painful or Too Strong, Does Not Provide Sufficient Pain Relief, Inconvenient, Interruption Precluding Completion of the Session, Initiated Treatment by Mistake, Other). Percentages will be based on number of FAS subjects. Subjects who had >1 interruption will contribute one observation for every type of interruption.

9.5 DEVICE OBSERVATION

The number of device observations, as well as the number and percent of FAS subjects experiencing a device observation will be presented by treatment group overall, by type of observation (device deficiency, other device observation, procedural observation), by time of occurrence (Activation/Randomization, Parameter Optimization, Randomized Testing, Implant, Revision, Explant, Follow Up, Unscheduled Visit, Long Term Follow Up), by component with observation (Patient Controller, Cuff Electrode, Generator, etc.), by observation (Contamination or product packaging damage, Product damage inside package, Mechanical breakage after implant, etc.), and what action was taken (No Intervention was required, Surgical Revision, Device Explanted, etc).

10 EFFECTIVENESS ANALYSES

10.1 PRIMARY EFFECTIVENESS VARIABLE

The hypothesis for this study is that the responder rate is significantly different in the Test ($\mathbf{T}_t$) group than the responder rate in the Control ($\mathbf{C}_t$) group during the Randomized Testing phase of the study (through Month 3).

$$H_0: P_T = P_C \text{ vs. } H_A: P_T > P_C,$$

where P is the proportion of responders (responder rate).

A responder is defined as any subject who attained significant pain reduction at the end of more than half of the treatment sessions during the parallel Randomized Testing phase (i.e., through Month 3) of the study. Specifically, a responder must attain ≥50% pain reduction at the 30-minute follow-up assessment in ≥50% of the treatment sessions during the Randomized Testing phase of the study.

H_0 will be tested at a one-sided 0.025 level of significance using a logistic regression analysis to compare treatment groups while controlling for the following covariates:

- Etiology (dysvascular, trauma, other)
- Amputation location (AKA, BKA)
- Pain Type (phantom, stump, both)
- Baseline Pain Intensity (<7, 7-10)[2]
- Baseline Pain Duration (episodic, persistent)

The FAS analysis set using available treatment sessions through Month 3 is the primary analysis set for this analysis. Specifically, FAS subjects who are randomized into the Test ($\mathbf{T}_t$) or Control ($\mathbf{C}_t$) group to receive treatment but who terminate prior to their scheduled Month 3 Visit (Day 91 + 14 days post-implantation) will be determined to be a responder or non-responder based on their available data prior to termination. This is the primary analysis for effectiveness. Table 3 identifies how treatment sessions should be used based on device log and eDiary data. Treatment sessions that are interrupted with rescue (p.r.n.) pain medications will utilize the assessment of pain at the time of the interruption when available.

[2] Defined as the average of the end-of-day worst pain scores from the subject's e-diary compliant eligibility window, consistent with the study eligibility criteria.

Table 3. Rules for Identifying Device Use Sessions on eDiary for Primary Endpoint Analysis

Usage for analysis		Record on Device Log (Test / Control Group)		
		Session duration = 30 min or 8 min ± 1 min	Session duration < 30 min or 8 min	None
Report on eDiary	0 and 30 min	*Use*	*Do not use*	*Do not use*
	0 min only without rescue meds use in <30 min	*Use as Non-response (leave as missing for sensitivity analyses)*	*Do not use*	*Do not use*
	0 min only with rescue meds use in <30 min	*Use pain score at interruption, if available. Otherwise, use as Non-response.*	*Use pain score at interruption, if available. Otherwise, use as Non-response.*	*Do not use*
	0 min pain score <4	*Do not use*	*Do not use*	*Do not use*
	None	*Do not use*	*Do not use*	*Do not use*

A statistical success on the primary endpoint is achieved if the null hypothesis is rejected and the estimate of P_T is larger than the estimate of P_C.

10.1.1 Responder Criterion

The percent change of pain intensity quantified by Numerical Rating Scale (NRS) has been widely used to define a positive response in a wide range of clinical studies of devices as well as drugs. The cutoff point in percent change for a positive response is usually 50% or more reduction in pain intensity for device studies, such as spinal cord stimulation for low back pain and occipital nerve stimulation for migraine headache (Kumar et al 2007, North 2011, and Saper 2011). This 50% reduction threshold is also defined as the benchmark for defining a "substantial improvement" in IMMPACT recommendations on interpreting the clinical importance of treatment outcomes in clinical studies of chronic pain (Dworkin et al 2008).

