Statistical Analysis Plan

Clinical Trial Protocol: NND-3002

Product Name: Low Emission NeoSync EEG Synchronized TMS Technology for the Treatment of Major Depressive Disorder (N.E.S.T.-MDD)

Sponsor: NeoSync, Inc.

Study Titles: **NND-3002**

A Prospective, Multicenter, Double-Blind, Sham-Controlled Adaptive Design Study to Confirm the Safety and Efficacy of NEST sTMS in Subjects with Major Depressive Disorder Who Have Not Responded to at Least One Antidepressant Medication in the Current Episode

Prepared by

Joseph Massaro, Ph.D.
34 Robinson Drive.
Bedford, MA 01730

January 26, 2018

ANALYSIS PLAN SIGNATURE PAGE

Protocol: **NND-3002**

Study Title: **A Prospective, Multicenter, Double-Blind, Sham-Controlled Adaptive Design Study to Confirm the Safety and Efficacy of NEST sTMS in Subjects with Major Depressive Disorder Who Have Not Responded to at Least One Antidepressant Medication in the Current Episode**

Protocol Date: 7 January 2018 (NND-3002)

The undersigned have reviewed this analysis plan and approve of it in its entirety.

Signature

Joseph M. Massaro, Ph.D.

Kathryn Rumrill
NeoSync, Inc.

Table of Contents

INTRODUCTION

This Statistical Analysis Plan describes the analysis to be performed for NeoSync, Inc. study NND-3002. The purpose of the Statistical Analysis Plan for this study is to provide a framework in which answers to the protocols' objectives may be achieved in a statistically rigorous fashion, without bias or analytical deficiencies. Specifically, this Plan has the following purpose:

To prospectively (a priori) outline the types of analyses and presentations of data that will form the basis for conclusions to be reached that will answer the studies' objectives outlined in the protocol, and to explain in detail how the data will be handled and analyzed, adhering to commonly accepted standards and practices of biostatistical analysis in the medical device industry.

STUDY OBJECTIVES

The primary objective of this study is to confirm the safety and effectiveness of synchronized TMS (sTMS) therapy in subjects with Major Depressive Disorder who have not responded adequately to at least one anti-depressant medication in the current episode.

STUDY DESIGN

Prospective, randomized, double-blinded, sham-controlled, parallel group adaptive design study to confirm the safety and efficacy of sTMS in subjects with major depressive disorder who have not responded to at least one anti-depressant medication in the current episode. This is a multicenter study in which approximately 120 subjects will be enrolled in the trial to obtain 92 evaluable subjects who fully complete the primary study phase of the trial. Randomized subjects will be treated 5 days per week for 6 weeks. After signing informed consent, subjects who qualify for enrollment will discontinue use of their current antidepressant treatment (if applicable). Subjects must have discontinued the antidepressant medication a minimum of 1 week prior to their baseline measurements and initiation of treatment with the randomized treatment (sTMS or sham). Following wash-out of the antidepressant medication, an additional evaluation will be performed to determine whether the protocol eligibility criteria are still met before randomized treatment.

Qualified subjects will be randomized in a 1:1 manner to either sTMS or sham treatment groups. Treatment will be initiated on Day 1 of the study. Subjects will visit the clinic for 5 daily treatment sessions per week for a total of 6 "treatment" weeks. A visit interval of 5-10 calendar days has been set for each "treatment" week. Treatment will be discontinued at the end of Treatment Week 6 (30 treatment sessions). Subjects will be clinically evaluated for safety and efficacy at the end of each of the six weekly treatment courses. At the end of Treatment Week 6, subjects will have completed the primary study phase treatments and will be offered either open label sTMS therapy (if criteria are met) or alternate treatment as clinically indicated.

TIMING OF SUBJECT ENROLLMENT

Once a subject has been consented, the screening process begins. The screening/baseline assessments may be split into two or more days, one pre-washout that will occur before the antidepressant medication washout period, and one post-washout, which will be done after washout but before randomization. The post-washout screening will be used as the baseline for

the study. Once eligibility requirements in both pre-washout and post-washout screenings have been met, the subject is enrolled in the study.

WITHDRAWAL AND REPLACEMENT OF SUBJECTS

Subjects have the right to withdraw from the study at any time, for any reason, without jeopardizing their medical care. The Investigator, the IRB and NeoSync, Inc. also have the right to withdraw subjects or terminate the trial for the following reasons:

- Occurrence of unacceptable risk to the subjects enrolled in the study, including significant worsening of depression.
- A decision on the part of the Sponsor to suspend or discontinue testing, evaluation, or development of the product.

The Investigator may use the weekly clinical evaluation to determine if the subject's symptoms or quality of life have significantly deteriorated to the point where withdrawal from the study would be in the subject's best interests.

A study conducted at a single study site or a single study site in a multicenter study may also warrant termination under the following conditions:

- Failure of the Investigator to comply with pertinent regulations of appropriate regulatory authorities.
- Insufficient adherence to protocol requirements.
- Submission of knowingly false information from the research facility to the Sponsor, study monitor, or appropriate regulatory authority.

