Speech Entrainment for Aphaisa Recovery (SpARc)
Statistical Analysis Plan (SAP)

Version 5
July 2024

Prepared by

Jordan J. Elm, PhD
Christy Cassarly, PhD

Statistical and Data Management Center (SDMC)
Data Coordination Unit
Department of Public Health Science
Medical University of South Carolina
Charleston, SC 29425

Summary of Changes

Version 5; July 2024

- Added Title Page and Summary of Changes.
- Section 7. Clarified that the analyses will adjust for the *centrally scored* baseline WAB-AQ. Updated primary analysis model to include baseline VPM as a covariate to increase efficiency.
- Section 8. Specified variables to use to define nearest neighbors for imputation of missing data.
- Section 11. Updated secondary analysis plan to mixed model approach to account for different timepoints.
- Section 15. Updated exploratory analysis plan to mixed model approach to account for different timepoints.

Version 4; September 2023

- Section 2. Updated "time from stroke onset" to "age at stroke onset" in the list of baseline variables to be compared by treatment group.
- Section 12. Added clearer language describing that the safety monitoring plan includes testing for worsening only of the primary outcome.

Version 3; June 2022

- Section 6. Changed "permuted block design" to "minimization" as minimization is more similar to the hierarchical restrictive covariate adaptive randomization procedure used in SpARc. Clarified that the randomization approach used in SpARc provides a *higher* level of allocation randomness than minimization.
- Section 7. Clarified that the primary outcome is calculated as the change in *average VPM* as VPM is measured on two tasks at each visit. Added the following sentence: "Empirical robust sandwich estimators will be used to account for the randomization procedure in order to appropriately estimate the marginal treatment effect." Added the EMPIRICAL keyword to the PROC MIXED code. Corrected PROC MIXED syntax to include main effects for arm and visit.
- Section 11. Added individual components of the primary outcome (VPM on Narrative task and Procedural task) as secondary endpoints.
- Section 12. Added new monitoring plan for worsening in the primary outcome.

Version 2; August 2021

- Section 6. Updated the randomization procedure to be used from "asymptotic maximal procedure" to "hierarchical restrictive covariate adaptive randomization procedure" given the small sample size and number of baseline covariates.
- Section 7. Added the following sentence: "As an exploratory analysis, the combined SET duration versus control comparison will be tested at a two-sided alpha of 0.05."
- References were deleted as the only cited reference was to the asymptotic maximal procedure.

1. SYNPOSIS

Speech entrainment for Aphasia Recovery (SpARc) is a prospective, controlled, randomized, assessor blinded, phase II clinical trial (n=80). The primary objective is to estimate the dose (duration of treatment in weeks) of SET with the highest effect size on speech fluency Verbs Per Minute (VPM) as compared to control (no SET). We will enroll 80 patients and randomly assign them 1:1:1:1 to one of the 4 groups (A, B, C, D):

A - SET for 3 weeks (15 days, 1 hour daily, 5 x week)
B - SET for 4.5 weeks (22 days, 1 hour daily, 5 x week)
C - SET for 6 weeks (30 days, 1 hour daily, 5 x week)
D - no SET for 6 weeks (control condition)

2. GENERAL STATISTICAL CONSIDERATIONS

Patient Accountability & Compliance

A flowchart (CONSORT Diagram) will be created to present a summary of participant status. This flowchart will list the number of patients who were randomized to each treatment strategy (A, B, C, D). Then, within each group, it will list the numbers of patients who completed the study, withdrew consent, and lost to follow up.

Treatment Group Comparability

Summary statistics of baseline variables (e.g. demographics, age at stroke onset, aphasia type, WAB-R AQ, all baseline assessments) will be compared by treatment groups. Dichotomous variables will be summarized as number (%). Percentages will be calculated based on the number of participants with available data for that variable. Continuous variables will be summarized by the mean and standard deviation (SD). Ordinal data will be presented as median (IQR).

Preliminary Analysis

For all continuous variables, outliers will be explored. Extreme outliers will be queried to confirm that they are not erroneous before the data is locked for analysis, but outliers will not be removed from the analysis.

General Approach

In general, descriptive statistics will be presented by treatment group without p-values. Dichotomous variables will be summarized as number (%). Percentages will be calculated based on the number of participants with available data for that variable. Continuous variables will be summarized by the mean and standard deviation (SD). Ordinal data will be presented as median (IQR).