Examples for using the 50% cutoff point are: i) a change of pain intensity from 7 to 3 recorded 30 minutes after a treatment session represents a 4/7 = 57% ≥ 50% pain reduction, therefore a successful session; while ii) a change from 7 to 5 represents a 2/7 = 29% < 50% reduction, therefore a failed session. Further, examples for responder determination are: iii) during these 8 weeks, subject X had 40 successful sessions out of a total of 56 treatment sessions, i.e., a proportion of success of 40/56 = 71% ≥ 50%, thus a responder; whereas iv) subject Y had 13 successful sessions out of a total of 28 treatment sessions, i.e., a proportion of success of 13/28 = 46% < 50%, thus a non-responder.

The responder analysis will exclude any treatment session for which duration of treatment is less than the intended complete duration (<30min (Test) or <8min (Control)), or in which baseline pain score is <4. To assess the potential effect that these treatment duration exclusions may have on results, the

average change from pre-treatment session to end of session will be calculated across all such sessions that occur for each subject. Descriptive statistics of this per-subject average change will be presented for each treatment group.

10.1.2 Study Success Criterion

Study success will be determined by a superiority test on the difference between responder rates in the Control and Test groups at Month 3. The logistic regression model treatment effect must achieve statistical significance controlling for etiology (dysvascular, trauma, other), amputation location (AKA, BKA), pain type (phantom, stump, both), baseline pain intensity (<7, 7-10) and baseline pain duration (episodic, persistent), based on the FAS, i.e., all randomized subjects who receive study treatment (Tt or Ct) will be counted in each group. A study success will be determined if the responder rate is significantly higher in the Test group than the Control group at Month 3 with an overall one-sided significance level of 0.025.

10.1.3 Secondary Analyses of the Primary Effectiveness Endpoint

The above analyses will be repeated as follows:

A. On the FAS analysis set with imputation for premature withdrawal before Month 3 as discussed in the Missing Data section above (Section 7.3).
B. ITT analysis set with imputation for premature withdrawal before Month 3 as discussed in the Missing Data section above (including imputing data for randomized subjects who are not treated) (Section 7.3).
C. PP analysis set, if >5% were excluded from the FAS.

10.1.4 Primary Effectiveness Beyond Month 3 Through Month 12

As for the primary effectiveness endpoint, a subject will be considered a responder if they attain a significant pain reduction at the end of more than half of the treatment sessions subsequent to Month 3 through Month 12. Specifically, a responder must attain ≥50% pain reduction in ≥50% of the treatment sessions during the Crossover phase of the study (Month 3 through Month 12). The endpoint will be analyzed separately by the randomized treatment group as well as combined across treatment groups. A logistic regression model including treatment will be used to evaluate treatment effect as well as the baseline covariates and the Month 3 response outcome; another logistic regression model will only evaluate the impact of baseline covariates and the Month 3 response outcome for Month 3 through Month 12 response.

10.2 Secondary Effectiveness Variables Intended for a Labeling Claim

There are 5 secondary endpoints intended for a labeling claim, and they will be compared between treatments on the FAS population. There will be no imputation of missing data for secondary endpoints due to premature withdrawal; analyses will be based on available data.

In order to control type I error, the 5 secondary endpoints intended for labeling claim will be tested in a hierarchical, gatekeeping manner in the order specified below, each at a one-sided 0.025 level of significance, and only after the primary study effectiveness endpoint has been achieved. Further, each secondary endpoint numbered 2 through 5 below will be tested for significance between

treatments only if the previous secondary endpoint in the list has been shown to be *significantly beneficial for the Test treatment.* This hierarchical closed test procedure is used to account for multiple testing with the goal of controlling the "maximum overall Type I error rate", which is the maximum probability that has been prespecified that one or more null hypotheses are rejected incorrectly, at a one-sided 0.025 level of significance.

1. Change from baseline in Opioid Pain Medication Use (Morphine Equivalent Dose (MED)) at Month 3
2. Change from baseline in Brief Pain Inventory (BPI) at Month 3
3. Change from baseline in Short Form Health Survey (SF-12) Physical Component Summary (PCS) at Month 3
4. Change from baseline in Short Form Health Survey (SF-12) Mental Component Summary (MCS) at Month 3
5. Change from baseline in EuroQol (EQ-5D) at Month 3

The following five secondary effectiveness endpoints are intended for label claim and will be tested in the order specified below. Each successful test will prompt the next subsequent test. If at any point, an endpoint analysis is not significant, then labelling will not be sought for any further endpoints; p-values will be presented as descriptive statistics.

