Title: ReActiv8 Stimulation Therapy vs Optimal Medical Management: A Randomized Evaluation

NCT04803214

11-JAN-2024

Statistical Analysis Plan

For the RESTORE Post Market Study (Protocol 980012)

Sponsor

MML US Inc.
2159 India Street, Suite 200
San Diego, CA 92101

Rev D

11 JAN 2024

Revision History

Rev	Change Description	Effective Date
A	Initial Release	02 Jul 2021
B	Add leg pain description and NRS Add EQ-5D sub-analysis	11 Aug 2021
C	• Increase number of sites to up to 30 sites • Clarify that visit windows are a guideline, not a strict requirement • Remove the additional 30 days after randomization for the start of visit timing in the Control group • Clarifications on hypotheses and additional details on analysis methods.	07 Sep 2022
D	• Corrections and clarifications throughout to align with the RESTORE protocol (980012). • Change to Sponsor address	11 JAN 2024

Contents

1. Overview

This Statistical Analysis Plan (SAP) describes the statistical methods to be used during the reporting and analysis of data collected under the RESTORE Protocol (980012). This SAP has been reviewed by Biomedical Statistical Consulting and NAMSA.

2. Scope

This SAP should be read in conjunction with the protocol and case report forms (CRFs). This version of the SAP has been developed according to protocol 980012, Rev D. Any further changes to the protocol or CRFs may necessitate updates to the SAP.

3. Trial Design

The ReActiv8 Stimulation Therapy vs Optimal Medical Management: A Randomized Evaluation (RESTORE) study is a prospective, randomized study comparing ReActiv8 therapy to Optimal Medical Management (OMM). Patients meeting all eligibility criteria will be randomized to either the Treatment group (ReActiv8) or the Control group (OMM). Randomization will be performed according to a random permuted block design stratified by clinical site with a 1:1 allocation ratio.

The primary endpoint is a comparison of the change in Oswestry Disability Index (ODI) at 1 year between the Treatment and Control groups. After the 1 Year visit, all patients in the Control group may elect to receive a ReActiv8 device and will continue to be followed through the 2 Year visit. Patients who do not receive a device after the 1 Year visit will be exited from the study at that time.

A minimum of 204 evaluable patients is required to sufficiently power the primary endpoint. To allow for attrition, approximately 230 patients will be randomized at up to 30 clinical sites. To account for screen failures prior to randomization approximately 400 patients may be enrolled. A patient is enrolled in the study at the time of informed consent. Patients who are enrolled, meet the required inclusion and exclusion criteria, and continue consent will be randomized. Total study duration is anticipated to be about 4 years from first enrollment to the last 2 Year visit.

The study visit schedule will vary slightly between the groups to allow for those in the Treatment group to be implanted after the randomization visit. After the 1 Year visit, patients in the Control group may elect to be implanted. The study schema is provided in Figure 1.

