Biomedical Statistical Consulting

QUALITY IN DESIGN, IMPLEMENTATION, & PRESENTATION SINCE 1986

Statistical Analysis Plan

Applied VR HEOR Trial

A Decentralized, Randomized, Controlled Trial of
EaseVRx-8w+ for Chronic Low Back Pain

Sponsor:

AppliedVR Inc.
16760 Stagg Street, Unit 216
Van Nuys, CA 91406
Phone: 1-844-204-9093

Version: 2.0
June 15, 2022

Author:

David Maislin

TABLE OF CONTENTS

Statistical Analysis Plan Approval

Title: Applied VR HEOR Trial
Version: 2.0 | June 15, 2022
Author: David Maislin

Signature: Date:

This Statistical Analysis Plan has been reviewed by MCRA, LLC:

Name:

Position:

Signature: Date:

The Statistical Analysis Plan has been reviewed and approved by the Sponsor:

Name:

Position:

Signature: Date:

REVISION HISTORY

- Version 1.1 of this document has modified the randomization ratio from 2:1 to 1:1 and increased the maximum sample size to N=1000 total subjects. This changed was due to inadvertent 1:1 randomization being used rather than the 2:1 specified within version 1.0 of this document.

1 STATISTICAL ANALYSIS PLAN

1.1 COMMENT ON STUDY PROTOCOL

Please refer to the current Study Protocol for additional details on the overall design and conduct of the HEOR study. Where relevant to facilitate understanding, sections of the Study Protocol have been reproduced or referenced within the SAP.

1.2 OVERALL OBJECTIVE

The primary objective of the present study is to demonstrate the efficacy of the investigational product (EaseVRx) compared to control (ShamVR) in humans with Chronic Lower Back Pain (defined as moderate to severe pain lasting longer than three months) at screening. The primary endpoint is the absolute change in Pain (either Intensity or Interference) at Week 8 post treatment. Short term follow-up (weeks 1 through 8) and long-term follow-up (>Week 8 through Year 2) will be evaluated secondarily along with other pre-specified secondary endpoints. Finally, this study includes an economic evaluation using within and between study matching through insurance and claims databases.

1.3 STUDY DESIGN

This is a study evaluating the efficacy of the investigational product, EaseVRx, using multiple 3D VR scenes compared to a Sham 2-D VR device. The primary effectiveness analyses will compare groups of participants receiving EaseVRx versus Sham in the context of a mixed model for repeated measures (MMRM). Primary analyses will use a multiplicity testing approach given the two key primary endpoints, the change in Pain Intensity and the change in Pain Interference at the end of Week 8 (i.e., on Day 56). Only one of the endpoints need to be met for the study to claim overall study success. Key secondary endpoints include the change in Pain Intensity and Interference through a responder analysis where importance to the patient is defined to be a 30% improvement. Supplemental analyses are specified to examine Pain Interference and Intensity at follow-up visits adjusting for baseline pain, and another responder definition which defines a 1- and 2-point change in Interference and Intensity, respectively.

1.4 CONSIDERATIONS OF CONTROL GROUP

In compliance with VR-CORE clinical trial guidelines, an active control is selected that utilizes non immersive, 2D content within a VR headset as a rigorous VR placebo. The Sham VR headset will display 2D nature footage (e.g., wildlife in the savannah) with neutral music that was selected to be neither overly relaxing, aversive, nor distracting. The experience of Sham VR is like viewing nature scenes on a large-screen television and is not interactive. Sixteen videos will be rotated over the 56 sessions, with average duration of sessions closely matching those of EaseVRx.

Figure 1

1.5 RANDOMIZATION

To minimize bias in comparisons of experimental groups, subjects will be randomized to experimental condition (EaseVRx or Sham) in a 1:1:1:1 ratio. This approach assures relatively equal numbers of subjects in each experimental group over the course of the study while also minimizing the likelihood that study investigators will be able to infer the next randomization assignment in the sequence.

- 56-day pain relief skills VR program (EaseVRx)
- 56-day control VR condition (Sham VR)
- 56-day pain relief skills VR program followed by 56-day on demand availability of the pain relief skills VR experiences (EaseVRx plus extended on demand library)
- 56-day control VR program followed by 56-day on demand availability of the control VR program VR experiences (Sham VR plus extended on demand library)

The Curebase platform will be used to administer the blinded randomization program and ensure equal allocation to each group.

1.6 BLINDING

The primary endpoint of this study will be collected via Electronic Data Capture and securely transmitted to the Database. Curebase will maintain the database and be responsible for all aspects of data management.

Participants, study statisticians and investigators will be blinded to treatment group assignment until after the last subject contributes their primary outcome at week 8. Prior to the primary analysis, the database will be officially locked and signed off by the Sponsor and key stakeholders. Study participants will remain blinded to treatment group assignment until study completion.