10.2.1 Change from Baseline in Opioid Pain Medication Use (Morphine Equivalent Dose (MED)) at Month 3

Opioid pain medication use will be assessed using morphine equivalent dose (MED) (see https://www.cdc.gov/drugoverdose/pdf/calculating_total_daily_dose-a.pdf) from both rescue (p.r.n.) and routine opioid pain medications. The average daily morphine equivalent dose (MED/day) will be calculated for each subject across two weeks at Baseline and preceding Month 3. Descriptive statistics of the per-subject average daily MED at each visit and descriptive statistics of the change from baseline to Month 3 will be presented for each treatment group. The mean change from baseline in per-subject average MED/day will be compared between the Test and Control groups using analysis of variance (ANOVA) at a one-sided significance level of 0.025.

The hypothesis for this endpoint is that average change in Opioid Pain Medication use from Baseline to Month 3 is significantly less in the Test ($\mathbf{T}_t$) group than in the Control ($\mathbf{C}_t$) group.

$$H_0: D_T \geq D_C \text{ vs. } H_A: D_T < D_C,$$

where D is the mean change from baseline in per-subject average MED/day and H_0 will be rejected in favor of H_A when one-sided $p \leq 0.025$.

10.2.2 Change from Baseline in Pain Interference to Activities of Daily Living (ADL) at Month 3

Pain interference to the activities of daily living (ADL) due to limb pain will be assessed using the interference scale of the Brief Pain Inventory (BPI-Interference). The mean of the following 7 BPI items is used to calculate the BPI-Interference score: General Activity, Mood, Walking Ability, Normal Work, Relations with Other People, Sleep, Enjoyment of Life. Each item is scored on a scale

of 0 – 10, by intervals of one, where 0 indicates "Does not interfere" and 10 indicates "Completely Interferes". An individual missing >0% but <50% of the responses to these 7 items at a given visit will have missing responses imputed with the median of the remaining responses at that visit prior to calculating the 7-item average. An individual missing more than 50% of the responses at a given visit will be considered to have missing BPI-interference score at that visit. Descriptive statistics of the BPI-interference score will be presented at each visit, and descriptive statistics of the change from Baseline will be presented at Month 3, by treatment group.

The mean change in BPI-Interference summary score from Baseline to Month 3 will be compared between Test and Control groups using ANOVA at a one-sided significance level of 0.025. The hypothesis for this endpoint is that mean change in BPI interference summary score from Baseline to Month 3 is significantly lower in the Test (**T**t) group than in the Control (**C**t) group.

$$H_0: D_T \geq D_C \text{ vs. } H_A: D_T < D_C,$$

where D is the change from baseline in BPI-Interference summary score and H_0 will be rejected in favor of H_A at a one sided-significance level of 0.025.

10.2.3 Change from 12-Item Short Form Health Survey (SF-12) Physical Component Summary (PCS) at Month 3

Health-related quality of life (HR-QOL) will be assessed using SF-12 physical component summary (PCS). An individual missing responses on any PCS items at a given visit will have missing responses imputed with the Maximum Data Recovery option in the PRO CoRE scoring tool during the process of calculating the PCS score. Descriptive statistics of the PCS score will be presented at each visit, and descriptive statistics of the change from Baseline will be presented at Month 3, by treatment group.

The mean change in PCS score from Baseline to Month 3 will be compared between Test and Control group using ANOVA at a one-sided significance level of 0.025. The hypothesis for this endpoint is that mean change in PCS score from Baseline to Month 3 is significantly greater for the Test (**T**t) group than for the Control (**C**t) group.

$$H_0: D_T \leq D_C \text{ vs. } H_A: D_T > D_C,$$

where D is the change from baseline in PCS score and H_0 will be rejected in favor of H_A at a one sided-significance level of 0.025.

10.2.4 Change from Baseline in 12-Item Short Form Health Survey (SF-12) Mental Component Summary (MCS) at Month 3

HR-QOL will be assessed using SF-12 mental component summary (MCS), The analyses on this endpoint will be carried out in the same manner as the SF-12 PCS.

