BioStats
STATISTICAL CONSULTING LTD.

STATISTICAL ANALYSIS PLAN

NEUROLIEF LTD

MOOD STUDY (PROTOCOL CL-CIP-201)

29 MAY 2024
VERSION 01

AUTHORIZATION				
	NAME	**TITLE**	**SIGNATURE**	**DATE**
Issued by:	Netanel Deutsch, MA	Biostatistician		29-May-2024
Reviewed by:	Lisa Deutsch, Ph.D.	Biostatistician		29-May-2024
Reviewed by:	Yaron Gruper, M.Sc.	Clinical Studies Senior Manager		29-May-2024
Approved by:	Michal Kedar-Datel, B.sc.	VP Clinical, Regulatory & Quality		29-May-2024

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse Event
ATRF	Antidepressant Treatment Resistance Form
ATIF	Antidepressant Treatment Intolerance Form
C-SSRS	Columbia – Suicide Severity Rating Scale
CFR	Code of Federal Regulation
CGI-S	Clinical Global Impression scales - Severity
CGI-I	Clinical Global Impression scales – Improvement
CRF	Case Report Form
eCRF	Electronic Case Report Form
eCOT -AS	external Combined Occipital and Trigeminal Afferent Stimulation
EDC	Electronic Data Capture
FDA	Food and Drug Administration
HDRS	Hamilton Depression Rating Scale
IFU	Instruction For Use
ISO	International Organization of Standardization
ITT	Intent to Treat
LTFU	Lost To Follow Up
MDD	Major Depressive Disorder
MDRS/MADRS	Montgomery-Asberg Depression Rating Score
mITT	modified Intent to Treat
mYITT	Modified Young Intent to Treat
NTF	Note To File
PP	Per Protocol
QIDS	Quick Inventory of Depressive Symptomatology
Q-LES-Q	Quality of Life Enjoyment and Satisfaction Questionnaire
RDC	Remote Data Capture
SAP	Statistical Analysis Plan

1. INTRODUCTION

This Statistical Analysis Plan (SAP) is a more detailed companion to the Statistical Methods section of the study titled "The MOOD study – external Combined Occipital and Trigeminal Nerve Stimulation (eCOT-NS) for the treatment of Major Depressive Disorder (MDD)" and provides a comprehensive description of the analysis sets, endpoints, methods, and data analyses to be used. This SAP prevails when differences exist in descriptions or explanations provided in the protocol and the SAP. We shall address analysis strategies for the open label stage of the study. Addressed as well is a plan for prognostic factor analysis to identify potential baseline parameters related to the primary endpoint and its derivatives.

Relivion®DP, an external Combined Occipital and Trigeminal Afferent Stimulation (eCOT-AS) is proposed as a novel treatment for MDD. The combination of neuromodulation of both the occipital and the trigeminal nerve branches non-invasively is made possible for the first time, after having been successfully performed either separately or in invasive procedures.

Major depressive disorder is one of the leading causes of disability worldwide. However, treatment options are still limited, therefore new approaches are needed to enhance clinical improvement. The non-invasive combined neuromodulation of both the occipital and trigeminal nerve branches is a safe, self-administered novel technology with high potential for treating MDD patients. Based on this rationale, Neurolief developed the Relivion®DP, a neuro-stimulator applying combined occipital and trigeminal afferent stimulation for treatment of major depressive disorder (MDD). Following completion of an early clinical study showing promising results, Neurolief designed the proposed clinical trial to evaluate the safety and efficacy of the Relivion®DP in a two-arms controlled, double blinded randomized study.

The MOOD Study will evaluate the safety and efficacy of a self-administered treatment for MDD using an external combined occipital and trigeminal afferent stimulator (Relivion®DP).

This pivotal clinical investigation is intended to support future marketing applications (US FDA and CE) for Neurolief's Relivion®DP device.

2. STUDY DESIGN AND OBJECTIVE

2.1 Study Objectives

The main objective of the study is to evaluate the safety and efficacy of a self-administered treatment for MDD using an external combined occipital and trigeminal nerve stimulator (Relivion®DP). This pivotal clinical investigation is intended to support future marketing applications (US FDA and CE) for Neurolief's Relivion®DP device.

2.2 Study Design

2.2.1 Experimental Design

This is a prospective, randomized, double-blind, parallel-group, sham controlled, multi-center clinical investigation followed by an active open label extension period. All enrolled patients are those treated with the Relivion®DP device either by active treatment or sham treatment. The treatment will be done according to the IFU. Patients' duration in the study will be up to 20 weeks (including the screening period).

In this study for the double-blind phase, the Relivion®DP was provided in therapeutic and non-therapeutic modes, for the Active and Sham groups respectively:

- Active Treatment: For the therapeutic mode the device will be preset to the following parameters: stimulation waveform- symmetrical biphasic, phase duration 130-300 microseconds, pulse frequency 80Hz, trigeminal stimulation intensity – up to 6.7mA, Occipital stimulation intensity– up to 18mA.
- Sham Control: For the non-therapeutic mode the device will be preset to the following parameters: Stimulation waveform- symmetrical biphasic, phase duration 100 microseconds, pulse frequency 0.33 Hz, trigeminal stimulation intensity up to 5mA, occipital stimulation

intensity up to 10mA for 3 minutes, then gradually decrease to 0.2 mA and stays constant throughout the remaining treatment duration.