3. PRIMARY HYPOTHESES

- Primary Endpoint:
  VPM at 3 months post treatment

We hypothesize that, compared to patients who receive control, patients who receive SET for 3, 4.5, or 6 weeks for aphasia have improved VPM at 3 months and that there is an optimal duration of SET therapy. However, this study was not designed to formally test SET versus control. This study represents the first step in a series of studies which will ultimately test this hypothesis. However, for this study no formal hypothesis tests are planned. The primary analyses will be descriptive statistics of VPM at 3 months and will estimate the effect size of SET versus control adjusting for baseline.

4. SAMPLE SIZE CONSIDERATIONS

Sample Size Considerations for Comparison with Control group
This study is not powered to detect statistically significant differences between the chosen SET dose and the control group. This will be the focus of a later and dedicated comparison study once the best dose has been defined. The sample size was determined by administrative reasons.

Sample Size Considerations for Effect Size of SET versus Control group
For each treatment group versus control we will estimate the standardized mean difference (effect size). The criteria needed to accept any duration is that the effect size is at least 0.36 (a "small" effect size). The rationale for this criteria is that if the pooled standard deviation estimate is large, because of the small sample size, then the effect size may be smaller than expected. We consider a 20% improvement from control to be the MCID. This corresponds to a 1 point improvement in VPM (5x20%=1). Given the MCID=1 and the common SD of change in VPM estimate from pilot data of 2, we expect the effect size to be ½=0.5. However, with n=20 per group, observed SD may be as large as 2.74 (i.e. a one-sided upper 95% confidence interval for SD=2 is 2.74), thus the observed effect size may be only 1/2.74=0.36.

5. POPULATIONS FOR ANALYSES

The analyses will be analyzed under the intent-to-treat principle (ITT). Under this principle, the evaluable sample will include all participants who are randomized.

6. RANDOMIZATION

A "Real-Time" randomization procedure will be implemented via the trial website on the WebDCU ™ System (Data Coordination Unit, located at MUSC). Using the WebDCU ™, the local site's study staff will enter baseline and eligibility information for each subject prior to enrollment. If a subject's eligibility status is confirmed, WebDCU™ will make the treatment assignment based on the current status of treatment group distribution at each site, age group (≤ 70, >70 years), and aphasia severity (mild/moderate, severe), as well as the overall balance of treatment assignments. A hierarchical restrictive covariate adaptive randomization procedure will be used to ensure that treatment groups are balanced within site and baseline covariates. This approach is similar to the commonly used minimization but reduces the proportion of deterministic assignments and provides a higher level of allocation randomness.

7. ANALYSIS OF THE PRIMARY ENDPOINT

The primary outcome is VPM which is a continuous value greater than 0. It is collected at baseline and after 3 months post treatment. The primary analysis will calculate the change in average VPM (from the Narrative Story Telling Task and Procedural Story Telling Task) from baseline. The statistical model will be a repeated measures model (SAS® MIXED procedure with REPEATED subcommand). The model will include the following fixed effects: categorical visit (1 week, 3, and 6 months post treatment period) by treatment group (class variable) interaction, site, baseline aphasia severity score (centrally scored), baseline VPM and age. The estimated difference (95% confidence intervals) in change between each SET duration group and control group will be reported. The standardized mean difference for each SET duration versus control will be used to select the best dose of SET. Since this study is not powered for direct comparison with control, no p-values will be reported.

Empirical robust sandwich estimators will be used to account for the randomization procedure in order to appropriately estimate the marginal treatment effect. The unstructured covariance matrix for repeated observations within subjects will be used. In case the model will not converge, the maximum-likelihood (ML) estimation method will be used instead of the default restricted ML (REML). If the model still does not converge, a simpler covariance structures with less parameters will be used, according to the following order: heterogeneous autoregressive(1) (ARH(1)), heterogeneous compound symmetry (CSH), autoregressive(1) (AR(1)), and compound symmetry (CS). The estimated means at the 3 month post visit of the change from baseline in VPM will be compared for the SET 3 week group versus control, SET 4.5 week versus control, SET 6 week versus control, and the combined SET durations versus control. The SAS code for this analysis is as follows:

```
PROC MIXED EMPIRICAL;
  CLASS SUBJID ARM VISIT CENTER;
  MODEL CHANGE=ARM|VISIT CENTER BASE_WABAQ BASE_VPM AGE;
  REPEATED VISITNUM / TYPE=UN SUBJECT=SUBJID R RCORR;
    LSMEANS ARM*VISIT/ cl;
RUN;
```

The adjusted means and 95% confidence intervals for each follow-up assessment visit and treatment arm will be plotted. The primary time point of interest is the 3 months post, but the 1 week and 6 month assessments will be explored to assess whether benefits are sustained over time. As an exploratory analysis, the combined SET duration versus control comparison will be tested at a two-sided alpha of 0.05.