10.2.5 Change from Baseline in EuroQol-5D (EQ-5D) at Month 3

HR-QOL will be assessed using EQ-5D summary index. The primary variable of interest is the EQ-VAS, scored 0-100 with a higher value indicating better health. A summary of responders per five

levels (No problems, Slight Problems, Moderate Problems, Severe Problems, Unable) within each category (Mobility, Self-care, Usual activity, Pain/Discomfort, Anxiety/Depression) will be presented over time. A summary of responders by No Problems versus Any Problems within each category will be presented over time as well. The analyses on this endpoint will be carried out in the same manner as the SF-12 PCS.

10.3 Secondary Effectiveness Variables Not Intended For A Labeling Claim

The following secondary effectiveness endpoints are not intended for labeling claims. No adjustment to the significance level will be made for treatment comparisons on these endpoints. They will be compared between treatments on the FAS population (primary) and Per-Protocol population (secondary). There will be no imputation of missing data due to premature withdrawal; analyses will be based on available data.

- Responder in Control Subjects at Month 6
- Pain Relief after 2 Hours
- Pain Days per Week
- Change from baseline in Non-Opioid Analgesic Pain Medication Use through Month 12
- Change from baseline in Opioid Pain Medication Use (Morphine Equivalent Dose (MED)) through Month 12
- Change from baseline in Brief Pain Inventory (BPI) through Month 12
- EuroQol (EQ-5D) through Month 12
- Change from baseline Short Form Health Survey (SF-12) Physical Component Summary (PCS) through Month 12
- Change from baseline in Short Form Health Survey (SF-12) Mental Component Summary (PCS) through Month 12
- Patient Global Impression of Change (PGIC)
- Session Success Rate
- Composite Responder Rate (Reduction in Pain AND Absence of Increase in Medication Usage)

The following describes the analyses to be collected on each secondary effectiveness variable not intended for labeling claims. Note that only descriptive statistics (e.g., proportions, point estimates of mean, and standard deviation along with sample size) will be presented in the future product labeling for any of these secondary endpoints. P-values resulting from the analyses described below will not be included in the labeling but will be generated for internal company use and investigator publication.

10.3.1 Responder in Control Subjects at Month-6

The incidence of responder (defined as any subject with ≥50% pain reduction in ≥50% of the treatment sessions) will be presented for Control Subjects in the period from Month 3 to Month 6 (following

crossover and 3 months of treatment) and compared descriptively to the incidence of responder for Control Subjects at Month 3.

10.3.2 Pain Relief after 2 Hours

Sustained pain relief after treatment will be assessed using pain intensities reported 2 hours after each treatment session. Specifically, pain intensity will be reported immediately preceding a treatment session and again after 30 minutes and 2 hours (120 minutes) after a treatment session (pain intensity score missing at these time points will not be imputed and not included in the analysis). Percent change of pain intensity for each treatment session will be determined by the difference of pain intensity from before treatment to pain intensity after 30 minutes and after 2 hours, multiplied by 100.

Descriptive statistics of the mean percent change of pain intensity will be presented at 30 minutes and 2 hours for each treatment group across all treatment sessions through Month 3. Month 6 and Month 12. Within each treatment group, the mean percent change of pain intensity after 2 hours will be compared to the mean percent change of pain intensity after 30 minutes at a two-sided significance level of 0.05. Since subjects may contribute more than one treatment session to the analysis, this will be carried out using generalized estimating equation (GEE) linear regression within each treatment group. The dependent variable of the GEE model is the 2- hour percent change in pain intensity minus the 30-minute percent change in pain intensity. There is no independent variable; the assessment of interest is whether or not the treatment effect is significantly different from 0 at a two-sided 0.05 level of significance, which if it is, indicates a significant difference between 30-minutes and 2-hour with respect to mean percent change of pain intensity.