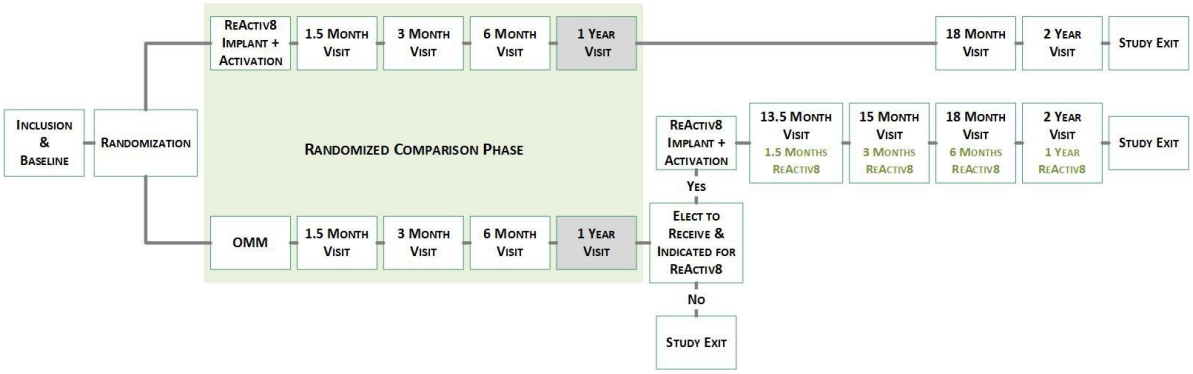

Figure 1: Study Schema

3.1. Randomization

Patients will be randomized at the Randomization Visit, approximately 14 days after baseline. Randomization will be performed according to a random permuted block design stratified by clinical site with a 1:1 allocation ratio for Treatment vs. Control.

3.2. Sample Size Rationale

The sample size for the study is determined under the following assumptions for the primary efficacy endpoint, which are based on a prior study of ReActiv8,[1] and approximately the MMRM test by its corresponding unadjusted t-test. All else being equal, this is a conservative estimate of power since control for baseline and inclusion of intermediate time points is not accounted for.

- Minimum power of 80%
- Type I error of 5%
- Assumed mean change in Treatment group: 18.2
- Assumed mean change in Control group: 12.2
- Pooled Standard Deviation: 15

Under the above assumptions, a minimum of 204 evaluable patients is required in order to demonstrate superiority of the ReActiv8 treatment. To account for attrition, approximately 230 patients will be enrolled and randomized.

4. Data Structure and Handling

4.1. Data Handling and Transfer

Programming of analysis datasets, tables, figures and listings will be conducted during the data management phase of the study. Tables, figures, and listings may be reviewed prior to final data lock for data review. Any data values requiring investigation or correction will be identified, and protocol deviations will be reviewed. The final run of outputs will take place after the data are deemed final.

4.2. Missing Data and Sensitivity Analyses

The primary and secondary efficacy endpoints will be assessed in the population of all randomized patients with at least one follow-up visit at or before the 1 Year visit, For the primary and secondary efficacy endpoints, mixed-effect repeated measure (MMRM) analyses will be conducted utilizing all available data through 1 Year. Additional sensitivity analyses will be conducted to assess the impact of missing data.

As a supporting analysis, the primary and secondary efficacy endpoints will be summarized in the completer cohorts consisting of all patients with both baseline and 1 Year data available. This will be summarized as the arithmetic mean change from baseline rather than the model-based estimation used for the primary analysis.

4.3. Visit Windows

For analysis purposes, data will be analyzed according to the visit assigned within the database.

4.4. Pooling of Data Across Trial Sites

Poolability of data will be assessed using the specified MMRM models for the primary analysis with fixed covariates for study site and the interaction term between treatment group and study site. The data will be considered poolable if the p-value for the interaction test is ≥ 0.15. If the p-value is < 0.15 additional analyses will be conducted to determine whether the site differences are due to imbalances in baseline factors that have evidence ($p < 0.20$) of an association with the primary outcome. If poolability across sites is not demonstrated, then results for the primary endpoint may be stratified and presented by site groupings.

5. Statistical Analyses

5.1. General Considerations

All statistical analyses will be conducted using SAS version 9.3 or later (SAS Institute Inc., Cary, NC) and other validated statistical software as required. Continuous variables will be summarized with means and standard deviations or as medians and interquartile ranges. Categorical variables will be summarized with the number and proportion of patients in each category. Binary outcomes will be presented as proportions with corresponding 95% confidence limits.

The primary efficacy endpoint and associated secondary efficacy endpoints are assessed at the 1 Year visit. Additional endpoints will be assessed at the 2 Year visit.