Subjects who withdraw from the study for any reason will not be replaced. Subjects who voluntarily withdraw from the study will be asked to complete the Week 6/Early Termination assessments. Subjects who are withdrawn due to adverse events will be followed until resolution or stabilization of the adverse event.

SAMPLE SIZE

The primary endpoint of the study is clinical response at Week 6, defined by 50% or greater reduction on the Hamilton Rating Scale for Depression (HAM-D[17]) from baseline (Day 0). The null and alternative hypotheses of interest are:

$$H0: \pi A - \pi S = 0$$

vs.

$$Ha: \pi A - \pi S \neq 0$$

where πA is the proportion of subjects who show 50% or greater improvement in HAM-D[17] total score from baseline (Day 0) to Week 6 for the active sTMS treatment and πS is the equivalent value for the sham-controlled group.

The calculation of sample size requirements for this confirmatory study is based on the results observed in the per-protocol (PP) analysis of non-treatment-naïve subjects who completed Week

6 HAM-D[17] assessments in the previous NEST trial NND-3001.

In that population, 34% of subjects treated with active sTMS were categorized as responders, whereas only 8% of sham treated subjects responded. Assuming true population response rates of 34% for sTMS and 8% for Sham, power analysis indicates that a PP sample size of 92 subjects (i.e., 92 subjects, or 46 subjects per treatment group, without major protocol violations and with HAM-D[17] measured at Week 6) yields 81% power to demonstrate superiority of sTMS over Sham, where the statistical test is Fisher's exact test at a two-sided 0.05 level of significance.

The sample size ratio for the active treatment group to sham-controlled group is 1:1. Based on a conservative estimate of published studies, approximately 23% of enrolled subjects may not be part of the PP population in each treatment group. Thus, a total of 120 subjects (60 per group) will therefore be enrolled in the study to account for potential drop-outs and protocol noncompliance prior to Week 6.

SCHEDULE OF VISITS

Table A on the following page displays the schedule of visits and procedures for protocol NND-3002. Full details can be found in the protocol.

TABLE A: STUDY FLOWCHART

ACTIVITY	SCREEN	BASELINE (DAY <0)	WK1 TX DAY 5	WK2 TX DAY 10	WK3 TX DAY 15	WK4 TX DAY 20	WK5 TX DAY 25	WK 6 / EARLY TERM TX DAY 30	WK7 OPEN LABEL TX DAY 5	WK8 OPEN LABEL TX DAY 10	WK9 OPEN LABEL TX DAY 15	WK 10,11,1 OPEN LABE TX DAY 20,25,30
Informed Consent	X											
Inclusion/Exclusion Criteria	X	X										
Mini Mental Status Exam (Version 2)	X											
MINI	X											
ATHQ	X											
Blinding Questionnaire			X					X				
Randomization		X										
NeoSync EEG	X	X										
Med/Psych History	X											
Vital Signs		X						X				
Labs[a]	X											
Pregnancy Test (HCG)[b]	X											
Urine Drug Screen[b]	X											
Weight		X						X				
Physical Exam[a]	X											
Demographics	X											
ECG[a]	X											
HAM-D/MADRS	X	X	X	X	X	X	X	X	X	X	X	X
IDS-SR		X	X	X	X	X	X	X	X	X	X	X
C-SSRS	X	X	X	X	X	X	X	X	X	X	X	X
Concomitant Med[c]	X	X	X	X	X	X	X	X	X	X	X	X
Adverse Events[c]	X	X	X	X	X	X	X	X	X	X	X	X

a Optional
b Must be performed at Screening and may be performed at any subsequent visit at the discretion of the Investigator
c In addition to the assessment visits outlined here, All subjects will have 5 treatment visits per week during treatment period changes in adverse events and concomitant medications should be recorded when made known to the site.
Note: The first treatment session may begin on the same day as baseline visit, after subject meets all criteria and is randomized to a treatment arm.

RANDOMIZATION

This study is randomized, double-blind, and sham-controlled to minimize bias. Subjects will be randomized following the post-washout screening/baseline analysis (Day 0) to one of the two treatment groups (sTMS vs. sham treatment), at a 1:1 ratio stratified by study center. The assignment is determined by a computer-generated randomization table defined prior to the start of the trial.

INITIAL SCREENING/BASELINE VISIT:

Potential study participants will be screened for eligibility per the protocol eligibility criteria. The Screening/Baseline visits may be split into two or more days. After informed consent has been signed and washout from previous medications is complete (when necessary), the following assessments will be performed as part of the subject eligibility assessment:

- MINI (Mini International Neuropsychiatric Interview)
- ATHQ (Antidepressant Treatment History Questionnaire)
- Mini Mental Status Exam (MMSE-2)
- Medical and psychiatric history
- Concomitant medication history
- Adverse Events
- Demographics
- NeoSync EEG
- Vital Signs (Heart Rate [HR], Blood Pressures [BP])
- Weight
- Pregnancy Test (HCG)
- Urine Toxicology Screen
- HAM-D/MADRS (Montgomery-Åsberg Depression Rating Scale)
- IDS-SR (Inventory of Depressive Symptomatology – Self Report)
- C-SSRS (Columbia Suicide Severity Rating Scale)
- Standard 12 lead ECG (optional)
- Physical Examination (optional)
- Laboratory Assessments (optional)

FOLLOW-UP VISITS

General Overview: The first treatment (Day 1) may occur on the same day as the baseline visit (Day 0) if all study criteria are met and randomization procedures are complete, or subjects will be asked to return on a subsequent day to begin the weekly treatment schedule. Subjects will come to the clinic for 5 daily treatment sessions per treatment week for a total of 30 treatment sessions over 6 treatment weeks. A consistent Monday through Friday weekly treatment schedule will be highly encouraged. However, a visit interval of 5-10 calendar days will be allowed for each treatment week to accommodate visit scheduling adjustments when needed. Treatment will be discontinued at the end of Week 6 (30 treatment sessions). Subjects will be clinically evaluated for safety and efficacy at the end of each of the six weekly treatment

courses. Subjects who have not remitted (remission defined as a HAM-D[17] score of ≤7 (Zimmerman et al., 2004)) at the Week 6 evaluation visit may be eligible to receive open label sTMS during the six-week follow-up phase. If so, additional weekly treatment evaluation visits shall be performed at the end of treatment Week 7, 8, 9, 10, 11 and 12. Subjects who do not meet response criterion (relative to pretreatment baseline, Day 0) following 3 weeks (15 treatments) of open label treatment may be offered additional concomitant antidepressant therapy for the final 3 weeks of open label treatment.

Alternatively, if the clinician feels it is not appropriate for the subject to participate in the open label sTMS follow-up, the subject chooses not to participate in the open label follow-up or if the subject has remitted in the blinded phase, subject will be deemed to have completed participation in this study.

Specifics of the study visits are as follows:

Double-Blind Treatment Visits (Week 1-6): After the subject has been deemed to be acceptable for inclusion in the trial and randomized to a treatment group, they may begin scheduled treatments. They will go to the physician's office 5 days per treatment week to complete the treatment procedure. In addition, subjects will be assessed for changes in adverse events, concomitant medications, or worsening of symptoms that may need physician intervention at each of the treatment visits. The first treatment visit may occur after randomization at the baseline visit given sufficient time.

Double-Blind Evaluation Visits (Week 1-5): In addition to the treatment visits, an evaluation will be conducted on the same day as the last treatment (5th treatment session, within 10 calendar days) of each weekly treatment cycle. The following assessments will be performed as part of that evaluation:

- HAM-D/MADRS
- IDS-SR
- C-SSRS
- Adverse Events
- Concomitant Medications
- Treatment Blinding Questionnaire(s) (Week 1)

Week 6 /Early Termination Visit

If assessments are done as part of an early discontinuation visit, subject should be asked to come in for the post-treatment follow-up visit.

- Vital Signs
- Weight
- HAM-D/MADRS
- IDS-SR
- C-SSRS
- Adverse Events
- Concomitant Medications

- Treatment Blinding Confirmation Questionnaire(s)
- Determination of eligibility for open label sTMS treatment

Open Label Treatment and Evaluation Visits (Week 7-12): In addition to the treatment visits, an evaluation visit will be conducted on the last day of each weekly treatment cycle. The following assessments will be performed as part of that evaluation:

- HAM-D/MADRS
- IDS-SR
- C-SSRS
- Adverse Events
- Concomitant Medications

STATISTICAL COHORTS

Intent to Treat

The Intent-to-Treat (ITT) Population is defined as all randomized patients. Following the ITT principle, subjects will be analyzed according to the treatment they were assigned to at randomization.

Safety

The Safety Population is defined as all subjects who received at least one study treatment. Subjects will be analyzed according to treatment received. This is the primary analysis population for safety.

Per Protocol

The Per-Protocol (PP) study sample is comprised of randomized subjects who receive the correct randomized treatments according to the protocol schedules without any major protocol violation. This is the primary analysis population per e-mail discussion between NeoSync and the Food and Drug Administration (FDA) on July 17, 2015. Major protocol violations include but are not limited to:

- Receiving treatments other than the randomized treatments.
- Treatments not administered according to study protocol.
- Subjects cannot miss more than 1 dose in a treatment week and their weekly visit interval cannot exceed 10 calendar days. Any subject who misses more than 6 doses in the 6 week treatment period is not considered compliant with the treatment and will not be included in the PP analysis population.
- Receiving concomitant medications that are not allowed per the study protocol; the prohibited medications are as follows:

  Use of the following concomitant medications is prohibited within 72 hours of the baseline EEG (Day 0), but is otherwise allowable in the study:
  - Anti-Migraine

- Benzodiazepines and Partial Benzodiazepine Agonists
- Non-Benzodiazepine Hypnotics
- Non-Sedating Antihistamines
- Cough Suppressants (Dextromethorphan, etc.)

Uses of the following concomitant medications are excluded throughout the study and should be washed out at least 1 week prior to baseline (Day 0):

- Anticholinergics
- Anticonvulsants
- Antidepressants (note: fluoxetine shall be washed out 4 weeks prior to baseline)
- Antihistamines (with sedating effects)
- Anti-Parkinsonian medications
- Antipsychotics
- Barbiturates
- Codeine or Oxycodone containing compounds
- Corticosteroids (tablet forms only)
- Cold Pills (with sedating ingredients)
- Medications that are known to reduce seizure thresholds:
    a. Accurbron/Accubron
    b. Doxepin
    c. Ambenonium Chloride
    d. Amitriptyline
    e. Imipramine
    f. Amoxapine
    g. Trazodone
    h. Amphetamine sulfate
    i. Baclofen
    j. Chlorpromazine
    k. Clozaril
    l. Cocaine
    m. Lithium
    n. Protostat
- Mood stabilizers
- Muscle relaxants
- Other sleep aid medications with anticholinergic properties (prescription OR over-the-counter)
- Opioid analgesics
- Pseudoephedrine combination forms that contain excluded medication (pseudoephedrine (e.g., Sudafed) in pure form is acceptable during the study)
- Psychostimulants

The following Nutritional supplements are excluded throughout the study and should be washed out at least 1 week prior to baseline (Day 0):

- CQ-10

- Ginkgo Biloba
- Ginseng containing supplements
- Sam-E
- St John's Wort
- High doses of Omega-3 fatty acids and fish oil supplements (note: this is acceptable if subject is taking ≤1000mg QD or reported taking the supplements for at least one month prior to starting the study).

- HAM-D17 not available at Week 6.

The final major protocol violations will be reviewed and defined prior to unblinding the treatment randomization.

STATISTICAL ANALYSES

General Overview of Methods

This is a randomized, double-blinded, sham-controlled, parallel group study to evaluate the effectiveness of sTMS in subjects with Major Depressive Disorder (MDD). The objective of this study is to establish the efficacy of sTMS in comparison to sham treatment in subjects with major depressive disorder.

The primary endpoint is clinical response, defined by 50% or greater reduction on the HAM-D17 total symptom scores following 6 weeks of double-blind treatment (Week6).

The data analysis on the double-blind portion will be performed when all randomized subjects have completed the double-blind Week 6 visit or have withdrawn from the study prior to completion of the double-blind phase. The data analysis on the open-label portion will be performed when all open-label subjects have completed the Week 12 visit or have withdrawn from the study prior to the open-label Week 12 visit.

The standard descriptive statistics such as mean, standard deviation, median, quartiles, minimum, and maximum will be used to summarize the continuous variables. For discrete variables, the frequency and percentage of patients in each category will be calculated.

Sites with fewer than 10 subjects in the PP group will be pooled with other centers by geographic region prior to carrying out analyses described below. All statistical tests will be two-sided with a significant level of 0.05. All confidence intervals will be two-sided with a confidence level of 95%. Statistical analyses will be conducted with SAS Version 9.4 or larger.

For the six-week double-blind phase, imputation of data on the primary endpoint will be performed as described below. For the six-week follow-up open-label phase (Weeks 6 – 12), all efficacy and safety summaries will be based on the all sTMS treated subject population, which consists of all subjects who have entered the six-week follow-up Phase. All efficacy analyses will be performed on the Observed Cases (OC) dataset only for each week in this open-label Phase. That is, no imputation will be used, and only data observed during the follow-up Phase will be summarized. Analyses in the open-label phase will be presented by double-blind

treatment group and for both treatment groups combined. Baseline for the double-blind phase will be defined as the last available value measured post-washout and prior to the first double-blind dosing. Baseline for the open-label phase will be defined as (a) the last post-washout assessment measured prior to the beginning of the double-blind treatment phase for patients randomized to sTMS in the double-blind phase; and (b) the last double-blind assessment measured prior to the beginning of the open-label phase for patients randomized to Sham in the double-blind phase.

Analysis of Primary Efficacy Endpoint

HAM-D/MDRS will be assessed at screening, baseline, Week 1-6 Evaluation Visits, and during the open-label phase through Week 12 evaluation visit. The primary efficacy endpoint is clinical response, defined by 50% or greater reduction on the HAM-D[17] from baseline to Week 6. The efficacy analyses will be performed on the PP, ITT with available data, and ITT with imputed data (see below for imputation methods) analysis populations. The PP analysis population is the primary analysis population for efficacy per e-mail discussion between NeoSync and the FDA on July 17, 2015.

The null hypothesis is that the active sTMS treatment group does not differ from the sham group with respect to the proportion of patients with clinical response at Week 6. The corresponding alternative hypothesis is that the active sTMS treatment group differs from the sham group with respect to the proportion of patients with clinical response at Week 6.

Specifically, the primary efficacy analysis will be a test of the hypotheses:

$$H0: \pi A - \pi S = 0$$

and

$$Ha: \pi A - \pi S \neq 0$$

where πA is the proportion of subjects who show 50% or greater improvement in mean HAM-D17 total score from baseline (Day 0) to Week 6 for the active treatment and πS is the equivalent value for the sham-controlled group. The test of the null hypothesis will be based on Fisher's exact test at a two-sided 0.05 level of significance. In addition, the observed response rates and two-sided 95% confidence intervals of the risk difference between treatments will be presented, calculated using the Wilson method discussed in Newcombe (1998).

An assessment of treatment-by-investigative center interaction on the proportion of clinical responders at Week 6 will be carried out using the Breslow-Day test. Sites with fewer than 10 subjects in the PP group will be pooled with other centers by geographic region prior to carrying out analyses (the same pooled sites will then be used for the analyses on the ITT analysis population). The pooling will be carried out before the blind is broken. A 0.15 level of significance will be used to assess the interaction. A non-significant interaction or an interaction significant but only quantitative in nature will support the pooling of results across study centers for the final analysis.

For the six-week open-label follow-up phase, the proportion of patients with a clinical response at each open-label visit for each double-blind treatment and for both treatments combined will be presented. The proportion of patients with a clinical response at each visit will be compared between double-blind treatment in a similar manner as above. These open-label analyses are

considered exploratory.

INTERIM ANALYSES

Due to uncertainty regarding the expected sham response rate and resulting treatment effect differential, this study will utilize an adaptive design that allows one interim sample size re-estimation after a minimum of 70 subjects have been enrolled and completed 6 weeks of treatment (or dropped out prior to 6 weeks). The pre-specified maximum allowable adjusted sample size following re-estimation will be 240 randomized, or two times the initial planned sample size of 120 enrolled, based on practical resource constraints and the desire to achieve a minimum differential of 12% between active and sham response rates. The re-estimation of sample size will be conducted by an independent biostatistician following a pre-specified plan as described by Mehta and Pocock, 2010 and presented to the independent data monitoring board (DMB).

Specifically, at the interim stage the conditional power (CP) for obtaining a significant beneficial effect of sTMS over sham with respect to the primary endpoint will be calculated by the independent statistician, using the protocol-specified planned sample size of 92 *evaluable* patients. This conditional power will be calculated using formula (6) in Mehta and Pocock (2010) under the assumption that the interim estimate of the risk difference is the true population risk difference. The study will not be stopped for overwhelming efficacy or futility at this interim stage. Based on Table 1 of Mehta and Pocock (2010), if the CP $\leq$ 36% or CP $\geq$ 90%, the study will continue to its original planned sample size of 92 evaluable subjects (120 enrolled subjects). Otherwise if 36% < CP <90% (the Mehta and Pocock "promising zone"), then the sample size may be increased up to 240 enrolled subjects (~184 evaluable patients) in order to obtain enough evaluable subjects to yield up to 90% CP for the primary analysis (if the maximum increase to 240 enrolled patients does not yield 90% CP, the sample size may still be increased to 240 to maximize CP, at the recommendation of the independent DMB). This approach to sample size increase does not inflate the overall Type I error.

The sponsor and investigators will remain blinded to the interim results for the duration of the ongoing study. After the interim analysis is carried out, the recommendation made by the DMB to the sponsor will only be to keep sample size as is or to increase enrolled sample size to a given value; the reason for the recommendation will not be given to the sponsor.

OTHER STUDY PARAMETERS AND ANALYSES

The following will be carried out for the ITT and PP analysis populations unless otherwise specified.

A. Subject Disposition

The number and percentage of subjects in each population will be presented by treatment group. The number and percentage of subjects in each of the ITT and PP populations who prematurely discontinued prior to Week 6 will be presented overall and by reason of discontinuation within each treatment group. The number and percentage of ITT and PP subjects will be presented by investigative center within each treatment group.

The number and percentage of ITT and PP subjects entering the open-label phase will be presented overall and by original double-blind treatment group.

B. Baseline Demographics

Descriptive statistics for demographic variables as well as baseline (post-washout and pre-start of double-blind phase) HAM-D17 total score and subscores, baseline Mini Mental Status Exam (MMSE) total score, and baseline values of other scales will be presented by treatment group for the ITT and PP populations. The standard descriptive statistics such as mean, standard deviation, median, quartiles, minimum, and maximum will be used to summarize the continuous variables. For discrete variables, the frequency and percentage of patients in each category will be calculated.

For the follow-up open-label phase, descriptive statistics for background and demographic variables will be presented according to previous double-blind treatment and for all follow-up subjects combined.

The following are the demographic variables collected at baseline:

Age

Sex (Male, Female)

Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Hispanic/Latino, Other)

Education Level (Some high school or less, High school diploma/GED, Vocational School, Some college, College degree, Professional or graduate degree)

The demographic profile of subjects enrolled in the present study will be compared in an exploratory descriptive manner to the non-Naïve cohort of subjects enrolled in the previous NEST study to assess comparability of study populations; this will be carried out for the ITT and PP analysis populations of the present study.

C. Medical History

The number and percentage of subjects with various medical histories, categorized by severity when available, will be presented by treatment group for the ITT and PP populations.

D. Mini International Neuropsychiatric Interview (MINI)

The number and percentage of subjects with each MINI diagnosis at baseline (suicidality, manic/hypomanic disorders, panic disorder, agoraphobia, social phobia, obsessive-compulsive disorder, posttraumatic stress disorder, alcohol dependence/abuse, substance dependence/abuse, psychotic disorders and mood disorder with psychotic features, anorexia nervosa, bulimia nervosa, generalized anxiety disorder, antisocial personality disorder) will be presented by treatment group for the ITT and PP populations. For patients entering the follow-up open-label phase, this will be

repeated and be presented according to previous double-blind treatment and for all follow-up subjects combined.

E. Antidepressant Treatment History

The following will be carried out for the ITT and PP analysis populations entering the double-blind phase.

The number and percentage of subjects within each of the following categories will be presented by treatment group: resistant to 2 antidepressant medications at adequate dose and duration, resistant to 3 antidepressant medications at adequate dose and duration, and resistant to 4 or more antidepressant medications. These number and percentages will also be presented by medication class (SSRI, SSNI, MAOI, NDRI, Other).

The number and percentage of subjects within each of the following categories will be presented by treatment group: inadequate dose or duration of 1 or more antidepressant medications, intolerant to 1 or more antidepressant medications. The number and percentage of patients who were intolerant to at least 1 medication or who experienced at least 1 inadequate trial will also be presented by medication class (SSRI, SSNI, MAOI, NDRI, Other).

Descriptive statistics of the approximate number of unsuccessful trials of adequate dose and duration, and of the approximate number of trials for which medication could not be tolerated, will be presented by treatment group.

For the follow-up open-label phase, similar analyses will be presented according to previous double-blind treatment and for all follow-up subjects combined.

F. Sensitivity Analyses on the Primary Efficacy Endpoint

The primary hypothesis in the double-blind phase will be tested using Fisher's Exact test on the ITT and PP analysis populations. For the ITT population, there may be missing data at Week 6 due to, for example, patients prematurely withdrawing from the study. The first analyses for the ITT population will use available data. As a supportive sensitivity analysis, multiple imputation will be used to assess the robustness of the primary endpoint findings relative to missing data. Specifically, multiple imputation using the linear regression approach will be used to impute missing HAM-D[17] across visits via SAS PROC MI; included as covariates in the linear model will be age, sex, race, education level, HAM-D17 measured at previous visits, investigative center, baseline MMSE, baseline IDS-SR, baseline C-SSRS, and ATHQ categorizations. Ten complete datasets will be created (the linear regression approach to multiple imputation assumes monotone missing data pattern; if such a pattern does not exist, then prior to each of the 10 linear regression imputations, any non-monotone missing data will first be imputed via the Monte-Carlo Markov Chain approach to create a monotone missing data pattern). Once 10 complete HAM-D[17] datasets are generated, the determination of whether the patient achieved imputed clinical response at Week 6 will be calculated for each patient with missing data within each imputed dataset. The z-test for proportions with continuity correction (which yields very similar results to Fisher's Exact test) will then be applied to each of these 10 datasets to compare treatments on

the incidence of Week 6 clinical response. The numerator and denominator of the 10 continuity-correct z -statistics will then be combined into a single continuity-corrected z-statistic using SAS PROC MIANALYZE. The resulting p-value from this z-statistic, assessing overall significance of treatment difference on clinical response rate at Week 6, will be calculated.

In addition, a tipping point analysis will be performed to assess the likelihood that the observed significance would no longer hold based on various assumed values of the missing clinical response data. Specifically, to start, at Week 6, all missing data will be assumed to be successes for both the sTMS and sham groups. Fisher's exact test will then be run comparing treatments on this imputed clinical response rate and the treatment comparison Week 6 p-value will be generated. Then, the analyses will be repeated where 1 missing sTMS patient is assumed to be a failure, and all remaining sTMS and all Sham patients are assumed to be successes. Then the analyses will be repeated where 2 missing sTMS patients are assumed to be a failure, and all remaining sTMS and all Sham patients are assumed to be successes. This process will continue until the "tipping point" is reached. The "tipping point" is the imputation point at which the Fisher's Exact test p-value changes from significance to non-significance.

The above process will then be repeated but where 1 missing Sham patient is assumed to be a failure throughout the above one-failure-at-a-time imputation for sTMS patients. It will then be repeated but where 2 missing Sham patients are assumed to be a failure throughout the imputations for sTMS patients, etc.

G. Subgroup Analyses

The robustness of the primary efficacy outcome will be confirmed via subgroup analyses. Subgroups of subjects will be identified according to their investigative center, gender, age (<median, ≥median), education, concomitant medication, ATHQ history and duration of illness (<median, ≥median). The Fisher's Exact test on the primary endpoint analysis will be repeated within the subgroups of subjects defined by levels of the variable (e.g. adults<median age, adults ≥ median age). The purpose of this subgroup analysis is not to find significant treatment effect within subgroups, but rather to assess consistency of treatment effect across subgroups. This will be performed on the ITT population with available data and imputed data, and on the PP analysis population.

H. Secondary Analyses for Label Considerations

The following analyses will be conducted for the ITT (available and imputed data) and PP populations.

The following secondary measures are to be considered for labeling. They will be analyzed using the gatekeeper (sequential) strategy to control for multiple comparisons by pre-specifying the order of testing (Westfall and Krishen, 2001). If the primary analysis is significant, these outcomes will be tested in the order listed below, each at the 0.05 alpha level. Testing for label consideration will cease when the first outcome results in a two-sided p-value greater than 0.05.

- Change from Baseline on the HAM-D[17] score at Week 6;
- Change from Baseline on the MADRS score at Week 6;
- Clinical response, defined by 50% or greater reduction on the MADRS total symptom score at Week 6.

Treatments will be compared on the mean change from baseline HAM-D[17] and MADRS scores at Week 6 using analysis of covariance adjusting for the baseline value of the given endpoint. Treatment group comparisons on the proportion of subjects who achieve MADRS clinical response at Week 6 will be carried out using Fisher's Exact test. Tests of hypotheses will be two-sided.

Assessment of investigator center-by-treatment interaction will be carried out using Breslow-Day test for the MADRS clinical response at Week 6 and using analysis of covariance with the independent variables of treatment group, investigative center, and the treatment-by-investigative interaction for change from baseline in HAM-D[17] and MADRS at Week 6. Again, a non-significant interaction (using a 0.15 level of significance) or a significant interaction deemed only quantitative in nature supports the pooling of results across centers for the analyses on these secondary endpoints.

Multiple imputation for the ITT population for each endpoint will be carried out using linear regression in a similar manner as for the primary endpoint. For MADRS clinical response, the overall significance of the difference between treatment groups across the imputed datasets will be carried out in the same manner as for the primary efficacy endpoint. For the mean change from baseline at Week 6 for each of HAM-D[17] and MADRS, analysis of covariance comparing treatments adjusting for the baseline value of the endpoint will be carried on each imputed dataset with the results being combined across imputed datasets using PROC MIANALYZE to obtain one overall assessment of treatment difference on the mean endpoint, adjusting for baseline.

For the six-week follow-up open-label phase, double-blind treatments will be compared on the above endpoints at Week 12 in a similar manner as described above. Here, baseline will be defined as the last post-washout assessment measured prior to the beginning of the double-blind treatment phase for patients randomized to sTMS in the double-blind phase; baseline will be defined as the last double-blind assessment measured prior to the beginning of the open-label phase for patients randomized to Sham in the double-blind phase. There will be no assessment of interactions with investigative site, nor will there be imputation of missing data.

I. Exploratory Analyses

Exploratory efficacy variables are listed below. Analyses comparing treatments on these variables will be carried out on the ITT and PP populations; there will be no imputation of missing data, and there will be no assessment of center-by-treatment interactions.

- Change from baseline in HAM-D[17] score at each time point;
- Change from baseline in MADRS score at each time point;
- Remission at each visit, where remission is defined by a total score of ≤ 10 on the MADRS;
- Change from Baseline on the HAM-D[17] subscale scores (Maier, anxiety, retardation and sleep) at each time point;

- Change from Baseline on the HAM-D[24] and HAM-D[28] at each time point;
- Clinical response (as defined above for HAM-D[17]) for HAM-D[24] and HAM-D[28] at each time point;
- Change from Baseline on the IDS-SR30 scale at each time point.

Change from Baseline on the HAM-D[17], HAM-D[24] and HAM-D[28] scores, IDS-SR30 score, and MADRS score will be analyzed using analysis of covariance at each visit adjusting for the baseline value of the endpoint. Proportion of subjects with remission and with clinical response will be analyzed using Fisher's Exact test at each visit.

J. Safety Analysis

The assessment of safety will be based mainly on the frequency of adverse events recorded. Other safety data (e.g., vital signs and ECG) will not be collected systematically but will be considered as appropriate. All safety analyses will be performed on the Safety Population as defined above. Safety summaries will be performed for each of the active and Sham treatments.

Adverse events will be summarized by presenting, for each treatment group, the number and percentage of subjects having any treatment emergent adverse event overall and by each MedDRA preferred term and by each MedDRA system organ class. A patient experiencing more than one adverse event within a given system organ class and preferred term will be counted once within that system organ class and preferred term. Adverse events will be presented separately for the 6-week phase, for the follow-up open-label phase, and for the entire study. A treatment emergent adverse event in the double-blind phase is defined as an event starting or worsening in severity at or after the start of double-blind treatment; a treatment emergent adverse event in the open-label phase is defined as an event starting or worsening in severity at or after the start of open-label treatment. Note that the adverse event frequencies in the open-label phase will be presented by the treatment the patient received in the double-blind phase, as well as for all subjects entering the open-label phase as one group.

The above will be repeated for treatment emergent serious adverse events

Adverse events will be summarized in a similar manner by severity and relationship to study device. Patients with more than one adverse event within a given system organ class and preferred term will be categorized into the maximum category (e.g., maximum severity) experienced for that system organ class and preferred term.

Other safety information collected during this study is:

1. **Vital Signs and Weight:** Baseline, Week 6 Evaluation Visit

   Subjects' heart rate and blood pressure will be collected at Baseline and at the Week 6 Visit (or early termination visit if the subject discontinued prior to Week 6). Descriptive statistics of each variable will be presented at teach visit for the safety population by double-blind treatment group. Descriptive statistics of the change from baseline will also be presented by treatment group.

2. Urine Toxicology Panels: Screening

A urine toxicology dipstick screen will be performed at screening visit in order to rule out use of unprescribed psychoactive medications. All subjects with a positive test must be withdrawn from the study unless a subsequent negative test is obtained (this may only occur once prior to randomization). These subjects will be informed that urine samples may be obtained on additional occasions during the study, but will not be told how many samples nor about the time at which the samples will be collected. Sites may conduct an optional urine toxicology screen if the Investigator feels it is indicated at any time during a subject's study participation. Urine Toxicology Screening will be provided in data listings.

K. Other Assessments

1. Concomitant Medications are collected at Baseline and throughout the study through Week 12. All medications prescribed and/or taken by the subject will be collected. Any changes to these medications may be updated at any visit. Concomitant medications will be provided in data listings.

2. Verification of Blinding Questionnaire:

A treatment blinding questionnaire will be administered to treatment administrator ("treaters"), raters and patients at Treatment Day 5 (end of Week 1) and Treatment Day 30 (end of Week 6, or i.e. end of the double-blind phase). This brief questionnaire will ask participants to (a) provide a best guess of the treatment to which the patient was randomized (sTMS or Sham), (b) a reason for the guess (improvement or lack thereof; how the machine operated; something the subject, rater, or treatment administrator did or said; no reason); and (c) confidence level of the guess (Extremely; Considerably; Moderately; Slightly; Not at All).

From the "best guess" and its "confidence level", a blinding index will be calculated at each of Treatment Days 5 and 30, similar to the blinding index discussed in Bang et al. (2004)[1] and Kolahi et al. (2009)[2]. Specifically, for each treater, rater, and patient at each visit, a blinding index will be created as follows:

- If a treater/rater/patient correctly guessed the patient's randomized treatment group AND was extremely or considerably confident in the guess, then **the blinding index will be set to 1** for that treater/rater/patient

[1] Bang H, Ni L, Davis C E. Assessment of blinding in clinical trials. Controlled Clinical Trials. 2004; 25: 143-156.
[2] Kolahi J, Bang H, and Park J. Towards a proposal for assessment of blinding success in clinical trials: up-to-date review. Community Dentistry and Oral Epidemiology. 2009; 37: 477-484.

- If a treater/rater/patient correctly guessed the patient's randomized treatment group AND was moderately or slightly confident in the guess, then **the blinding index will be set to 0.5** for that treater/rater/patient
- If treater/rater/patient was "not at all confident" in their guess, then **the blinding index will be set to 0** for that treater/rater/patient regardless of the treatment group to which the patient was randomized.
- If a treater/rater/patient incorrectly guessed the patient's randomized treatment group AND was moderately or slightly confident in the guess, then **the blinding index will be set to -0.5** for that treater/rater/patient
- If a treater/rater/patient incorrectly guessed the patient's randomized treatment group AND was extremely or considerably confident in the guess, then **the blinding index will be set to -1** for that treater/rater/patient

Descriptive statistics (sample size, mean, standard deviation) of the treater, rater, and patient blinding index are presented at each visit (Treatment Day 5 and Treatment Day 30) for each treatment group and both treatments combined. A mean close to 0 indicates random blinding (or random guessing); a mean closer to 1 than 0 may indicate failure in blinding (majority of "guesses" were correct); a mean closer to -1 than 0 indicates mainly incorrect guesses were made.

L. Study Treatment Administered

A data listing of the study treatment administered will be provided. The cumulative number and percentage of subjects exposed to study treatment by Week, and descriptive statistics of the number of treatments and percentage of planned treatments overall and by each Week, will be provided. This will be presented by double-blind treatment group for each of the double-blind and open-label phases.

The number and percentage of subjects receiving concomitant psychotropic medications and significant non-drug therapy will be summarized by treatment group for the double-blind phase and for the open-label phase.

M. Data Listings

Listings of all data collected on the case report forms will be provided as an appendix to the clinical study report.