1.7 TIMING OF EVALUATIONS

Participants in all 4 cohorts will be evaluated at the following timepoints: pre-treatment, 2x/week during the treatment, immediately post-treatment (day 56 in all arms), 4-weeks post-treatment, 8-weeks post-treatment, 12-weeks post-treatment, 6 months post-treatment, 12 months post-treatment, 18 months post-treatment, 24 months post-treatment.

1.8 STUDY ANALYSIS SETS

1.8.1 INTENT TO TREAT ANALYSIS SET (ITT)

All subjects randomized will be included in the intent-to-treat (ITT) Analysis Set. The ITT analysis will only be considered a supportive and complementary analysis.

1.8.2 MODIFIED INTENT TO TREAT ANALYSIS SET (MITT)

For the primary analysis and safety evaluations, the mITT Analysis Set will be defined and presented as primary. The mITT Analysis Set includes all enrolled subjects who actually receive the device. If all enrolled subjects receive the device, the mITT and ITT Analysis Sets will coincide and will be referred to as the ITT Analysis Set.

1.8.3 PER PROTOCOL ANALYSIS SET (PP)

The primary endpoint and selected secondary endpoints will be repeated in the Per Protocol (PP) Analysis Set. The PP analysis set will exclude from the mITT analysis set subjects who did not receive a device, who have clinically significant violations of inclusion or exclusion criteria or who have post-treatment protocol violations that may reasonably be predicted to impact on effectiveness endpoints. For the purpose of defining this analysis set, treatment compliance is defined as using the device at least 42% within 8 weeks of treatment. Subjects using the device less than 42% will be excluded from this Analysis Set. Consistency with the mITT analysis will be evaluated and described in the Clinical Study Report.

1.8.4 COMPLETERS ANALYSIS SET (CC)

The primary endpoint and selected secondary endpoints will be repeated among those subjects completing the 56 days of treatment and have their final follow-up evaluation ("8-week outcome".) The CC analysis set will exclude from the mITT analysis set subjects who are missing their primary 8-week Pain evaluation. Consistency with the mITT analysis will be evaluated and described in the Clinical Study Report.

1.8.5 MATCHED ANALYSIS SET (MAS)

A Matched Analysis Set (MAS) is defined and described in the standalone Health Care Economics Protocol drafted by the Sponsor.

1.8.6 TRAINING CASES (TC)

There are no training subjects for this protocol.

2 EFFICACY EFFECT ANALYSES

2.1 DESCRIPTIVE ANALYSIS

Observed values within the four (4) experimental groups and differences among groups in primary and secondary endpoints will be summarized using descriptive statistics, including means, standard deviations, medians, minimum, and maximum values for continuous data, and frequencies and percentages for categorical data. 95% Confidence or Credible Intervals will be provided for all analysis and provided with clinical interpretations in the Clinical Study Report. Graphical analyses of most variables will be provided with summary statistics including standard error or posterior interval markings.

2.2 PRIMARY ENDPOINT

The primary study endpoint is the change in absolute Pain Intensity **or** Pain Interference at 8 weeks post treatment from baseline (i.e., pre-treatment). Evaluation of the primary endpoint will utilize both EaseVRx (EaseVRx-8w and EaseVRx-8w plus extended on-demand library) and ShamVR (Sham VR-8w and ShamVR-8w plus extended on-demand library) arms with the two EaseVRx and two Sham VR groups being combined resulting in a two-group comparison. This is justified given that differences between the arms only occur after 8 weeks when primary endpoint data will be collected and locked in the database.

2.3 PRIMARY HYPOTHESIS

This trial contains multiple primary endpoints (MPE). Study success will be declared if at least one endpoint is statistically significant correcting for multiplicity.

The primary effectiveness hypothesis is superiority of the investigational device relative to control in terms of the mean improvements (or reduction) in Pain Intensity **or** Pain Interference from baseline to Week 8.

The primary null and alternative hypotheses of superiority may be represented symbolically as:

Ho: $\delta(pain)_{EaseVRx}k - \delta(pain)_{Sham}k <= 0$, for all k

Ha: $\delta(pain)_{EaseVRx}k - \delta(pain)_{Sham}k > 0$, for at least one k

Where **$\delta(pain)_{EaseVRx}k$** and **$\delta(pain)_{Sham}k$** are the mean change (or reduction) from Baseline at Week 8 for EaseVRx and Sham, respectively. K indexes the two unordered Primary Endpoints, Pain Intensity and Pain Interference. When K is greater than 1, type 1 error increases. To control type 1 error, multiplicity adjustment is necessary and therefore is prespecified and described below.

2.4 TESTING THE PRIMARY SUPERIORITY HYPOTHESIS

Superiority testing will be performed using a mixed model for repeated measures (MMRM) analysis of covariance (ANCOVA) Model. The MMRM approach is a direct likelihood

approach requiring specialized statistical software. For this study, all MMRM parameters will be estimated using appropriate statistical software (see Software below). The MMRM model is notable for its inclusion of all available data from all eligible subjects and does not require their exclusion as in complete case analysis or arbitrary assignment of some value as in Last Observation Carried Forward (LOCF). Inclusion of outcome data from time points earlier than Week 8 informs the implicit imputation of missing Week 8 values through the covariance matrix used to model the random effects.