5.2. Subject Disposition

Subject disposition for randomized patients will be presented by:

- Summary of patients per visit
- Summary of early withdrawal and reason for early withdrawal
- Summary of randomized subjects not included in the primary analysis population

5.3. Demographics and Baseline Characteristics

Demographics and baseline characteristics of the data set will be summarized by randomization group. These parameters will include (but not be limited to):

- Age
- Gender
- Ethnicity
- BMI
- Medical history
- Low Back pain descriptive characteristics
- Duration of back pain
- Work status
- Treatments being used to treat low back pain

5.4. Primary Efficacy Endpoint

5.4.1. Primary Analysis

Oswestry Disability Index (ODI) is a disease specific assessment of the disabling effects of back pain.[48] The ODI covers 1 item on pain and 9 items on activities of daily living (personal care, lifting, walking, sitting, standing, sleeping, sex life, social life, and traveling). ODI is reported as a score from 0 to 100.

The relevance of changes in ODI has been explored by many authors, and discussions abound on Minimal Clinical Important Difference (MCID), Minimally Important Change (MIC) and the like. Fairbank, et al., suggest a minimally important change is 15 points.[2]

Hypothesis Test

The primary superiority hypothesis is that the mean change in ODI from baseline to 1 Year is more negative (i.e., superior to) for patients treated with ReActiv8 Stimulation Therapy compared to patients treated with OMM. Symbolically, the primary effectiveness hypotheses can be represented as follows:

Ho: $\delta_T - \delta_C \geq 0$

Ha: $\delta_T - \delta_C < 0$

Where δ_T and δ_C are the true mean changes in ODI from baseline to 1 Year for ReActiv8 Stimulation Therapy (T) and for patients treated with OMM alone (C), respectively. See section 5.7 for a discussion on Mixed Model Repeated Measures. See discussion in Section 5.8 defining the definition of the primary estimand $\delta_T - \delta_C$.

The two-sided p-value for the equality null hypothesis will be reported as well as the two-sided adjusted 95% confidence interval for the difference in mean changes. Superiority will only be claimed if $p<0.05$ and the mean change is more negative (greater improvement) for the Treatment group compared to the Control group. This is equivalent to specifying a one-sided test at $\alpha=0.025$.

5.5. Secondary Efficacy Endpoints

All tests of significance will be tested utilizing an overall Type I error of 5%. To control type 1 error, the p-values for superiority at 1 Year will only be interpreted as inferential if the superiority for Month 12 ODI is demonstrated at 2-sided type 1 error rate of $\alpha = 0.05$ (with hypothesized directionality). The Hochberg method (Hochberg 1988[3]) will be used to control type 1 error among these two secondary endpoints. To implement the Hochberg method, the larger of the two p-values, each derived from the corresponding MMRM, will be compared to 2-sided 0.05. If the larger of the two p-values is smaller than 0.05, superiority for both secondary endpoints. If the larger of the two p-values is not smaller than 0.05, then the smaller of two p-values will be compared to two-sided 0.05/2 = 0.025; and if less than 0.025, superiority will be claimed for the one secondary endpoint. If the smaller is not less than 0.025, the superiority will not be claimed for either secondary endpoint.

5.5.1. Low Back Pain – NRS

Average Low Back Pain will be measured using the 11-point Numerical Rating Scale for Low Back Pain, as recommended by IMMPACT.[40] Specifically, the NRS scale for "average low back pain in the last 24 hours" will be used.

The endpoint will compare change from baseline in NRS between Treatment and Control groups at the 1 Year visit.

Hypothesis Test

The hypothesis is that the mean change in NRS from baseline to 1 Year is more negative (i.e., superior to) for patients treated with ReActiv8 Stimulation Therapy compared to patients treated with OMM. Symbolically, the hypotheses can be represented as follows:

Ho: $\delta_T - \delta_C \geq 0$

Ha: $\delta_T - \delta_C < 0$

Where δ_T and δ_C are the true mean changes in NRS from baseline to 1 Year for ReActiv8 Stimulation Therapy (T) and for patients treated with OMM (C), respectively. See section 5.7 for a discussion on Mixed Model Repeated Measures. See discussion in Section 5.8 defining the definition of the primary estimand, $\delta_T - \delta_C$.