Both Active and Sham stimulation devices are packaged and labeled identically to maintain blinding of both the subject and study staff.

The double blind Relivion®DP treatment will last 60±20 minutes (preferably in two equal sessions of 30±10 minutes each) per day for 5-7 days a week (intensity level up to 50) for a period of 8±1 weeks.

For purposes of the active open label phase, only active treatments were applied. Open label phase treatment regimens would be as follows, according to the subject HDRS score at the end of the double-blind phase, for a period of 8±1 weeks:

Maintenance treatment (for HDRS remitter subjects): 40±10 minutes 3-4 times a week (intensity level up to 50).

Daily treatment (for HDRS responder subjects): 5-7 days a week 60±20 minutes per day (preferably in two equal sessions of 30±10 minutes each).

An interim analysis was performed after 86 (~80% of the calculated sample size) evaluable subjects were accrued, mainly for sample size re-assessment. The primary, as well as the safety endpoints were assessed in the ITT, mITT, PP and mYITT analysis sets. The interim analysis indicated "Favorable" results, therefore the Data Safety Monitoring committee (DMC) issued their formal recommendation that the study should continue enrollment with the original planned sample size of 124, without the need to increase the sample size.

2.2.2 Planned Sample Size and Study Hypothesis

The initial planned sample size was a total of 124 subjects, aged 22-70, randomized into two treatment arms, active Relivion®DP, and sham.

An additional up to 36 subjects (30+15% dropout) were planned to be randomized in the age range 18-21, to assess efficacy in this age range as well.

In this study, we will test the following hypotheses:

- H_0: $M_R = M_S$
- H_1: $M_R \neq M_S$

Where M_R and M_S represent the mean change from baseline in HDRS17 total score to week 8 in subjects suffering from MDD in the active treatment group and sham group, respectively

3. DATA MANAGMENT

All data will be collected using a paper CRF and/or Remote Electronic Data Collection (RDC/EDC) system. The clinical sites will use hard copy and/or electronic case report forms (eCRFs) to document the information required by the study protocol.

The EDC allows for the secure collection, transmission, validation, monitoring and real-time administration of study data gathered at investigative sites. The system allows password-restricted access to clinical trial information based on individuals' roles and responsibilities. The EDC are compliant with 21 CFR Part 11 and FDA's "Guidance: Computerized Systems Used in Clinical Trials".

Data reported on the CRF/eCRF should be driven from source documents (Section 14.5 in the study protocol) and be consistent with these source documents. Editing of data is done under full audit trail. Paper CRF worksheets may be provided to the investigational site to assist with data collection for the required fields.

Required data will be recorded on CRF/eCRFs by authorized site personnel as indicated on the Delegation of Authority Log. The Investigator will ensure that all eCRFs are completed promptly, completely, and accurately. Information on case report forms must conform to the information in the source documents.

The Sponsor will oversee and/or perform all data management functions. Data management functions include database development, system maintenance, data queries, and report generation.

Investigators must ensure that clinical records clearly indicate that the subject has been enrolled in a clinical investigation. Regulations require that Investigators maintain information in the study subject's medical records to corroborate data collected on the CRF/eCRF. In order to comply with these regulatory requirements, the study site will manage the following information, and

retain/manage it as required to make it available to monitors, auditors and/or regulatory bodies. Complete hospital and clinical medical records for all study subjects should include all study required procedures, labs and assessments as noted in Study schedule and Site Collection Data.

4. STATISTICAL SOFTWARE

All statistical analyses and data presentations, including tabulations and listings, will be performed using the SAS version 9.4 (SAS Institute, Cary NC, USA) software.

5. SAS® PROGRAMS VALIDATION

All SAS® programs used for analyses described in this document will be validated by double programming **or** code review before the final analysis as per BioStats SOP's.

6. RANDOMIZATION AND BLINDING

Up to the date of the finalization of this SAP the treatment code was not revealed. Subjects will be prospectively randomized into the clinical study. Randomization will occur only after the subject provides informed consent, completes all required screening and baseline procedures, and satisfies the study eligibility criteria. Eligible subjects will be randomly allocated (with a 1:1 ratio) to one of the following 2 treatment groups based on a randomization scheme using the permuted block method stratified by center:

- Active stimulation.
- Sham stimulation.

The Relivion®DP Active and Sham devices will look the same and will be provided in the same packaging bearing the same labeling. Hence, both subjects and study personnel will remain blinded to the assigned treatment group. Furthermore, the sponsor, investigator, MDD assessor and device trainer are all blinded to the treatment assignment.

7. MISSING DATA HANDLING & SENSITIVITY ANALYSES

The study outcomes will not be evaluated for patients who drop out prior to randomization. The primary endpoint will be analyzed using a repeated measures ANCOVA model which can handle data missing at random and we do not expect a high proportion of dropouts until the primary endpoint is measured. Therefore, for this evaluation no imputation of missing data is considered beyond the model estimates. Additionally, imputation of data such as by LOCF may harm linearity in such models, therefore, for this evaluation no imputation of missing data is considered beyond the model estimates.

Nevertheless, a sensitivity analysis based upon last observed values (LOV) may be performed. The Last Observed Value (LOV) is defined as the last available post baseline visit data up to and including the last treatment visit or termination visit. For categorical variables (such as response and remission rates at week 8) the LOCF concept will be applied. Baseline characteristics of patients who drop out will be evaluated by study group to evaluate the potential reason for differential drop out if needed.