In addition, the mean percent change from pre-treatment (to 30 minutes and to 2 hours) in the Test group will be compared to the mean percent change from pre-treatment (to 30 minutes and to 2 hours) in the Control group through Month 3. Month 6 and Month 12 using a mixed model repeated measures (MMRM) model with terms for treatment, time point (30 minutes, 2 hours), and the treatment by time point interaction. The model will also include pre-treatment pain intensity score as a covariate and subject will be a random effect. An unstructured covariance matrix will be used for the within-subject correlation. Kenward-Rogers' approximation will be used to estimate degrees of freedom. The SLICE option will be employed to examine differences in means across treatments at each specific time point (30 minutes and 2 hours). Two-sided 95% confidence intervals of the mean difference between treatments of the percent change will be constructed using MMRM model estimates for the mean. Tests of hypotheses will be two-sided at the 0.05 significance level and will be based on the contrasts between Test and Control within the MMRM model. The specific PROC MIXED statements that will be used are:

```
PROC MIXED DATA=dataset;
CLASS usubjid trt01pn time;
MODEL chg=trt01pn time trt01pn*time base/DDFM=KENWARDROGER;
        repeated time / subject=usubjid type=UN;
SLICE trt01pn*time/PDIFF CL SLICEBY=time;
run;
```

where USUBJID is subject id, TRT01PN is treatment group indicator, TIME is indicator of time (30 minutes, 2 hours), BASE is the pre-treatment pain intensity score (replace variable names with actual variable names in the dataset), and the dependent variable CHG is the percent change of pain intensity from pre-treatment.

10.3.3 Pain Days Per Week

The number of pain days will be assessed using the intensity scale of the Brief Pain Inventory (BPI-Intensity). A pain day will be defined as a day for which the worst daily pain was moderate-to-severe (Numerical Rating Scale [NRS] ≥ 4). The number of pain days per week will be averaged across two weeks at Baseline and two weeks prior to Month 3, Month 6 and Month 12 for each subject. Descriptive statistics of this per-subject average pain days/week, and of the change from baseline in per-subject average pain days/week, will be presented at each visit within each treatment group and for both treatment groups combined.

Within each treatment group and for both treatment groups combined, the mean of the per-subject average number of pain days per week will be compared between Baseline and each of Month 3, Month 6 and Month 12 using paired t-tests at a two-sided significance level of 0.05. In addition, the mean change in per-subject average number of pain days from Baseline to Month 3 in the Test group will be compared to the mean change in per-subject average number of pain days from Baseline to Month 3 in the Control group using analysis of covariance (ANCOVA) adjusting for Baseline average number of pain days.

10.3.4 Change from Baseline in Opioid Pain Medication Use (Morphine Equivalent Dose (MED)) at Follow-up

Opioid pain medication use will be assessed using morphine equivalent dose (MED) from both rescue (p.r.n.) and routine opioid pain medications. The average daily morphine equivalent dose (MED/day) will be calculated for each subject across two weeks at Baseline and two-weeks preceding Month 3. Month 6, and Month 12. Descriptive statistics of the per-subject average daily MED at each visit and descriptive statistics of the change from Baseline to each post-Baseline visit with respect to per-subject average daily MED will be presented for each treatment group and for both treatment groups combined. The average daily MED will be compared between Baseline and each of Month 3, Month 6 and Month 12 using paired t-tests at a two-sided significance level of 0.05 within each treatment group and for both treatment groups combined only beyond Month 3 (i.e., crossover).

10.3.5 Change from Baseline in Non-opioid Pain Medication Use at Follow-up

Non-opioid pain medication use will be assessed using dose in milligrams (or equivalent unit) for both rescue (p.r.n.) and routine medications. The average daily dose in milligrams of non-opioid medications will be calculated for each reported medication type across two weeks at Baseline and two-weeks preceding each of Month 3, Month 6, and Month 12. Analyses will be carried out in the same manner as Opioid Pain Medication Use at Follow-up described above for each reported medication type, i.e., ibuprofen use will be compared between Baseline and Month 3, Month 6 and Month 12. In addition, the mean change in per-subject average daily dose from Baseline to Month 3 in the Test group will be compared to that in the Control group using ANCOVA adjusting for Baseline per-subject average daily pain medication dose for the respective medication being analyzed at a two-sided 0.05 level of significance.

Subjects using non-opioid pain medications that cannot be converted to dose in milligrams will be summarized as frequency of subjects using these medications at each of Month 3, Month 6 and Month 12.

10.3.6 Change from Baseline in Pain Interference to Activities of Daily Living (ADL) at Follow-up

Pain interference to the activities of daily living (ADL) due to limb pain will be assessed using the BPI-Interference scale. Within each treatment group, descriptive statistics of the BPI-Interference summary score will be presented at Baseline and at each of Month 3, Month 6 and Month 12; descriptive statistics of the change from Baseline will be presented at each visit. Within each treatment group and for both treatment groups combined, the mean change in pain interference will be compared between Baseline and each post-baseline visit using paired t-tests at a two-sided significance level of 0.05.