The two-sided p-value for the equality null hypothesis will be reported as well as the two-sided adjusted 95% confidence interval for the difference in mean changes. Superiority will only be claimed if $p<0.05$ and the mean change is more negative (greater improvement) for the Treatment group compared to the Control group. This is equivalent to specifying a one-sided test at $\alpha=0.025$.

5.5.2. EQ-5D

The EQ-5D-5L (referred to as EQ-5D throughout this document) is a quality-of-life questionnaire comprising of the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression.

Population norms for EQ-5D for various populations have been reported, and categorized by age, gender and social class. EQ-5D is reported as an index up to 1.00. An EQ-5D index score of <0 is possible in some circumstances.

The endpoint will compare change from baseline in EQ-5D between Treatment and Control groups at the 1 Year visit.

Hypothesis Test

The hypothesis is that the mean change in EQ-5D from baseline to 1 Year is more positive (i.e., superior to) for patients treated with ReActiv8 Stimulation Therapy compared to patients treated with OMM. Symbolically, the primary effectiveness hypotheses can be represented as follows:

Ho: $\delta_T - \delta_C \leq 0$

Ha: $\delta_T - \delta_C > 0$

Where δ_T and δ_C are the true mean changes in EQ-5D from baseline to 1 Year for ReActiv8 Stimulation Therapy (T) and for patients treated with OMM alone (C), respectively. See section 5.7 for a discussion on Mixed Model Repeated Measures. See discussion in Section 5.8 defining the definition of the primary estimand, $\delta_T - \delta_C$.

The two-sided p-value for the equality null hypothesis will be reported as well as the two-sided adjusted 95% confidence interval for the difference in mean changes. Superiority will only be claimed if $p<0.05$ and the mean change is more positive (greater improvement) for the Treatment group compared to the Control group. This is equivalent to specifying a one-sided test at $\alpha=0.025$.

5.6. Tertiary Endpoints

5.6.1. Percent Pain Relief (PPR)

Percent Pain Relief (PPR) is a question asked to the patient in which the patient is asked to report the percent pain relief at the time of the current visit compared to the pain at baseline.

The endpoint will compare PPR at the 1 Year visit between Treatment and Control groups.

5.6.2. Subject Global Impression of Change (SGIC)

Subject Global Impression of Change (SGIC) is based on the "Subject Global Impression of Change" as described by Farrar.[65] The patient is presented with a questionnaire with 7 choices.

The endpoint will compare SGIC as a binary outcome of "Very Much Improved" or "Much Improved" vs all other responses at the 1 Year visit between Treatment and Control groups.

5.6.3. Treatment Satisfaction

Treatment satisfaction will be measured by a single question asking if the patient is satisfied with the treatment.

The endpoint will compare Treatment Satisfaction at the 1 Year visit between Treatment and Control groups.

5.6.4. ODI and NRS Composite

Patients suffering from CLBP are continuously balancing their activity level with their level of pain. As their condition improves, patients make personal choices on whether to increase their level of activity while tolerating a certain level of pain, or to continue with the same level of activity as earlier but with less discomfort, or somewhere in between. These choices are based on the patients' individual circumstances and preferences. Therefore, when evaluating a therapy for CLBP, an evaluation of improvements in pain interpreted in conjunction with functional improvements allows for a more comprehensive assessment of the therapy.

This endpoint will compare the percent of patients with a reduction in ODI ≥ 15 or a reduction in NRS ≥ 50% (and no increase in either measure) at the 1 Year visit between Treatment and Control groups.