The random effects in an MMRM model are based on the so-called 'unstructured covariance matrix' in which variances over time are allowed to vary as are the pair-wise correlations over time.

The values over time include all Pain scores, namely at each and all of the 2x week follow-up time points through Week 8.

All Pain values will be included in the analysis even if they are obtained out-of-window. The fraction of out-of-window visits will be summarized at each time point.

The contrast indicated by the primary hypotheses above, $\boldsymbol{\delta(pain)_{EaseVRx} - \delta(pain)_{Sham}}$ will be estimated as the treatment group contrast from baseline to Week 8 in the Pain score derived from the MMRM. The null hypothesis is that the true value of this contrast is equal to zero. The same approach will be taken for testing both Pain Interference and Pain Intensity.

2.4.1 COVARIATES IN THE MMRM

The MMRM model must contain an indicator variable for treatment group, a categorical time factor, and treatment group by time interaction. It is also necessary to include the baseline value of the Pain outcome variable (either pain or function) so that significance levels will apply equally to values over time and to changes from baseline.

It is often considered useful to include additional baseline covariates to further reduce potential bias from missing values. Candidate variables include age, sex, and BMI.

Exploratory analyses will be provided that add (or subtract) additional baseline covariates to the model in exploratory analyses only if any of these variables appear with clinically significant imbalance between treatment groups. The purpose of these analyses is to determine the extent to which such imbalance impacts on estimated treatment group differences and associated significance levels.

Should this model not converge, the primary outcome will be determined using a simplified covariance structure and only include treatment, group, and treatment group interaction. In this case, the Clinical Study will report the convergence diagnostics without additional multicity adjustment arising from model switching.

2.4.2 DETAILS OF THE MMRM

As stated above, the covariance matrix for MMRM is an unstructured covariance matrix in which each of the variances (at every follow-up) are free to vary as are the covariances.

That is, the covariance matrix for the MMRM will not include any random effects per se, but accounts for correlations among errors by specifying the form of the covariance matrix. The MMRM employs an unstructured covariance to produce inferences that are valid under the so-called 'missing at random' (MAR) assumption which is more generally true than the 'missing completely at random' assumption that is required for validity of analyses restricted to complete cases.

The longitudinal model will be specified as a repeated measures model that expresses the Pain score as a linear function of treatment, time, treatment-by-time interaction, and a covariate of the baseline pain measurement.

The model is mathematically expressed as:

$\mathbf{Y_i = X_i\beta + e_i}$

Where:

- $\mathbf{Y_i}$ is change from baseline to Week 8 in Pain
- $\boldsymbol{\beta}$ is a vector of fixed-effect regression parameters which includes:
    - μ, the overall mean
    - θ, the treatment effect
    - $\boldsymbol{\tau}$, a vector of post-baseline time effects
    - $\boldsymbol{\eta}$, a vector of treatment-by-time interaction effects
    - $\boldsymbol{\varphi}$, a vector of covariate effect parameters that only includes baseline Pain score but in general could include additional baseline variable
- $\mathbf{X}$ is a design matrix for the fixed effects, and
- $\mathbf{e}$ is the error vector with
    - $E(\mathbf{e}) = \mathbf{0}$ and
    - $Var(\mathbf{e})=\sigma_e^2 V_{unstructured}$

The specific model to be used may be expressed as:

$\mathbf{Y = b_{int} + b_{base} + b_T + b_1 + \ldots + b_{56} + b_T b_1 + \ldots + b_T b_{56}}$

With parameters as explained below.

Parameter	Represents
Y	Change from baseline
b_{int}	Model intercept
b_{base}	Baseline value
b_T	Indicator for Active Treatment
$b_{1, ..., } b_{56}$	Indicator for each timepoint, subscripts index categorical visit with 1 being the first day of treatment, and 56 being the final day of treatment.
$b_T b_1$, …, $b_T b_{56}$	Indicators allowing varying treatment differences at each timepoint

From this model estimates of treatment effect at specific visits and specifically at Week 8 difference may be obtained as follows.

$$\mathbf{T_{W8} = b_{int} + b_{base} + b_T + b_T b_{56}}$$

$$\mathbf{S_{W8} = b_{int} + b_{base}}$$

$$\mathbf{\Delta_{W8} = T_{56} - S_{56} = b_T + b_T b_{56}}$$

Where T_{56} and S_{56} are the expected changes from baseline to Week 8 in the Pain scores in the active and sham control groups, respectively. In terms of these model parameters, the superiority hypotheses are:

H_0: $\Delta_{W8} = 0$; H_1: $\Delta_{W8} > 0$ when larger values reflect improvement or

H_0: $\Delta_{W8} = 0$; H_1: $\Delta_{W8} < 0$ when smaller values reflect improvement

2.5 TESTING THE SECONDARY ENDPOINTS

Analyses of secondary endpoints will focus on presenting between group differences and 95% confidence intervals from statistical models appropriate for the distribution of the endpoint of interest. For continuous endpoints with repeated measurements (e.g., Change in Pain scores at all timepoints post-dosing), analyses will be performed utilizing the MMRM approach as the primary analysis. The endpoints will also be evaluated using the Values over time adjusted for baseline differences approach as the primary analysis.