10.3.7 Change from Baseline in EuroQol-5D (EQ-5D) at Follow-up

The EQ-5D summary index will be analyzed across visits in the same manner as the Pain Interference to ADL above.

10.3.8 Change from Baseline in 12-Item Short From Health Survey (SF-12) at Follow-up

The SF-12 PCS and MCS will be analyzed across visits in the same manner as the Pain Interference to ADL above.

10.3.9 Patient Global Impression of Change (PGIC) at Follow-up

Patient global impression of change (PGIC) will be assessed using surveys conducted during office visits at Month 3, Month 6 and Month 12. The number and percentage of subjects reporting “Much Improved” or “Very Much Improved” at Month 3, Month 6 and Month 12 will be reported within each treatment group and for both treatment groups combined. The percentage of subjects reporting “Much Improved” or “Very Much Improved” at Month 3 will be compared between treatment groups (Test vs. Control) using the logistic regression model.

10.3.10 Session Success Rate

A successful session will be assessed using the NRS reported at the beginning and end of each session (30-minute follow-up). A session that results in ≥50% pain reduction will be considered a successful session. For this analysis, missing observations for a particular treatment session will be considered a failure for that session. Treatment sessions that are interrupted with rescue (p.r.n.) pain medications will utilize the assessment of pain at the time of the interruption when available. Otherwise, there will be no imputation of missing data.

The proportion of successful sessions will be calculated for each subject. Descriptive statistics of this per-subject proportion will be presented for each treatment group at Month 3. The mean per-subject proportion will be compared between treatment groups (Test vs. Control) at Month 3 using ANOVA adjusting for clinical site at a significance level of 0.05.

10.3.11 Composite Responder Rate (Reduction in pain AND Reduction in medication usage)

A composite responder is defined as any subject who attains ≥50% pain reduction in ≥50% of the treatment sessions during the Randomized Testing phase of the study (Primary Endpoint Responder) AND has either 1) a decrease from Baseline in either average daily morphine equivalent dose OR average daily non-opioid pain medication use OR 2) uses no pain medications over the period of two weeks prior to follow-up assessment. A Composite Responder must also have no addition of new opioid or new non-opioid medications over those defined in the baseline medication contract.

The proportion of Composite Responders at Month 3 will be compared between treatment groups (Test vs. Control) using the same logistic regression model as for the primary effectiveness endpoint at a two-sided significance level of 0.05.

10.4 Additional Study Outcomes

10.4.1 Technical Success

Technical success is defined as implantation and activation of the study device. The percentage of subjects and devices in which technical success is achieved (even after multiple attempts) will be summarized within each treatment group with two-sided 95% exact confidence intervals of the percentage. The number and percentage of technical success across all attempts will also be presented by treatment group. The summaries of technical success will be done in the ITT population.

10.4.2 Prosthetic Use

Prosthetic use will be assessed using the daily self-reported number of hours of prosthetic leg use in the subset of subjects who were using a prosthesis at baseline. The average hours of prosthetic use per week will be calculated for each subject using data from the two-week period before each of Baseline, Month 3, Month 6 and Month 12.

Within each treatment group and both treatment groups combined, descriptive statistics of the per-subject weekly average hours of prosthetic use and of the change from Baseline in per-subject

weekly average hours of prosthetic use will be presented at each visit. Also, within each treatment group, the mean of the per-subject weekly average hours of prosthetic use will be compared at Month 3, Month 6 and Month 12 to Baseline at a two-sided significance level of 0.05 using paired t-tests. In addition, the mean change in per-subject weekly average number of hours of prosthetic use from Baseline to Month 3 in the Test group will be compared to that in the Control group using ANCOVA adjusting for study center and Baseline per-subject weekly average prosthetic use. The summaries of prosthetic use will be done in the ITT population.

11 SAFETY ANALYSES

The IPA will review and adjudicate all deaths, all procedure related SAEs, any SAEs associated with the target limb or abdominal implant site, and any treatment or device related SAEs. All analyses of safety data will utilize the IPA-adjudicated data where available. Otherwise, investigator determinations of relatedness will be used. Adverse Event rates will be monitored for the entire duration of the study by an independent Data Monitoring Committee (DMC) comprised of medical and statistical expert reviewers. All safety analyses will be carried out on the Safety Analysis Set unless otherwise specified.