5.6.5. Leg Pain NRS

Average Leg Pain will be measured using an 11-point Numerical Rating Scale for Leg Pain. The patient will rate his/her average leg pain in the last 24 hours on a scale from zero to ten, where zero is no pain and ten is the worst imaginable pain

The endpoint will compare change from baseline in Leg Pain NRS between Treatment and Control groups at the 1 Year visit

5.7. Mixed Model for Repeated Measures

These hypotheses will be tested using the 1 Year contrast estimated from a mixed model for repeated measures (MMRM) (Verbeke and Molenberghs 2000)[4]. The MMRM will include baseline endpoint value (ODI, NRS or EQ-5D), treatment group, visit, and visit by treatment group interaction. Changes from baseline to Months 1.5, 3, and 6, as well as from baseline to 1 Year are included in the MMRM as a function of the fixed effects. A random site effect will be included to account for the randomization within site. A random site effect is appropriate as we wish to generalize to patients from any site and specifically from sites included in this study. Patients lost-to-follow-up due to reasons related to lack of effectiveness or subsequent to a related AE will utilize baseline observation carried forward imputation of missing values; otherwise, missing values are implicitly imputed while estimating the parameters of the MMRM. MMRM produces unbiased estimates when there is missing follow-up when missing-at-random (MAR) assumption is true (Rubin 1976[5]). The MAR assumption is that the likelihood of missing is statistically independent of the distribution of the unobserved missing outcome conditional on observed data and is most reasonable for missingness not associated with lack of efficacy.

5.8. Defining the Primary Estimand

Further refinement of primary estimand, $\delta_T - \delta_C$, and implications for handling missing data and clinical status measures after relevant 'intercurrent' events are now discussed in the context of FDA Guidance E9(R1)[6]. The primary estimand is defined as the ***treatment group difference in (adjusted) mean changes from baseline to 1 Year among patients who start their randomized treatment and continue on that treatment throughout the 1 Year follow-up period.***

Reasons for missing after initial treatment initiation for patients with at least one follow-up visit (at Months 1.5, 3 6, and/or 1 Year) will be categorized as:

- Withdrawal due to lack of effectiveness
- Crossover from control to ReActiv8 due to lack of effectiveness
- Withdrawal due to removal of ReActiv8 for reasons other than lack of effectiveness (e.g., to obtain an MRI)
- Withdrawal due to adverse event
- Missing due to loss-to-follow-up with no known reason

Randomized patients who do not initiate a study treatment or do not have any follow-up data through 1 Year will be categorized as:

- Randomized to ReActiv8 but withdrawn prior to implant
- Randomized to ReActiv8 but withdrawn before any follow-up visit
- Randomized to OMM but withdrawn before any follow-up visit

E9(R1) sets up the argument for specifying clinically meaningful estimands as follows:

> *It remains undisputed that randomization is a cornerstone of controlled clinical trials and that analysis should aim at exploiting the advantages of randomization to the greatest extent possible. However, the question remains whether estimating an effect in accordance with the ITT principle always represents the treatment effect of greatest relevance to regulatory and clinical decision making. The framework outlined in this addendum gives a basis for describing different treatment effects and some points to consider for the design and analysis of trials to give estimates of these treatment effects that are reliable for decision-making.*

Relevant comments from E9(R1) include:

- Discontinuation of randomized treatment represents an intercurrent event to be addressed in the precise specification of the trial objective through the estimand. Study withdrawal gives rise to missing data to be addressed in the statistical analysis.
- Switching treatment represents an intercurrent event, and the clinical question of interest in respect to that event should be clear.
- In addition, regarding the distinct consequences of different intercurrent events, events such as discontinuation of treatment, switching between treatments, or use of an additional medication may render the later measurements of the variable irrelevant or difficult to interpret even when they can be collected. Measurements after a subject dies do not exist.
- Hypothetical Strategies: A scenario is envisaged in which the intercurrent event would not occur: the value of the variable to reflect the clinical question of interest is the value that the variable would have taken in the hypothetical scenario defined.