The primary endpoint analysis, using the same methods described, will also be performed on the PP and ITT analysis sets as sensitivity analyses.

8. EFFICACY ENDPOINTS

8.1 Primary Efficacy Endpoint

Change from baseline to week-8 post Relivion®DP treatment initiation in depressive symptoms as measured by **HDRS-17 total score**.

8.2 Secondary Efficacy Endpoints

Secondary efficacy endpoints include:

- **Proportion of responder subjects** - defined as the percent of subjects achieving at least 50% reduction from baseline in their HDRS-17 total score, 8 weeks post Relivion®DP treatment initiation.

- **Proportion of subjects achieving remission** - defined as the percent of subjects with HDRS-17 score≤7 at 8 weeks post Relivion®DP treatment initiation.
- **Change from baseline** to week-8 post Relivion®DP treatment initiation **in depressive symptoms** as measured by **MADRS** total score.

8.3 Tertiary Efficacy Endpoints

Tertiary Efficacy endpoints include:

- **Change from baseline** to 8 weeks post Relivion®DP treatment initiation **in the Clinical Global Impression Severity scale (CGI-S) score**.
- **Clinical Global Impression Improvement (CGI-I) score** over time **to Visit 4**.
- **Change from baseline** to week-8 post Relivion®DP treatment initiation **in the total score of the Quick Inventory of Depressive Symptomatology (self-rated)** (QIDS-SR-16).
- **Change from baseline** to week-8 post Relivion®DP treatment initiation **in depressive symptoms** as measured by **HDRS-21** total score.
- **HDRS-17 Total score** to week-8 post Relivion®DP treatment initiation.
- **Proportion of subjects achieving clinically substantial improvement in HDRS-17 -** defined as the percent of subjects with reduction in HDRS-17 score of at least 7[1] points at 8 weeks post Relivion®DP treatment initiation.
- **HDRS-17 Category Shift from Visit 2 to Visit 4 –** defined as the category change from baseline to week-8 post Relivion®DP treatment initiation in HDRS-17; Categories definition: 5: Very severe 26-52, 4: Severe 20-25, 3: Moderate 14-19, 2: Mild 13-8 and 1: 0-7 Remission.
- **Change from baseline Q-LES-Q score at Visit** 4 – measured as the difference between the baseline and visit 4 Q-LES-Q percentage of maximum total scores.

[1] Rush AJ et al. Clinically Significant Changes in the 17- and 6-Item Hamilton Rating Scales for Depression: A STAR*D Report. Neuropsychiatric Disease and Treatment 2021:17 2333–2345

8.4 Safety Endpoints

Safety of the study device following study treatment: Incidence of adverse events and serious adverse events related or unrelated to the study device [Time Frame: up to 18 weeks post treatment initiation].

9. SIGNIFICANCE LEVELS AND HANDLING OF TYPE I ERROR

9.1 Type I error

The overall significance level for this study is 5% using two-tailed tests, except for treatment by site interaction that will be tested at a significance level of 10% and the test for early efficacy at the interim look.

9.2 Hierarchy Approach for The Primary and Secondary Endpoint Analysis

The hierarchy approach will be adopted for the primary and secondary endpoints to control type I error due to multiple endpoint testing. Thus, the primary endpoint will first be tested and only if $p \leq .05$, will the secondary endpoints be tested. For the three secondary performance endpoints, the Benjamini–Hochberg step-up method will be used to adjust the p-values. If the primary and the secondary endpoints are found statistically significant then the hierarchy will be extended to include the tertiary endpoints for the seven tertiary performance endpoints, the Benjamini–Hochberg step-up method will be used to adjust the p-values.

10. DATA ANALYSIS SETS

Each of the following analysis sets will be evaluated for the principal statistical analyses:

10.1 Intent to Treat Analysis Set (ITT)

The intent-to-treat (ITT) population will include all subjects enrolled in the study, aged 22 to 70, as randomized. This population will include all subjects who receive

at least one active/sham treatment. According to the ITT principle all subjects will be analyzed in the treatment group as assigned by randomization.

There was one randomization error; subject 06-BR-001 received an “A” group device from the site team by mistake instead of a “B” group device as per the randomization, a note to file was issued. This subject is included in the ITT set with treatment group as randomized in group “B”. In addition, two subjects were Screen failures who were randomized in error and two subjects under 22 years of age were removed from the ITT set.

Following is the list of subjects excluded from the ITT set with main reason:

Subject ID	Reason for Exclusion
08-JS-005	Screen failure, was randomized in error. No device training (NTF)- not in ITT as well
09-IS-012	Screen failure, was randomized in error. No device training (NTF)- not in ITT as well
11-JV-012	19 years old
04-XM-027	21 years old

10.2 Modified Intent to Treat Analysis Set (mITT)

The modified intent-to-treat (mITT) population will include all subjects from the ITT set enrolled in the study, who completed the minimal stimulation time required by the protocol and have no post randomization exclusion criteria. The mITT analysis set will be analyzed in the treatment group as treated.

There was one randomization error; subject 06-BR-001 received an “A” group device from the site team by mistake instead of a “B” group device as per the randomization, a note to file was issued. This subject is included in the mITT set with treatment group as treated in group “A”.