11.1 PRIMARY SAFETY VARIABLE

The primary safety endpoint is the incidence of all serious adverse events, including Serious Adverse Events (SAEs), Serious Adverse Device Events (SADEs), and Unanticipated (Serious) Adverse Device Events (UADE), starting or worsening from the time of implant through three (3) months post-implant. The primary safety endpoint will be analyzed at the conclusion of the Randomized Testing phase of the study, after all active subjects complete the Month 3 visit. The number and percent of subjects with at least one SAE will be presented overall and by adverse event type (abscess, allergic reaction, etc., as indicated on the adverse event CRF page) within each treatment group. The two-sided exact 95% confidence interval of the percent of subjects with at least one SAE will also be presented. Adverse events will be additionally presented in the same manner for Month 3 through Month 12

The number and percent of subjects experiencing at least one SAE related to each of procedure, device, treatment, or “other” (for example, disease progression, medication related) will be presented within each treatment group overall and by adverse event type. The two-sided exact 95% confidence interval of the percent of subjects with at least one SAE related to each of procedure, device, treatment or “other” will also be presented. The rates for the “other” category will be descriptively compared to the rates seen in Conventional Medical Management (CMM) of Amputation Pain to better understand which AEs can be ascribed to treatment or device effects. Table 3 provides a breakdown of how procedure-related, device-related, treatment-related, and “other”-related are defined.

Table 3. Primary Safety Endpoint SAE Categorization

Event Categories
• Procedure related, occurring intra-operatively or within 14 days post-operatively.
• Device (hardware) related
• Treatment related, resolved after treatment is stopped.
• Other, for example, disease progression or medication-related.

11.2 Additional Safety Parameters

The number and percent (and two-sided exact 95% confidence interval of the percent) of subjects with at least one adverse event (AE) including non-serious AEs, non-serious adverse device effects (ADE), SAEs, SADEs, and UADEs, that started or worsened from time of initiation of implant procedure through Month 12, will be presented. The analyses described in the previous section on SAEs through Month 3 will be repeated for all adverse events from Month 3 through Month 12, with results also presented for both treatment groups combined. In addition, the number and percent (and two-sided exact 95% confidence interval of the percent) of subjects with at least one AE from Month 3 thorough Month 12 will be summarized by treatment group and both treatment groups combined on the following subgroups of adverse events, overall and by adverse event type:

- Serious; (SAE/SADE/UADE)
- Non-serious; (AE/ADE)
- Serious Device-or-Procedure-related
- Serious Procedure-related
- Serious Device-related
- Treatment related –permanent
- Treatment related – reversible

Events will be considered permanent if the adverse event does not resolve by Month 12 despite medical management, changes in device programming, or stopping treatment**.**

In addition, the number and percent of subjects (and two-sided exact 95% confidence interval of the percent) with AEs and SAEs occurring post informed consent through initiation of implant procedure will be presented overall and by type of AE. This assessment will be reported for both the Enrolled-not-Implanted and the Enrolled-and-Implanted populations.

11.3 Deaths

Should any subjects die during the course of the QUEST trial, relevant information (including treatment group, SAE leading to death, study day of death relative to implant) will be supplied in a data listing.

12 OTHER PLANNED ANALYSES

12.1 PLANNED SUBGROUP ANALYSES

Exploratory data analysis comparing treatments on the primary effectiveness endpoint and on the secondary endpoints for labeling will be performed within each of the following subgroups for the FAS analysis set without data imputation. The purpose of this analyses is not to detect a significant treatment effect within each subgroup, but rather to assess consistency of treatment comparison results across subgroups. To that end, treatment-by-subgroup interaction will also be assessed for each subgroup variable below using the Breslow-Day test for categorical variables, and an ANCOVA model adjusting for subgroup, treatment and the subgroup by treatment interaction for continuous variables.