Endpoints assessed at a single timepoint will be compared using consistent approaches applied in the context of appropriate methods for cross-sectional data, including linear regression for continuous secondary endpoints.

Statistical significance for the following secondary endpoints will be determined based on a Hochberg step-up procedure (see below).

- ODI
- PROMIS sleep disturbance
- PROMIS anxiety
- PROMIS depression

All other endpoints may be provided with nominal p-values and reported as nominal in the Clinical study report and publications.

All secondary endpoints will include nominal 95% Confidence or Credible Intervals to allow for clinical interpretation in the Clinical Study Report. Analyses will be reported in both tabular and graphical form.

2.6 RESPONDER ANALYSIS JUSTIFICATION

If the study obtains success around the primary endpoint, we will conduct supplementary analysis to evaluate the sensitivity of these study results using supplemental responder analyses. Specifically, we will evaluate whether a greater portion of participants who received active treatment had clinically important differences (CIDs) compared to those in the sham arm. CID thresholds are commonly used to answer the question of whether statistically significant differences in groups are driven by meaningful improvements in individuals. Broadly, there are two classes of CIDs for evaluating change from baseline. One approach is based on percent change in score; the other is based on specified point change in score. We will conduct responder analyses using both approaches. The table below the responder definitions for each measure Brief Pain Inventory (BPI) Measure (Average Pain and Pain Interference) and each Supplementary Analysis.

Table 1: Responder Definition for Brief Pain Inventory Measures of Average Pain and Pain Interference

	Responder Analysis 1	Responder Analysis 2
BPI Average Pain	30% decrement in score compared to baseline	Decrease of 2 points on 0-10 scale (one item)
BPI Pain Interference	30% decrement in score compared to baseline	Decrease of 2-point on 0-10 (average of 7 items scored 0-10)

The choice of the above thresholds is based on research precedent and recommendations. Regarding the 30% threshold, the IMMPACT consensus statement concluded that changes of 30% to 36% represent 'much better,' 'much improved,' or 'meaningful' decreases in chronic pain", p. 111 (Dworkin, 2008). With respect to single item numerical rating scales (NRS) such as BPI Average Pain, another relevant study was conducted by Farrar and colleagues (Farrar, 2000). They evaluated a range of percentage change thresholds for a 0-10 NRS: ≥10, 15, 25, 33, 40, 50, 60, 75.5 and concluded that a threshold of 33% had the best balance of specificity, sensitivity, and accuracy. Subsequently, studies cited Farrar to support a 30% threshold, even though Farrar did not specifically evaluate a 30% threshold. However, on a 0-10-point scale, the point changes are identical with either a 30% or 33% threshold. Another study of relevance is one by Ostelo and colleagues conducted in the context of low back pain (Ostelo, 2005). This study concluded that a 23% change in NRS score was a reasonable threshold for a minimally clinically important change for patients with chronic

low back pain. A 30% criterion has been used for scores based on the Brief Pain Inventory (BPI) short form (Deere, 2021) and for the BPI Pain Interference subscale (Bomalaski, 2018). Another study used the 30% criterion for scores on the BPI Sleep Interference Scale (BPISI) scale (Mehta, 2016). A study in patients with fibromyalgia estimated a slightly higher CID threshold of 32.3% for BPI Average Pain (Meese, 2011, p. 824).

Studies exploring and applying point-based CID thresholds for BPI Average Pain and BPI Pain Interference typically use 1-2 points. Based on prior work, Krebs and colleagues estimated a 1-point CID threshold for both BPI Average Pain and Pain Interference (Krebs, 2010). They then used these estimates in a randomized clinical trial of pain function in patients with chronic back pain or hip or knee arthritis (Krebs, 2018). In contrast, Mease and colleagues estimated clinically important change in BPI Average Pain and BPI Interference scores in a sample of patients with fibromyalgia and derived a CID value for both measures of ≈ 2 score points (Mease, 2011).

Given the amount of evidence and precedence for the 30% criterion for measures of pain severity and interference including those of the BPI, we believe a 30% score change is an appropriate threshold for clinically meaningful change in scores. For the score point CID criterion, an improvement of either 1 or 2 points on the 11-point scales could be justified. We chose to go with the more stringent 2-point CID for both measures.

2.7 EXPLORATORY HEALTHCARE UTILIZATION AND ECONOMIC ANALYSIS

This section provides a high-level summary of the Health Care Economics portion of this trial. The following text is taken from the Sponsor's summary. Please consult the standalone Health Care Economics Protocol for additional details of these analyses.