Following is the list of subjects excluded from the mITT set with main reason:

Subject ID	Reason for Exclusion
14-BO-005	Consent withdrawn
12-RB-004	Consent withdrawn
12-MP-005	Excluded due to compromised blinding, did not complete the minimal stimulation time required by the protocol
12-SM-015	Consent withdrawn
12-NM-016	Consent withdrawn

08-RP-003	Consent withdrawn
07-AV-001	Device failure at Double-Blind and open label phases (NTF)
07-GS-003	randomized in error (eligibility) but treated through the Double-Blind phase
06-SH-003	Consent withdrawn
05-DR-002	Consent withdrawn
04-RW-005	Consent withdrawn
02-MO-002	Device deficiency resulted with the subject got partial or no effective stimulation at all during the Double-Blind phase (NTF)
09-ST-009	Did not complete the minimal stimulation time required by the protocol
12-RN-017	Consent withdrawn
07-DA-007	Consent withdrawn
15-EH-002	LTFU that arrived for v4 after 16 days without performing treatments
14-DT-013	Consent withdrawn
15-RS-005	Consent withdrawn
08-DP-019	Consent withdrawn
12-JD-021	Consent withdrawn
10-TR-004	Consent withdrawn
12-BP-022	Consent withdrawn
12-MC-024	Did not complete the minimal stimulation time required by the protocol
12-AD-025	Did not complete the minimal stimulation time required by the protocol
08-ER-022	Consent withdrawn
05-EB-009	Lost to follow up (LTFU)
08-RB-024	Consent withdrawn

10.3 Per Protocol Analysis Set (PP)

The per protocol set includes all subjects from the mITT set that completed the 8-week treatment period (without withdrawal) and were without any major protocol deviations.

Following is a list of subjects not included in the PP set with main reason for exclusion:

Subject	Reason for Exclusion
06-NC-002	Change in dose (over 25%) of depression medication. Protocol deviations were documented for the subject.

10.4 Modified Young Intent to Treat Analysis Set (mYITT)

The modified younger intent-to-treat (mYITT) population will include all subjects from the ITT set enrolled in the study including subjects aged 18-21, who completed the minimal stimulation time required by the protocol and have no post randomization exclusion criteria. The mYITT analysis set will be analyzed in the treatment group as treated. Since only two subjects were under the age of 22, the mYITT set will not be analyzed separately and those subjects will not be included in the other analysis sets.

10.5 Statistical Analysis of Analysis Set

Safety assessments will be performed on the ITT analysis set. The mITT cohort will serve as the principal data analysis set for the primary, secondary and tertiary efficacy endpoints analyses. The primary and secondary efficacy assessments will also be performed on the per protocol (PP) and the ITT analysis sets to show consistency of study results.

11. STATISTICAL ANALYSIS

11.1 General Considerations

Statistical analyses will be performed using SAS® v9.4 (SAS Institute, Cary NC, USA). Statistical analyses and reporting will be performed in compliance with FDA Guidance 21 CFR part 812 and E9, ISO 14155, and MDR 2017/745. Study data will be summarized with descriptive statistics and presented in tables and figures. Continuous variables will be summarized by a mean, standard deviation, minimum, median and maximum and categorical variables by a count and percentage. For comparison of means (continuous variables), the two-sample t-test or the Wilcoxon rank sum test will be used as appropriate. For comparison of proportions (categorical variables), the chi-squared test or Fisher's exact test will be used as appropriate. If multiple measurements are taken in a single subject, statistics described below will be appropriately modified to accommodate the

within subject correlation Deviations from the planned analysis will be described, with proper justification, in the clinical study report.

11.2 Demographic and Case Characteristics

Demographic and baseline condition related characteristics will be tabulated and compared between the study groups by data type. Continuous variables will be summarized by a mean, standard deviation, minimum, median and maximum and categorical variables by a count and percentage. The statistical evaluation of baseline characteristics will include all available data from the ITT population.

These data will include:

- Demographic data
- Medical history
- Psychiatric history
- Previous and Current Psychotropic Medication and/or other treatments (i.e. psychotherapy, other medical devices)
- Type, dose and frequency of concomitant medications

11.3 Disposition and Subject Tolerability

The number of subjects who entered the study and completed each stage of the study will be provided and compared between the study groups, as well as the reasons for all post randomization discontinuations, grouped by major reason, e.g., lost to follow-up, adverse event, poor compliance, did not administer any treatment (with reasons). A list of discontinued patients, protocol deviations and subjects excluded from the efficacy analysis will be provided as well.

Time to withdrawal will also be assessed and presented by Kaplan-Meier curves and will be compared using the Log-Rank test if relevant.

11.4 Prognostic Factor Analysis

This section is intended to provide the architecture for this evaluation and does not jeopardize the sanctity of this study because this evaluation is being conducted independent of the randomized treatment assignment. Although

randomization creates asymptotic balance in important prognostic factors, including baseline values of the outcome variable in finite samples an imbalance in such factors may occur notwithstanding randomization, this represents the difference between the expectation of a random process and its realization. Depending crucially on the correlation between the baseline covariate and the outcome variable, this chance imbalance may not only create a potential bias in crude estimates of treatment effect in the outcome variable but may also affect the precision with which such an effect is measured and the statistical power of the analysis. Adjustment through the ANCOVA model is recommended to reduce risk of bias whilst also improving the precision of estimates and the power of the statistical test (Egbewale et al. BMC Medical Research Methodology 2014, 14:49). Specifically, if there was a significant difference in the HDRS-17 or in the secondary endpoints response and remission rates detected between the active and sham groups at 8 weeks post treatment initiation, but a retrospective analysis determined that the effect may have been influenced by one or more covariates not defined in the randomization scheme, the conclusion drawn from the unadjusted results could be challenged (Egbewale et al. BMC Medical Research Methodology 2014, 14:49).