- Etiology (dysvascular, trauma, other)
- Amputation location (AKA, BKA)
- Pain Type (phantom, stump, both)
- Baseline Pain Intensity (5-6, 7-10)
- Baseline Pain Duration (episodic, persistent)
- Protocol version (v1.* [v1.5, v1.7, v1.8, v1.9], v2.0, v2.1)
- Number of electrodes implanted (1, 2)
- Prosthetic (user, non-user)
- Age (less than 55, greater than 55)
- Sex (male, female)
- Race (White, black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Number of Treatment sessions during randomized phase (above median, at or below median)
- Timing of administration of treatment (above median, at or below median)
- Treatment Session Duration (for session-based analysis: 0 min, <30 min, 30 min)
- Proportion of Pain episodes in which treatment is initiated during randomized phase (above median, at or below median)
- Timing of Pain episodes in which treatment is initiated (month 1 vs. month 2 post randomization)
- Pain score at initiation of treatment (for session-based analysis) Rescue Medication (day of/prior to treatment, during treatment)
- Use of Non-Contract Medication (yes, no)
- Baseline lidocaine injection concentration (1%, 2%)

12.2 COMPLIER-AVERAGE CAUSAL EFFECT (CACE) ANALYSIS.

The CACE estimates the causal effect for compliers, under certain assumptions – for example, assuming that the proportion of "always compliers" is the same in the control group as it is in the treatment group. This analysis uses the entire FAS dataset. In its basic form, the CACE treatment effect = FAS / Pc, where Pc is the proportion of compliers in the treatment group and FAS is the

treatment effect under full analysis set as described above. Specifically, the estimate of the difference in responder rate between randomized treatments for the FAS analysis set will be divided by Pc. This will be done for the FAS effect size based on available data and based on imputed data.

12.3 Frequency and Timing of Treatment Analysis

Using the FAS population, multivariable logistic regressions will be conducted as supportive and informational analyses to examine the effect of the number of treatments, and the timing of their administration on the primary endpoint adjusting for the number of treatment sessions. The first logistic model will include an interaction term for the number of treatment sessions by treatment group. The second logistic model will include an interaction term for the number of treatment sessions in the first month by treatment group. The third logistic model will include an interaction term for the number of treatment sessions in the first two months by treatment group.

12.4 Blinding Analysis

A 3-tiered auxiliary questionnaire will be administered for evaluating blinding (masking). The study subjects will state their belief regarding their group assignment.

Subject unblinding shall be evaluated based on the expected frequencies of the subject's responses to the questionnaire. The expected frequencies shall be derived from the questionnaire's 3x2 frequency table where columns denote the actual treatment provided to the subject and rows denote the subject's belief regarding treatment. The frequency table is summarized in Table 4.

Table 4. 3x2 frequency table for 3-tiered blinding questionnaire.

Believed Treatment (i.e., guessed)	Actual Treatment	
	Test (Tt)	Control (Ct)
Test (Tt)	nT|T	nc|T
Control (Ct)	nT|c	nc|c
Don't Know (dk)	nT|dk	nc|dk

The questionnaire will be administered at the conclusion of the Randomized Testing phase. In addition, since a subject may become aware of their group assignment by virtue of observing the effects of treatment (or lack thereof), the questionnaire will also be administered shortly after commencement of the Randomized Testing phase and intermittently throughout the study. Successive assessment of the subjects' blinding will allow for characterization of therapeutic unblinding by means of multivariate analysis with time.

Several indices will be used to investigate the quality of the blind. The James Blinding Index (BI) will be used to assess the blinding of the study overall. The James BI is a modification of Cohen's kappa (K), a commonly used method to assess observed agreement (i.e., between Believed and Actual Treatment) verses agreement expected by chance. The James BI is designed specifically for the situation of blinding assessment, where correct guesses support unblinding and therefore assigned a

weight of 0, incorrect guesses are moderately supportive of blinding and therefore assigned a weight of 0.5, and "Don't Know" responses are in fact most supportive of blinding and therefore assigned a weight of 1. The James BI can attain values in the interval [0, 1] with higher values denoting increasing levels of blinding. Therefore, BI = 1 indicates perfect blinding and BI = 0 an unblinded study.

The above frequency table can also be used to calculate the other most prominent blinding index, the Bang BI, which addresses several critiques of the James BI. In contrast to the James BI, the Bang BI gives less weight to "Don't Know" responses and more to decisive responses. It is also treatment-arm specific, so it can detect different levels of blinding for subjects randomized to Test (Tt) and Control (Ct) treatment groups. Finally, it is sensitive to "reverse unblinding" in which subjects consistently guess the incorrect treatment assignment.

Descriptive statistics of the James BI and the Bang BI will be provided at randomization commencement, Month 1, Month 3 and Month 6, by treatment group and for both treatment groups combined.

Appendices