> *To assess resource use associated with chronic low back pain, participants will be run through a secure Datavant matching process to be linked to their healthcare claims data if it is available in the Komodo data set. The claims will provide descriptive characterizations of resource use and costs associated with chronic low back pain. For patient-level changes in resource use, and for comparisons between study arms, only the participants with complete claims files will be assessed. Claims-matched data will capture all interactions with the healthcare system that generate an insurance claim record, which includes such things as physician visits, interventions such as steroid injections, surgery or physical therapy, emergent use of services, medications, and diagnostic procedures. Claims adjudication often lags a few months, so the most robust analyses of utilization will be performed about 6 months after the time point of interest. Health economic modeling will be performed after 1 year to assess the cost and economic outcomes of the EaseVRx intervention.*
>
> *The Komodo claims data provides a real-world source of utilization and cost parameters associated with claims of chronic low back pain. As such, it can be used more broadly to either compare the study participants' resource use and*

costs to the broader population of patients with chronic low back pain, or a "synthetic arm" can be constructed using available demographic data to match any of the study arms.

2.8 SAMPLE SIZE

2.8.1 MULTIPLE PRIMARY ENDPOINT POWER ANALYSIS

Clinical trial simulation was used to derive the statistical operating characteristics of this trial (FDA Guidance 2010). These simulations provide calculations of Type I error, power, and sample sizes for the primary superiority study claim under the default reference assumptions and a range of sensitivity analyses to ensure the trial is conservatively designed.

There are many ways to generate simulated clinical trials for the use in statistical power calculations. Regardless of the method, the simulated data have no impact on the actual trial. For this trial, simulated datasets were generated by drawing from random bivariate normal distributions with specified means, SDs, and correlations between the two primary endpoints, Pain Intensity and Pain Interference. The model parameters (the mean, SD, and correlations) were based on the summary statistics taken from the previous EaseVRx trial to support the initial default assumptions.

Note, the previous trial measured interference in four domains from the DoD/VA Pain Scale, Sleep, Mood, Activity, and Stress. The future trial will collect these measurements as an average of 7 domains through the BPI survey. For sample size estimation, the average of the 4 DoD/VA Pain Scale domains collected in the previous trial will be used as an estimate for the BPI survey. As will be discussed below, the sample size calculations have been appropriately made more conservative to deal with possible differences arising from the surveys.

Unadjusted results from the previous trial are shown in the following table.

Table 2: Previous Trial Results for Power Analysis

	EaseVRx						Sham					
		Intensity		Interference*		Cor		Intensity		Interference*		Cor
	N	Mean	SD	Mean	SD	Rho	N	Mean	SD	Mean	SD	Rho
Baseline (BL)**	94	5.10	1.24	4.78	1.91	0.72	94	5.24	1.14	5.08	1.64	0.71
Week 8 (W8)	84	2.98	1.90	2.26	2.13	0.81	84	3.95	2.03	3.24	2.31	0.85
Change (BL to W8)	84	-2.12	1.64	-2.56	1.99	0.76	84	-1.21	1.68	-1.82	1.97	0.77

*Calculated as mean of Mood, Activity, Stress and Sleep.
** Refers to data collected at relative day -7, which is the average of 3 timepoints between -14 and 0 days before first treatment.

This table shows that the mean changes for Pain Intensity are -0.91 higher for EaseVRx, and the mean change for Pain Interference is -0.74 higher for EaseVRx.

Using the observed data in the previous trial and shown in the table immediately following this paragraph, clinical trial simulation was used to generate thousands of future trials where the group level summary statistics (means, SDs and correlation between endpoints) matched observed data. For every trial, two sample t-tests were performed and statistics from each recorded. Multiplicity adjustment was applied to each of the p-values (testing the group difference < 0). Each trial matrix was discretized into either success or failure with every trial being assumed a success if either p-value was less than 0.025. The binary trial success from 10,000 trials was stored, and the fraction of all simulated trials equal to the trial's power.

In this framework, the previously observed data is used in the assumptions. It is assumed that the difference observed in the last trial is clinically meaningful and worth detecting. That is, is the group difference in Intensity of -0.91 (EaseVRx= -2.12 vs. ShamVR= -1.21) and Interference of -0.74 (EaseVRX= -2.56 vs. ShamVR=-1.82) clinically meaningful and worth studying. The current design assumes that it is worth detecting.

Note that the assumed correlation at 0.76 and 0.77 does not significantly impact the total sample size in this study and is generally understood to be conservative. If the endpoints were entirely not correlated, power would go up. When the endpoints are assumed to be perfectly correlated, power drops, but only ~1.5%.

Before LTF and Matching adjustment, with 240 total subjects (120 and 120 in the 1:1 randomization) provide >95.0% power to detect either an improvement in Intensity or Interference with a two-sample t-test testing the difference in means is less than 0 for the

week 8 change value. The p-values are adjusted for multiplicity using a Hochberg correction. Study success is defined and will be declared if either p-value is < 0.025.

The trial size of N= 240 does not include adjustment for LTF or the expected match rate. When the Claims match rate is only 50% and Lost to Follow-up is 15%, the study sample size is N= 565. When the match rate is 33% and Lost to Follow-up is 15%, the study sample size is N= 855.