With these E9(R1) considerations in mind, defining of the primary analysis set and handling of missing data and data observed subsequent to a relevant intercurrent events are specified as follows:

- Only patients randomized to ReActiv8 who are implanted with the device and who have at least one follow-up assessment (at Months 1.5, 3, 6, and/or 1 Year) will be included in a Full Analysis Set (FAS) to be used in testing of primary and secondary effectiveness endpoints.
- Only control patients who have at least one follow-up assessment (at Months 1.5, 3, 6, and/or 1 Year) will be included in the FAS.
- Patients who are withdrawn from the study due to lack of effectiveness prior to the primary efficacy assessment at 1 Year will have their subsequent clinical status indicators defined by baseline values (i.e., baseline observation carried forward or BOCF) for the primary and secondary endpoint analyses. A sensitivity analysis will be conducted in which their subsequent clinical status indicators are defined by last observation carried forward (LOCF). Similar methodology will be employed for analyses after the 1 Year time point.

- Control patients who are withdrawn prior to the 1 Year visit to receive ReActiv8 commercially will have their subsequent clinical status indicators defined by baseline values (i.e., BOCF). A sensitivity analysis will be conducted in which their subsequent clinical status indicators are defined by LOCF. The rationale for employing BOCF in primary analyses is the hypothesis that such patients experience negligible improvement in clinical status and therefore desire what they perceive as a potentially effective treatment, ReActiv8.
- Patients withdrawn from treatment due to an adverse event prior to the primary efficacy assessment at 1 Year will have their subsequent clinical status indicators defined by baseline values (i.e., baseline observation carried forward) for the primary and secondary endpoint analyses. A sensitivity analysis will be conducted in which their subsequent clinical status indicators are defined by last observation carried forward. The rationale for employing BOCF in primary analyses is the hypothesis that patients who experience AE severe enough to be withdrawn from their study treatment cannot benefit from the treatment and so the expected change is zero. Similar methodology will be employed for analyses after the 1 Year time point.
- Patients who have their ReActiv8 removed for reasons other than lack of effectiveness prior to the primary efficacy assessment at 1 Year will have their subsequent clinical status implicitly imputed through the use of the MMRM by assuming MAR for the primary and secondary endpoint analyses.
- Patients in either group who have missing primary or secondary endpoint data due to loss-to-follow-up with no known reason will have their subsequent clinical status implicitly imputed through the use of the MMRM by assuming MAR.

Residuals from the models will be evaluated to determine if the distributional assumptions concerning error and random effect distributions of the MMRM are tenable. The primary and secondary endpoints are bounded reducing the risk of extreme outliers. Normalizing transformations may be considered if required for validity of the MMRM. If the distributional assumptions of the MMRM are not tenable and no adequate normalizing transformation can be found, corresponding responder analyses will be performed. Responder definitions will be commonly used thresholds that are typically supported through comparison to MCID. Thus, for ODI, the responder definitions is an improvement in ODI of at least 15 points. For NRS, the responder definition is an improvement in NRS of at least 50%. Even if the parametric assumptions of the MMRM are met, descriptive comparisons between groups will utilize these responder definitions.

5.9. Safety Assessment

Reportable adverse events (AEs) are those related to the device, procedure, stimulation or other therapies utilized to treat LBP, and all serious adverse events (SAEs), whether related or not.

All reportable AEs will be documented and reported from the time of informed consent through the end of the study with summary statistics presented for observed rates. No formal statistical hypotheses will be tested in the safety assessment.

5.10. Supporting Analyses

The supporting efficacy analyses will include:

- Comparison between the Treatment and Control groups in the of responder rate in ODI at the 1 Year visit, where a change ≥15 points is considered a responder
- Comparison between the Treatment and Control groups in the cumulative proportion of responder rate (a comparison of ranks of the percentage of "responders" across the range of all possible response thresholds) in ODI at the 1 Year visit,
- Comparison between the Treatment and Control groups in the of responder rate in NRS at the 1 Year visit, where a change ≥50% is considered a responder
- Comparison between the Treatment and Control groups in the cumulative proportion of responder rate (a comparison of ranks of the percentage of "responders" across the range of all possible response thresholds) in NRS at the 1 Year visit
- Within group changes in all analyses will also be assessed at the 2 Year visit
- Health economic outcome measures at the 1- and 2 Year visits will be compared to baseline
- Activity monitoring will be collected on a subset of patients and assessed through the 2 Year visit and compared to baseline
- Additional ad hoc analyses may also be conducted.