The protocol and Statistical Analysis Plan (SAP) are clear regarding the tests for poolability (12.9.6 of the study protocol) and subset analyses (12.9.1 of the study protocol) that should be conducted. From Section 12.9.6 of the study protocol:

"Subgroup analysis of the primary efficacy endpoint by comparing US vs OUS centers, will be used to evaluate the poolability of the results. Treatment by center interaction will be tested in the primary analysis model at a significance level of 10%. If the interaction is not significant, it will be removed from the model. If the interaction is found significant, it will also be removed from the model and the reason for this interaction will be further explored and rationalized. This evaluation may include demographic features, symptoms at presentation, and clinical and treatment history, and center comparability in the features found to be associated with the primary efficacy variable.

In the case that poolability is questionable, the reasons for differential treatment effect, such as subject and clinical characteristics, will be investigated and reported."

From Section 12.9.1 of the study protocol:

"Subset Analyses of Primary Endpoint:

Significant or important variables, such as demographic or other baseline patient characteristics or concomitant medications (depending on use), will be entered as additional covariates in the primary efficacy endpoint ANCOVA model to evaluate their effect on the change from baseline HDRS-17 score.

- *Use of concomitant medications .*
- *US vs. OUS sites*
- *Blinding assessment - The effect of potential unblinding will be evaluated as a sensitivity analysis, where an additional binary variable identifying subjects unblinded versus those not un-blinded will be added to the ANCOVA model as a covariate. "*

The prognostic factor analysis is an extension of the examination of poolability and subsets, with an important difference not addressed in sections 11.5.2 and 11.5.3 of the SAP. Specifically, that the list of factors that may play a role in response have been prospectively defined and a threshold for inclusion has been defined. Before the final analysis is conducted on the locked data, the baseline factors will be examined using the locked database and significant covariates may be incorporated into the final model for analysis if deemed appropriate.

The following factors will be tested:

- Age.
- Sex at Birth
- Years of education.
- Employment status.
- ATRF score.
- ATIF score.
- Age of first episode.
- # of episodes since initial diagnosis.

- Other device use yes/no.
- Duration of other device used.
- Current episode duration.
- Psychotherapy yes/no.
- Duration of Psychotherapy.
- Each HDRS 17/21 item separately at baseline.
- Total HDRS 17/21 scores at baseline.
- Each separate MADRS item at baseline.
- Total MADRS scores at baseline.
- Baseline CGI-S scores.
- Baseline C-SSRS score Lifetime/ past 12 months.
- Baseline QIDS-16 score.

To assess correlation of the listed variables with the primary efficacy endpoint change from baseline HDRS-17 score to week 8 an ANOVA model will be used, entering each variable univariately as the independent variable. The same will be done for change from baseline HDRS-17 score to week 4.

To assess correlation of the listed variables for the binary variables (response and remission) a logistic regression model will be used, entering each variable univariately as the independent variable. The same will be done for change from baseline HDRS-17 score to week 4.

11.5 Endpoints Analyses

11.5.1 Primary Efficacy Endpoint Analyses

The primary efficacy endpoint is the change from baseline to 8 weeks post treatment initiation in HDRS-17 score.

The change in HDRS-17 from baseline to 8 weeks post treatment initiation will be compared between the treatment groups using a repeated measures analysis of covariance (ANCOVA, SAS® MIXED procedure). The model will include the following fixed effects: treatment group, visit, treatment group by visit interaction

with Baseline HDRS-17, and center entered as covariates. If covariates are identified in the prognostic factor analysis, they may be added to the model as well. Baseline HDRS-17 scores will be entered as a continuous variable so that the potential for co-linearity problems will be minimized. The treatment group by center interaction will be evaluated as well, but not as part of the principal statistical evaluation.

The unstructured covariance matrix structure will be used. If the model does not converge, then either the compound symmetry or autoregressive (whichever model has the lower AIC statistic) covariance matrix structure will be used instead. At this time point (up to 8 weeks) we do not expect a high proportion of dropouts, as a placebo effect of the sham treatment is expected and was taken into consideration in the design of the study. Thus, any missing data at 8 weeks post treatment initiation can be considered missing at random. Therefore, since repeated measures ANOVA is also an imputation method, for this evaluation no other method of imputation of missing data is considered beyond the model estimates. Nevertheless, should the missing at random assumption prove to be incorrect, a sensitivity analysis using other methods for data imputation mentioned in Section 7 may be performed.

Mock SAS code for the analysis:

```
Proc Mixed data=HDRS17;
  class visit group center subjectid;
  model change_HDRS17 = baseline_HRDS17 visit group visit*group;
  repeated / type=un subject=subjectid;
  lsmeans group*visit / pdiff cl;
  random center;
run;
```

The principal statistical analysis will be a comparison between the treatment groups, derived from the visit by treatment group interaction term from the model. The adjusted mean change from baseline in HDRS-17 scores at 8 weeks post treatment initiation will be estimated from the model (LS Means) interaction term for each group, as well as the difference between the adjusted means, and will be presented together with 95% confidence intervals. The null hypothesis will be rejected in favor of the alternative hypothesis and the study deemed

successful if the p-value is <0.05 and the mean change HDRS-17 in the active group is higher than that of the sham group.

The primary analysis model uses likelihood-based inference that is valid under a missing at random assumption.

Cohen's D will be calculated and presented as well as a measure of the clinical effect size of the change from baseline HDRS-17 at week 8. An effect size can also be calculated using methods described in "Effect sizes for growth-modeling analysis for controlled clinical trial in the same metric as for classical analysis", Psychol Methods. 2009 March; 14(1): 43-53), where ES = difference between LSmeans / Pooled SD of baseline HDRS-17 score.

11.5.2 Subset Analyses

The primary endpoint will be presented descriptively for the following subsets in addition they will be entered as additional covariates in the primary efficacy endpoint ANCOVA model to evaluate their effect on the change from baseline HDRS-17 score:

- Age Group (by median age)
- Sex (at birth)
- Baseline HDRS-17 severity ($\geq$ 26, 20-25, 17-19)
- Baseline Resistance, ATRF 1 vs. 2-4
- Concomitant anti-anxiety medication use, Yes/No
- Concomitant insomnia (includes anti-anxiety medications used to treat insomnia at bedtime) medication use, Yes/No
- Concomitant anticonvulsant medication use, Yes/No
- Comorbidities migraine and pain disorders, Yes/No
- Past history of ECT, TMS or VNS treatment, Yes/No
- Subjects on subtherapeutic dose of antidepressant, Yes/No
- Blinding assessment - The effect of potential unblinding will be evaluated as a sensitivity analysis, where an additional binary variable identifying subjects correctly guessed their treatment assignment versus those who did not will be added to the ANCOVA model as a covariate.

11.5.3 Pooling

Subgroup analysis of the primary efficacy endpoint by comparing clinical centers will be used to evaluate the poolability of the results. Treatment by center interaction will be tested in the primary analysis model at a significance level of 10%. This interaction term is not part of the primary endpoint model. If the interaction is found significant the reason for this interaction will be further explored and rationalized. This evaluation may include demographic features, symptoms at presentation, and clinical and treatment history, and center comparability in the features found to be associated with the primary efficacy variable.

In the case that poolability is questionable, the reasons for differential treatment effect, such as subject and clinical characteristics, will be investigated and reported.

11.5.4 Secondary Efficacy Endpoint Analyses:

- Proportion of responder subjects- defined as the percent of subjects achieving at least 50% reduction from baseline in their HDRS-17 total score, 8 weeks post Relivion®DP treatment initiation - will be summarized by a count and percentage and compared between the groups with a chi-squared test and a Fisher's exact test. The odds ratio with 95% Wald confidence interval will be presented as well. As an additional measure of effect size the NNT will be presented. If covariates are identified in the prognostic factor analysis, logistic regression will be performed as well adjusting for the identified factors.
- Proportion of subjects achieving remission- defined as the percent of subjects with HDRS-17 score≤7 at 8 weeks post Relivion®DP treatment initiation- will be summarized by a count and percentage and compared between the groups with a chi-squared test and a Fisher's exact test. The odds ratio with 95% Wald confidence interval will be presented as well. As an additional measure of effect size the NNT will be presented. If covariates are identified in the prognostic factor analysis logistic regression will be performed as well adjusting for the identified factors.

- Mean change in depressive symptoms, measured by MADRS total score, from baseline to week-8 post treatment initiation - will be compared between the groups using a similar model as the primary endpoint.

11.5.5 Tertiary Endpoints Analyses

- Mean Change in the CGI-S score - at 8 weeks post treatment initiation– will be compared between the groups using a similar model as the primary endpoint.
- Mean CGI-I scores at 8 weeks post treatment initiation will be compared between the groups using a similar model as the primary endpoint but with the value at each visit modeled instead of the change.
- Mean change from baseline in total score of the Quick Inventory of Depressive Symptomatology self-rated (QIDS-SR-16) score at 8 weeks post treatment initiation– will be compared between the groups using a similar model as the primary endpoint.
- Mean change in depressive symptoms, measured by HDRS-21 total score, from baseline to week 8 post Relivion®DP treatment initiation will be compared between the groups using a similar model as the primary endpoint.
- Proportion of subjects with a clinically substantial reduction in HDRS-17 score defined as the percent of subjects with reduction in HDRS-17 score of at least 7 points at 8 weeks post Relivion®DP treatment initiation - will be summarized by a count and percentage and compared between the groups with a chi-squared test and a Fisher's exact test. The odds ratio with 95% Wald confidence interval will be presented as well. As an additional measure of effect size the NNT will be presented.
- HDRS-17 total score at visit 4 (week-8 post Relivion®DP treatment initiation) will be compared between the groups using a similar model as the primary endpoint using the HDRS-17 score at each visit instead of the change.
- HDRS-17 Category Shift from Visit 2 to Visit 4 – defined as the category change from baseline to week-8 post Relivion®DP treatment initiation in HDRS-17; Categories definition: 5: Very severe 26-52, 4: Severe 20-25, 3: Moderate 14-19, 2: Mild 13-8 and 1: 0-7 Remission, will be summarized by a count and

percentage and compared between the groups with a Cochran-Armitage trend test.

- Mean change in Q-LES-Q score at visit 4 – measured as the difference between the baseline and visit 4 Q-LES-Q percentage of maximum total scores, will be compared between the groups using analysis of variance (ANOVA).

11.5.6 Safety Analyses

Descriptive statistics will be presented per study group for all safety parameters.

The primary safety variable, the cumulative incidence (and 95% CI) of device related adverse events (AEs) throughout the study, will be presented in tabular format and will include incidence tables by severity.

Adverse event rates will be compared between the study groups with a Fisher's exact test.

Serious adverse events will be listed and discussed individually.

Treatment tolerability will be compared between the study groups. The number and percent of subjects who fail to complete the study and the number and percent of subjects who fail to complete the study because of Adverse Events will be presented as well.

11.5.7 Additional Exploratory Endpoints and Analyses - Open Label Stage

1. The following endpoints will only be evaluated in the Relivion®DP treatment group for the mITT analysis set:

    - Change from Baseline to 16 weeks post Relivion®DP treatment initiation in HDRS-17 (Visit 6).

      The change from baseline HDRS-17 score will be modeled with a repeated measures ANCOVA model. The change will be modeled as a function of visit (categorical) with baseline value entered as a covariate. LSmean changes in HDRS-17 scores will be estimated from the model per visit and for the difference between visits and will be presented with 95% confidence intervals as well as level of significance (testing the null hypothesis

LSmean=0). The change from baseline in HDRS-17 score at visit 6 will also be compared to the change at visit 4.

- **Proportion of HDRS-17 responder subjects to visit 6.** The percent of subjects achieving at least 50% reduction from baseline in HDRS-17 total score to each visit in the original active group will be summarized by a count and percentage and compared between the visits using a GEE model (repeated measures model for binary data) programmed in the SAS GENMOD Procedure which will account for the within subject correlation, baseline HDRS-17 score will be entered as a covariate.
- **Proportion of subjects achieving clinically substantial improvement in HDRS-17 scores to visit 6**, defined as the percent of subjects with reduction in HDRS-17 score of at least 7 points at 16 weeks post Relivion®DP treatment initiation.

  The percent of subjects achieving clinically substantial improvement in HDRS-17 scores at each visit in the original active group will be summarized by a count and percentage and compared between the visits using a GEE model (repeated measures model for binary data) programmed in the SAS GENMOD Procedure which will account for the within subject correlation, baseline HDRS-17 score will be entered as a covariate.
- **Proportion of subjects achieving remission at visit 6** - defined as the percent of subjects with HDRS17 score≤7 at 16 weeks post Relivion®DP treatment initiation in the original active group. The percent of subjects achieving remission at each visit in the original active group will be summarized by a count and percentage and compared between the visits using a GEE model (repeated measures model for binary data) programmed in the SAS GENMOD Procedure which will account for the within subject correlation, baseline HDRS-17 score will be entered as a covariate.
- **HDRS-17 Category Shift from Visit 2 to Visit 6** – defined as the category change from baseline to week-16 post Relivion®DP treatment initiation in HDRS-17; Categories definition: 5: Very severe 26-52, 4: Severe 20-25, 3:

Moderate 14-19, 2: Mild 13-8 and 1: 0-7 Remission. The category shift will be tabulated and summarized by a count and percent.

- **Proportion of new responder subjects** - defined as the percent of subjects who did not have a remission (per HDRS-17) at visit 4, achieving at least 50% reduction from baseline (Visit 2) in HDRS-17 total score 16 weeks post Relivion®DP active treatment initiation (Visit 6). A count and percentage with two-sided exact 95% confidence interval will be presented.
- **Proportion of new subjects achieving remission** - defined as the percent of subjects who did not have a remission (per HDRS-17) at visit 4, achieving a HDRS-17 total score≤7 points 16 weeks post Relivion®DP active treatment initiation (Visit 6). A count and percentage with two-sided exact 95% confidence interval will be presented.
- **Change from baseline** to 16 weeks (Visit 6) post Relivion®DP treatment initiation **in the Clinical Global Impression Severity scale (CGI-S) score.** The change from baseline CGI-S score will be modeled with a repeated measures ANCOVA model. The change will be modeled as a function of visit (categorical) with baseline value entered as a covariate. LSmean CGI-S changes from baseline will be estimated from the model per visit and for the difference between visits and will be presented with 95% confidence intervals as well as level of significance (testing the null hypothesis LSmean=0). The change at visit 6 will also be compared to the change at visit 4.
- **Clinical Global Impression Improvement (CGI-I) score** over time **to Visit 6**. The CGI-I score (from visit 3) will be modeled with a repeated measures ANCOVA model. The score will be modeled as a function of visit (categorical). LSmean CGI-I scores will be estimated from the model per visit and for the difference between visits and will be presented with 95% confidence intervals as well as level of significance (testing the null hypothesis LSmean=0). The CGI-I score at visit 6 will be compared to the score at visit 4.

- **Change from baseline to week-16 (visit 6) post Relivion®DP treatment initiation in the total score of the Quick Inventory of Depressive Symptomatology (self-rated) (QIDS-SR-16).** The change from baseline QIDS-SR-16 score will be modeled with a repeated measures ANCOVA model. The change will be modeled as a function of visit (categorical) with baseline value entered as a covariate. LSmean QIDS-SR-16 changes from baseline will be estimated from the model per visit and for the difference between visits and will be presented with 95% confidence intervals as well as level of significance (testing the null hypothesis LSmean=0). The change at visit 6 will also be compared to the change at visit 4.
- **Change from baseline Q-LES-Q score at Visit 6** – measured as the difference between the baseline and visit 6 Q-LES-Q percentage of maximum total scores. The change from baseline Q-LES-Q score will be modeled with a repeated measures ANCOVA model. The change will be modeled as a function of visit (categorical) with baseline value entered as a covariate. LSmean Q-LES-Q changes from baseline will be estimated from the model per visit and for the difference between visits and will be presented with 95% confidence intervals as well as level of significance (testing the null hypothesis LSmean=0). The change at visit 6 will also be compared to the change at visit 4.

2. The following endpoints will be analyzed on all subjects that completed visit 6 grouped together regardless of prior double-blind treatment:
   - **SIQ (Satisfaction Questionnaire) Question #2**: Proportion of subjects rating the subjects convenience (score of 1, 2 and 3) to undergo a depression treatment that required a daily treatment (5-7 days a week) or maintenance treatment, defined as Very convenient, Somewhat convenient or Neutral, neither convenient nor inconvenient using a simple Likert-type scale whereas 1 = Very convenient, 2 = Somewhat convenient, 3 = Neutral, neither convenient nor inconvenient, 4 = Somewhat inconvenient, 5 = Very inconvenient, will be summarized by a count and

percentage and will be presented with two-sided exact 95% confidence interval.

- **SIQ (Satisfaction Questionnaire) Question #3**: Proportion of subjects rating the subjects convenience (score of 1, 2 and 3) to undergo a home based self-operated depression treatment, defined as Very convenient, Somewhat convenient or Neutral, neither convenient nor inconvenient using a simple Likert-type scale whereas 1 = Very convenient, 2 = Somewhat convenient, 3 = Neutral, neither convenient nor inconvenient, 4 = Somewhat inconvenient, 5 = Very inconvenient, will be summarized by a count and percentage and will be presented with two-sided exact 95% confidence interval.
- **SIQ (Satisfaction Questionnaire) Question #6**: Proportion of subjects rating the subjects ease (score of 1, 2 and 3) to administer the Relivion®DP treatment, defined as Very easy, easy , Neutral, using a simple Likert-type scale whereas 1 = Very easy, 2 = Easy, 3 = Neutral, 4 = Challenging, 5 = Difficult, will be summarized by a count and percentage and will be presented with two-sided exact 95% confidence interval.
- **SIQ (Satisfaction Questionnaire) Question #11**: Proportion of subjects satisfied from the study device effect on their depression status (score of 1 and 2), defined as Very satisfied and Satisfied, using a simple Likert-type scale whereas 1 = Very satisfied, 2 = Satisfied, 3 = Unsure, 4 = Dissatisfied, 5 = Very dissatisfied, will be summarized by a count and percentage and will be presented with two-sided exact 95% confidence interval.

3. The following endpoint will be evaluated in all subjects who did not achieve remission in the double-blind phase for the mITT analysis set regardless of prior double-blind treatment:

   - **Change from visit 4 to 16 weeks post Relivion®DP treatment initiation in HDRS-17** (Visit 6). The change from visit 4 HDRS-17 score will be modeled with an ANCOVA model with baseline value entered as a covariate. LSmean changes in HDRS-17 scores will be estimated from the model and will be

presented with 95% confidence intervals as well as level of significance (testing the null hypothesis LSmean=0).

STATISTICAL

Final Audit Report 2024-05-29

Created:	2024-05-29
By:	Lilach Aizen (Lilach.Aizen@neurolief.com)
Status:	Signed
Transaction ID:	CBJCHBCAABAAcKzPbHHLGM-xTDT9BogVTL12s5EmyqHM

"STATISTICAL" History

Document created by Lilach Aizen (Lilach.Aizen@neurolief.com)
2024-05-29 - 11:16:33 AM GMT

Document emailed to Netanel Deutsch (netanel@ldbiostats.com) for signature
2024-05-29 - 11:16:39 AM GMT

Document emailed to Lisa Deutsch (lisa@ldbiostats.com) for signature
2024-05-29 - 11:16:39 AM GMT

Document emailed to Yaron Gruper (yaron.gruper@neurolief.com) for signature
2024-05-29 - 11:16:39 AM GMT

Document emailed to Michal Kedar-Datel (michal.kedar-datel@neurolief.com) for signature
2024-05-29 - 11:16:39 AM GMT

Email viewed by Lisa Deutsch (lisa@ldbiostats.com)
2024-05-29 - 11:17:44 AM GMT

Document e-signed by Lisa Deutsch (lisa@ldbiostats.com)
Signature Date: 2024-05-29 - 11:18:49 AM GMT - Time Source: server

Email viewed by Yaron Gruper (yaron.gruper@neurolief.com)
2024-05-29 - 11:29:52 AM GMT

Document e-signed by Yaron Gruper (yaron.gruper@neurolief.com)
Signature Date: 2024-05-29 - 11:31:46 AM GMT - Time Source: server

Email viewed by Michal Kedar-Datel (michal.kedar-datel@neurolief.com)
2024-05-29 - 12:16:36 PM GMT

Document e-signed by Michal Kedar-Datel (michal.kedar-datel@neurolief.com)
Signature Date: 2024-05-29 - 12:21:54 PM GMT - Time Source: server

Email viewed by Netanel Deutsch (netanel@ldbiostats.com)

2024-05-29 - 3:31:37 PM GMT

Document e-signed by Netanel Deutsch (netanel@ldbiostats.com)

Signature Date: 2024-05-29 - 3:32:12 PM GMT - Time Source: server

Agreement completed.

2024-05-29 - 3:32:12 PM GMT