- 240 / .500 = 480 accounting for 50% matching only
- 480 / .850 = ~565 accounting for 50% matching + LTF
- 240 / .330 = ~730 accounting for 33% matching only
- 730 / .850 = ~855 accounting for 33% matching + LTF

The sample size is further increased to allow for testing of the secondary endpoint of 30% improvement in either Pain Interference or Intensity at longer time points. As this study extends to 24 months' follow up, the Lost to Follow-up rate for the secondary endpoints may exceed 15%. Given these two factors (less power 30% responder analysis and possible additional LTF), the trial may enroll up to 145 subjects beyond the minimum sample size for a total study size of N= 1000.

2.8.2 COMMENT ON STATISTICAL POWER IN THIS TRIAL

- It is noted that this trial is statistically powered with 240 subjects before adjusting for Lost to Follow-up or adjustment for matching. Therefore, this trial may detect differences that are small and therefore not meaningful for any individual patient. To prospectively guard against capitalizing on differences not necessarily important to patients, a set of responder analysis are specified so that proportional differences between the groups will be provided such that each subject will only contribute "success" when they have individually achieved a meaningful improvement.

2.9 HANDLING OF MISSING DATA AND SPECIFICATION OF THE ESTIMAND

Primary effectiveness data may be unavailable for several reasons. This section describes the implications and handling of unavailable primary outcome data on the estimand of interest. For this study, the estimates of interest are the group mean differences in changes from baseline to 8 weeks in Pain Intensity and Pain Interference had subjects completed 8 weeks of follow-up and not withdrawing for lack of treatment effectiveness.

The long term estimands of interest are the change in Pain Interference and Pain Intensity at 4 weeks post treatment, 8 weeks post treatment, 12 weeks post treatment, 6 months post treatment, 12 months post treatment, 18 months post treatment and 24 months post treatment.

2.9.1 HANDLING DATA FOLLOWING SUBJECT WITHDRAWAL DUE TO LACK OF TREATMENT EFFECTIVENESS

Discontinuation due to lack of treatment effectiveness may occur. This discontinuation is an "intercurrent event" in the sense of Draft Guidance E9(R1). "[I]ntercurrent events such as discontinuation or switching of treatment, or use of rescue medication, may in some circumstances render the later measurements of the variable irrelevant or difficult to interpret even when it can be collected." It is necessary to specify how data from subjects with intercurrent events will be handled to provide specificity about the estimand.

E9(R1) discusses several approaches to handling data after an intercurrent event. One suggested approach is to incorporate the intercurrent event as a failure mode in a composite endpoint requiring clinically meaningful improvements. However, if the primary effectiveness endpoints are continuous changes from baseline rather than a responder analysis, another approach is necessary.

Another approach to handling intercurrent events is to adopt a "treatment policy" estimand. This approach requires data collection after the intercurrent event. If data is not collected after the intercurrent event, it is not possible to adopt a "treatment policy" estimand. Moreover, the treatment policy estimate does not directly pertain to the effectiveness of the investigational device, but rather pertains to the comparison of starting off with the investigational device and starting off with the control device but allowing changes to the treatment regimen as deemed medically necessary. The treatment policy approach is not consistent with the estimand defined above. Adopting the treatment policy approach changes the primary question the study is designed to address.

The hypothetical approach is "envisaged in which the intercurrent event would not occur: the value to reflect that scientific question of interest is that which the variable would have taken in the hypothetical scenario defined." That is, the "treatment effect of interest might concern the outcome if the subsequent active treatment had not been administered."

The hypothetical approach is consistent with the estimand defined above and so will be used for this study. It is reasonably assumed that subjects who discontinue due to lack of effectiveness do not experience clinically significant improvement in pain. Therefore, to include these subjects in the primary analysis, a baseline carried forward approach will be used implying that subjects who discontinue due to lack of effectiveness are given a value of zero for their changes from baseline. A sensitivity analysis will be performed in which these subjects are given the last observation prior to the discontinuation as their Primary endpoint (week 8) value.

2.9.2 HANDLING OF SUBJECTS LOST TO FOLLOW-UP WITHOUT AT LEAST ONE EVALUATION

Subjects lost to follow-up may produce bias in this study. Reasons for discontinuation will be evaluated and characterized. Subjects who do not contribute any follow-up data will not be imputed using multiple imputation.

2.9.3 HANDLING OF SUBJECTS LOST TO FOLLOW-UP WITH AT LEAST ONE EVALUATION

- If there are any patients with no follow-up (defined as not having at least one completed follow-up survey), multiple imputation (MI) (Rubin and Schenker 1991) will be used to create 20 multiply imputed data sets. This will allow patients with no follow-up to be included by imputing their follow-up outcomes in a probabilistic fashion starting with predictions based on baseline values only. The primary device group comparison using the same MMRM will be performed in at least 20 multiply imputed completed data sets with results combined using the Rubin formula, which can be implemented in SAS PROC MIANALYZE. The MI model will contain patient baseline demographic variables (age, sex, BMI, and household income) and baseline values of primary and secondary patient reported outcomes as well as values of the primary endpoint evaluated over time. It is likely that the missing data pattern will monotonic or nearly so. If so, then monotonic regression methods will be used to obtain 'proper' imputations using SAS PROC MI with sporadic, low frequency missingness determined using single imputation prior to MI. If the missing value pattern is more than trivially non-monotonic, a fully conditional specification (FCS) MI approach that allows for non-monotonic missing be employed, i.e., multivariate imputation by chained equations (MICE), which can also be implement in SAS PROC MI.