5.10.1. Health Economics

Health economic outcome measures to be recorded will include:

- Work status, work-days missed and ability to do their work the year prior to baseline, and at the 1- and 2 Year visits
- Health care utilization (office visits, hospital visits, emergency room visits, and other therapies such as physical therapy) the year prior to baseline and at the 1- and 2 Year visits

5.10.2. Activity Data Collection

Activity data will be collected on a subset of patients who:

- have an activity tracker that they have used regularly for at least three months,
- agree to wear the activity tracker for the duration of the study, and
- agree to have activity data transmitted to be saved in the database.

After signing consent and meeting enrollment criteria, activity data from at least 2 weeks prior to enrollment will be downloaded. Activity data will be collected continuously while the patient is wearing the activity tracker and downloaded periodically by the patient. To help ensure complete data are obtained, a download of the data will also occur at each study visit.

The time period from Baseline through Activation will serve as the Baseline assessment of activity, with the following exceptions:

- Patients in the Treatment group would be expected to be moving less during the healing process post-procedure; therefore, activity data will be excluded between the Implant and Activation visits (approximately 2 weeks).
- Patients in the Control group will have the 2-week time period prior to Activation excluded to align with the timeframe excluded for the Treatment group.

Activity data to be collected include:

- Distance and Number of Steps
- Stationary Time
- Sleep Data

5.11. Subgroup Analyses and Predictors of Response

The primary efficacy endpoint will be analyzed in the subgroups defined below to demonstrate consistency of treatment effect. The specified MMRM model will be used to test the interaction of randomization assignment and subgroup/covariate. Subgroups to be analyzed will include, at a minimum:

- Gender
- Age (as both a continuous variable and grouped in ranges)
- Ethnicity
- Baseline NRS
- Baseline EQ-5D
- BMI (as both a continuous variable and grouped in ranges)
- Duration of work status adversely affected by CLBP

Multivariable analysis to explore predictors of response with response defined as an improvement of 15 points or more in ODI will be conducted in the Treatment group patients. Univariable regression models with (at a minimum) covariates as defined below will be analyzed to determine predictors of response. All covariates with an association with response ($p<0.20$) will be included in a multivariable logistic regression model.

- Duration of back pain (as both a continuous variable and grouped in ranges)
- Baseline ODI (as both a continuous variable and grouped in ranges)
- Baseline EQ-5D (as both a continuous variable and grouped in ranges)
- Baseline VAS (as both a continuous variable and grouped in ranges)
- Baseline opioid use
- Age
- Gender
- BMI
- Compliance with stimulation (the number of minutes that stimulation was actually delivered divided by the number of minutes that could have been delivered)

[1] Gilligan, C., Rauck, R., Rathmell, J, et al. An Implantable Restorative Neurostimulator for Refractory Mechanical Chronic Low Back Pain - A Randomized Sham Controlled Trial. *Pain* 2021 Oct 1;162(10):2486-2498.

[2] Fairbank, J. & Pynsent, P. B. The Oswestry Disability Index. *Spine (Phila. Pa. 1976)*. 25, 2940–52; discussion 2952 (2000).

[3] Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 1988;75:800-2. 10.1093/biomet/75.4.800

[4] Verbeke G and Molenberghs G. Linear Mixed Models for Longitudinal Data, New York: Springer 2000.

[5] Rubin DB. Inference and missing data. Biometrika 1976. 63:581-592.

[6] E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials, May 2021 https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical