Study Title:

Improving Outcomes for Patients With Sleep-Disordered Breathing and Insufficient Sleep (RESTORE)

ClinicalTrials.gov Identifier (NCT):

NCT04279834

Document Date:

August 15th, 2023

UPDATED STATISTICAL ANALYSIS PLAN: July 2023

Preliminary Data Analyses: First, we will examine the distribution of each variable, check for out-of-range values, outliers and abnormal values using graphical methods (e.g., boxplots and histograms) and descriptive summaries to ensure that all values are within expected ranges. We will verify that the distributions of measures meet the assumptions of the statistical tests to be used, applying a formal test such as the Shapiro-Wilk's test.[1] Transformations will be used when distributional assumptions are not fulfilled for inferential tests. Tests will be conducted to identify potential relationship of baseline demographic and clinical variables to our dependent variables and to see whether they are balanced between groups. If a baseline variable is not balanced between groups and is correlated with the dependent variable (r>.30), we will include this variable as a covariate in subsequent analyses. Since we propose to test two primary hypotheses for two specific aims, the Hochberg step-up multiplicity adjustment will be used with a two-tailed family wise alpha-level of 0.05.[2] All other tests described below will each have a two-tailed alpha-level of 0.05. Outcome analyses will be based on the principle of intention to treat (ITT).[3]

Revised Primary Aims and Hypothesis

Specific Aim 1: To demonstrate the effectiveness of STE compared to PAP in **reducing daytime sleepiness and fatigue** among patients with insufficient sleep and mild-to-moderate SDB.
Primary Hypothesis: STE will be superior to PAP therapy in reducing daytime sleepiness and fatigue.

Secondary Aim 2: To demonstrate the effectiveness of STE compared to PAP in **increasing objective alertness** among patients with insufficient sleep and mild-to-moderate SDB.
Secondary Hypothesis: STE will be superior to PAP therapy in increasing objective alertness, as measured by number of psychomotor vigilance testing (PVT) lapses.

Analysis of Primary Hypotheses (superiority of STE compared to PAP in reducing sleepiness): We hypothesize that STE will be superior to PAP in reducing daytime sleepiness (measured by EPWORTH sleepiness scale in PH2). The primary hypothesis will be tested using mixed-effects linear regression analyses.[75] Model will include up to 3 repeated assessments (months 0, 1 and 3) as the dependent variable. Each model will include a random intercept and slope and fixed effects for treatments, and time, as well as the stratification variables: gender, age etc. A likelihood ratio test will examine the incremental contribution of the treatment by time interaction, which represents the interaction of interest for primary hypothesis, testing for a differential treatment effect over time. Since only one primary hypothesis is used no alpha adjustment is necessary and a two-tailed alpha-level of 0.05 is used. In addition, likelihood ratio tests will be used to compare the model fit with that having a first-order autoregressive (AR1) covariance structure (see Hedeker and Gibbons).[77] The study is designed and powered to test the two primary hypotheses. However, we expect that its data will yield valuable information for hypotheses generation, which can guide the design of subsequent confirmatory studies.

Effect size	ICC		
	0.40	**0.50**	**0.60**
0.45	0.85	0.78	0.68
0.50	0.91	0.84	0.75
0.55	0.95	0.92	0.79
0.60	0.99	0.96	0.88

Table A: Power table, Effect size is expressed as the standardized between group difference at the end of 3-month.

Power Analysis/Required Sample Size

Statistical power analyses examined the sample size requirements to detect clinically meaningful group differences on each of the two primary outcomes over time (i.e., the slope). This represents the differential course for the two intervention cells over time. Statistical power was estimated in a simulation study using SAS PROC MIXED. The sample size for the study was based on the N needed to detect the two-way interaction at two sided alpha=0.05. The main outcome time point will be the 3-month follow-up. Sample size calculation was based on ESS as the main outcome, allowing for 10% attrition by 3-month follow-up. When assessing power, Cohen's d is used to describe the size of an effect, which can be detected for the specified sample size and power. For tests comparing two groups, Cohen's d=of .20 is described as small, .50 as medium, and .80 as a large effect.[77] Based on the results, the protocol proposes recruitment of 110 participants (55/arm).

Table A shows the statistical power to detect standardized interaction effects of various magnitudes (0.45 to 0.60) for the proposed N's and a two-tailed alpha-level of 0.025. (The magnitude of interactions is expressed as standardized differences at 3 month that would be seen with differential slopes). The intraclass correlation coefficient varied (ICC: .40, .50, .60); and in an effort to account for attrition, the total number of observations per participant was increased by 10%. The proposed final sample size of N=122 (61/arm) will provide sufficient statistical power to detect differences in slopes that results for all effect size in Table A, with $0.45 \geq$ ICC.

For secondary hypothesis that STE will be superior to PAP in increasing objective alertness (measured by PVT lapses in PH1), a similar mixed effect model as that of primary hypothesis model will be used. Albeit secondary hypothesis is not powered.

Exploratory analyses within treatment groups:

PAP: To evaluate the potential mechanisms through which PAP may improve outcomes, we will examine the degree of PAP use as a predictor of change in each of the main outcomes from baseline to post-treatment. Our exploratory hypothesis will be greater improvement in outcome will be correlated with greater PAP usage. In addition, we will use cluster analysis to discover additional subgroups based on clinical and demographic variables and then study the relationship of these subgroups with outcomes. This exploratory analysis may result in additional discovery of underlying factors and mechanisms to improve upon intervention itself or its usage in the future. The exploratory analysis is not powered, but may result in preliminary findings which need to be established in future via appropriately powered studies.

STE: To evaluate the potential mechanisms through which STE may improve outcomes, we will examine the degree of increase in total sleep time (based on both sleep diaries and wrist actigraphy) as predictors of change in each of the main outcomes from baseline to post-treatment. Similar to PAP, we will perform exploratory analysis with STE and outcomes.

Attrition and missing data

Attrition introduces bias and reduces power, precision and generalizability.[79] In keeping with ITT principles,[69] we distinguish between intervention termination and study termination. Accordingly, we will make every effort to continue assessments for the entire course of randomized treatment, even among those who do not continue with randomized treatment.[80] We anticipate attrition to be ≤10%. The proposed mixed-effects models will incorporate all available data, even from subjects who do not complete the trial. Mixed-effects models yield valid inferences assuming ignorable attrition.[81] To evaluate missing data assumptions and determine the best analytic strategy, the methods recommended by Schafer and Graham[82] will be used. First, for a small proportion of missing data and evidence that they are Missing at Random (MAR), the available data will be analyzed using maximum likelihood-based estimation procedures. Second, for a non-trivial amount of missing data and evidence that they are MAR, multiple imputation for repeated measurements will be used to generate complete data. Third, for a non-random missing data mechanism (i.e., MNAR), a pattern mixture model will be used to examine response to treatment among participants with various dropout patterns and will be implemented using a longitudinal strategy.[83] In addition, a sensitivity analysis will be carried out for identifiable variables at baseline

that have a correlation of >0.30 with attrition status. Estimates of the treatment effect from the models described above will be compared with models that also include the main effects of dropout pattern. Estimates of the treatment effect from the models described above will be compared with models that also include the main effects of dropout pattern. This approach will be used to examine the sensitivity of the results to attrition.