2.10 COMPARISONS AT BASELINE AND COVARIATE ANALYSIS

Baseline factors will be compared among experimental groups, including pre-dose Pain, ODI, PROMIS sleep disturbance, anxiety and depression, age, sex, race/ethnicity, education, household income, and BMI. Continuous variables will be described using means, standard deviations, median, minimum, and maximum values, and compared among groups using analysis of variance (ANOVA). Categorical variables will be described using counts and percentages and compared among groups using chi-squared or Fisher's exact tests. Where appropriate, between group differences will be described along with 95% confidence intervals. Any baseline variables that differ among experimental groups and are associated with the outcome of interest will be included in the MMRM model to remove potential bias and improve precision.

2.11 SUBGROUP ANALYSIS

Differences in the efficacy of the investigational product will be evaluated within pre-specified subgroups of interest, defined based on key clinical factors. The primary subgroup analysis in this project will be evaluating differences based on sex (male vs. female), number of comorbidities (<=1, >1), race/ethnicity (African American, Caucasian, Hispanic, Asian), SES (above or below median), and insurance type. Analyses will include repeating the examination of the primary efficacy endpoint differences among groups within male and female, separately, as well as conducting formal statistical interaction tests evaluating evidence for effect modification in the context of the MMRM that includes all subjects. Evaluation of whether the between group differences over time are different based on strata will be based on testing the significance of the [sex]-by-[experimental group]-by-[time] interaction term in a model that includes all lower-ordered terms.

Exploratory subgroup analyses may be performed based on additional clinical characteristics (e.g., stratification based on baseline Pain [<40 vs. ≥40%]) deemed relevant by content experts to describe more fully an observed associations in primary and secondary analyses.

Heterogeneity of treatment effects will be evaluated for the named subgroups and the data presentation in the CSR will include the estimated change scores along with its 95% confidence interval (CI) within each subgroup. These results will be graphically displayed in a forest plot.

2.12 MULTIPLICITY

2.12.1 PRIMARY ENDPOINTS

A Hochberg "step-up" approach will be utilized to control overall type I error at the desired 5% level across the two primary efficacy analyses. To implement the Hochberg method, the p-values for the set of multiple null hypotheses are ordered from largest to smallest, and each p-value is compared to a sequentially decreasing alpha-level to determine whether the null hypothesis (and, potentially, subsequent hypotheses) should be rejected. Symbolically, for the set of p-values $\{p_1, \ldots p_k\}$ ordered from largest to smallest and testing the corresponding set of null hypotheses $\{H_{o1}, \ldots H_{ok}\}$, the Hochberg procedure is implemented as:

> **Step 1:** Evaluate whether $p_1 < \alpha$. If yes, reject H_{o1} and all subsequent null hypotheses $\{H_{o2}, \ldots H_{ok}\}$. Else, do not reject H_{o1} and go to Step 2.
>
> **Step 2:** Evaluate whether $p_2 < \alpha/2$. If yes, reject H_{o2} and all subsequent null hypotheses $\{H_{o3}, \ldots H_{ok}\}$. Else, do not reject H_{o2} and go to Step 3.
>
> […]
>
> **Step k:** Evaluate whether $p_k < \alpha/k$. If yes, reject H_{ok}. Else, none of the null hypotheses $\{H_{o1}, \ldots H_{ok}\}$ are rejected and stop.

As only the smallest p-value is compared to the traditional Bonferroni-corrected α-level, the Hochberg method is more statistically powerful for controlling type I error in the context of testing multiple null hypotheses. Hochberg also shows that this method is generally more statistically powerful than the Holms step-down procedure. This method of multiplicity was selected given the expectation that the endpoints are not negatively correlated. Should the observed data be negatively correlated, a bootstrapping approach for multiplicity will be provided in addition to the Hochberg correct. Therefore, the Hochberg step-up procedure achieves the goal of maintaining overall type I error at the desired α-level across a set of null hypotheses, while also improving statistical power over other approaches. The adjusted p-values from this procedure will be considered the primary study results.

2.12.2 SECONDARY ENDPOINTS

The following secondary endpoints will be adjusted for multiplicity using the same methodology as the primary endpoint:

- ODI
- PROMIS sleep disturbance
- PROMIS anxiety
- PROMIS depression

3 OTHER ELEMENTS OF THE ANALYSIS PLAN

3.1 SUBJECT DISPOSITION

Subject disposition will be reported in accordance with the 2004 FDA Guidance for Clinical Data Presentations for Orthopedic Device Applications (see References).

A subject Accounting Table will include at least the following elements:

Theoretical follow-up: The theoretical follow-up is the number of subjects that would have been examined if all subjects were evaluated on the exact anniversary of their respective initial treatment dates.

Cumulative deaths: Cumulative deaths up to the date of the exact anniversary defining the current interval. Deaths occurring after the exact anniversary are recorded in the next interval.

Not Yet Overdue: Includes subjects whose anniversary has occurred; however, clinical data has not yet been collected (e.g., BPI is unavailable) but the subject is still in the protocol specified follow-up window. Such subjects may yet be observed and so follow-up compliance estimates account for this by removing such subjects from the denominator as well as from the numerator when determining compliance ratios.

Expected due for clinic evaluation: This row is the number of subjects expected for a given time interval. These include the theoretical number of subjects who were due to be evaluated, less the number of subjects who died by that time interval and less the subjects in the "Not yet overdue" category.

Visit Compliance (%): These rows indicate the percentage of subjects compliant with the specified visit scheduled for all evaluated subjects among expected due subjects, and separately, for all subjects that are within the window among expected due subjects.

3.2 SAFETY ANALYSIS

Assessment of the safety of the investigational device will include an evaluation of the incidence and severity of complications and adverse reactions associated with the treatment. Primary safety comparisons will be performed using the mITT analysis set.

Adverse event rates will be summarized by type of AE and for specific AEs in two ways:

1. Per subject incidence of specific AEs and classes of AEs
2. By event, summarizing event counts by visit interval over time and in accordance with FDA Guidance (CDRH 2004).

Events will be summarized by severity and relatedness.

The summary tables will show the adverse events (in coded terms), the total number of events, and the number and the percentage of subjects affected in the investigational group

("subject wise evaluation") and stratified by relation to device and severity. Data of dropouts will not be presented separately, but possible bias will be discussed in the final report.

3.3 PROTOCOL DEVIATIONS

Protocol deviations will be summarized by experimental group.

3.4 SOFTWARE FOR DATA ANALYSIS

Analysis will be performed with SAS® software Package (Release 9.4 or higher) or R (version 4.1 or higher).

3.5 CHANGES TO THE ANALYSIS PLAN

All Changes to the Analysis Plan will be documented in the Clinical Study Report. Deviations to the analysis plan will be described in full in at least a supplemental section for publications.

4 REFERENCES

Guidance for Industry and FDA Staff. Clinical Data Presentations for Orthopedic Device Applications. Document issued on: December 2, 2004. https://www.fda.gov/media/71670/download

Rubin DB and Schenker N. Multiple imputation in health-care databases: an overview and some applications. Statistics in Medicine 10: 585-598, 1991.

Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain. 2008;9(2):105-21. pub 2007/12/07.

Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. Pain. 2000;88(3):287-94. Epub 2000/11/09.

Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best Pract Res Clin Rheumatol. 2005;19(4):593-607.

Anderson SP, Zylla DM, McGriff DM, Arneson TJ. Impact of Medical Cannabis on Patient-Reported Symptoms for Patients With Cancer Enrolled in Minnesota's Medical Cannabis Program. J Oncol Pract. 2019;15(4):e338-e45.

Deer TR, Gilmore CA, Desai MJ, Li SC, DePalma MJ, Hopkins TJ, Burgher AH, Spinner DA, Cohen SP, McGee MJ, Boggs JW. Percutaneous Peripheral Nerve Stimulation of the Medial Branch Nerves for the Treatment of Chronic Axial Back Pain in Patients After Radiofrequency Ablation. Pain Med. 2021 Mar 18;22(3):548-560.

Bomalaski MN BJ, Hoffman J, et al. . Predictors of response to telehealth CBT for chronic Pain in veterans with TBI. Arch Phys Med Rehabil 2018;99(10):E103-E104.

Mehta, N., Bucior, I., Bujanover, S. et al. Relationship between pain relief, reduction in pain-associated sleep interference, and overall impression of improvement in patients with postherpetic neuralgia treated with extended-release gabapentin. Health Qual Life Outcomes 14, 54 (2016).

Mease PJ, Spaeth M, Clauw DJ, Arnold LM, Bradley LA, Russell IJ, Kajdasz DK, Walker DJ, Chappell AS. Estimation of minimum clinically important difference for pain in fibromyalgia. Arthritis Care Res (Hoboken). 2011 Jun;63(6):821-6.

Krebs EE, Bair MJ, Damush TM, Tu W, Wu J, Kroenke K. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. Med Care. 2010 Nov;48(11):1007-14.

Krebs EE, Gravely A, Nugent S, Jensen AC, DeRonne B, Goldsmith ES, Kroenke K, Bair MJ, Noorbaloochi S. Effect of Opioid vs Nonopioid Medications on Pain-Related Function in Patients With Chronic Back Pain or Hip or Knee Osteoarthritis Pain: The SPACE Randomized Clinical Trial. JAMA. 2018 Mar 6;319(9):872-882.