8. MISSING DATA

Under the ITT principle, all patients who are randomized are included in the analysis. Therefore, missing data, especially in the primary outcome measure, can be problematic. All patients randomized will be included in the primary analysis regardless of whether or not they dropped out or discontinued treatment. For the primary analysis, a repeated measures linear mixed model will be fit; this is considered an implicit imputation approach if at least 1 post-baseline assessment is available. However, if baseline is the only assessment available (or if baseline is missing but follow-up available), then a nearest neighbor approach will be used to impute the missing values within aphasia type using age (≤, >70 years), and WAB-R AQ to define “nearest neighbors”.

9. PRE-SPECIFIED RULE TO SELECT BEST SET DURATION

1- Highest standardized mean difference (effect size) in VPM from baseline to three months post treatment
2- Adequate tolerability: group average participation in at least 80% of the treatment sessions

We will select the duration associated with the highest group average VPM at 3 months post treatment, but also had adequate tolerability.

10. TOLERABILITY

For each participant, the participation rate will be calculated as the number of completed treatment sessions divided by the expected number of treatment sessions (with allowances for out of window sessions). The average participation rate (95% confidence intervals) will be calculated for each treatment group. Adequate tolerability of a SET duration would be having a group average participation in at least 80% of the treatment sessions.

11. ANALYSIS OF THE SECONDARY ENDPOINTS

- Secondary Endpoint(s):
    - Quality of life assessment - SAQOL-39
    - Individual components of the primary outcome:
        - VPM on Narrative Story Telling Task
        - VPM on Procedural Story Telling Task

The secondary endpoints will be analyzed similarly to the primary endpoint. The same approach will be used to define the analysis sample and handling of missing data. The adjusted mean change from baseline of each SET group will be compared to control with 95% confidence intervals. The statistical model will be a repeated measures model (SAS® MIXED procedure with REPEATED subcommand) including the following fixed effects: categorical visit (1 week, 3, and 6 months post treatment period) by treatment group (class variable) interaction, site, baseline aphasia severity score (centrally scored), baseline value of the secondary endpoint, and age.

In addition, actual values and changes from baseline to each visit in secondary endpoints will be summarized using descriptive statistics. Since this study is not powered for direct comparison with control, no p-values will be reported.

12. SAFETY ANALYSES

<u>Monitoring Plan for Worsening in the Primary Outcome</u>

If SET leads to substantially worse outcome than the no SET arm, early stopping may be considered. The unblinded statistician will fit the primary analysis model semi-annually in order to compare VPM between the combined SET durations versus No SET to detect worsening only (one-sided α=0.025). As part of this analysis, outliers will be explored to ensure that the results are not driven by a specific

individual. Should the null hypothesis be rejected, an investigation of dose response among the SET arms based on qualitative evaluation (i.e. not statistical) when ordered in terms of dose level will occur. Should a dose response worsening be detected, the unblinded statistician will bring this to the attention of the DSMB and investigators.

Safety Monitoring

The unblinded statistician at the DCU will produce semi-annual DSMB reports and at the request of the DSMB. The Closed Session DSMB reports will show tables by partially blinded treatment group. All serious adverse events will be summarized by "preferred term" and associated system-organ class according to the MedDRA dictionary and by treatment group in terms of frequency of the event and number of subjects having the event. For the DSMB report, severity and relatedness to the study drug will also be shown by treatment group.

At the end of the trial, the cumulative incidences of the specific SAEs related to treatment, as well as all SAEs, will be compared across groups.

13. INTERIM ANALYSES

Given the small sample size, the study does not contain any planned interim analyses to stop the trial early for success or futility.

14. SUB-GROUP ANALYSIS

Recruitment and retention of females and minorities will be monitored by the DSMB and will be provided in the Final Report. Although we do not anticipate differential treatment effects based on sex, race, or ethnicity, our analyses will explore clinically important differences due to sex/race/ethnicity. A clinically important interaction of the treatment effect by sex, race, or ethnicity, regardless of the statistical significance, will be reported to the scientific community in the primary paper.

15. EXPLORATORY ANALYSES

The sensitivity of the treatment effect estimated in the primary analysis may be evaluated with exploratory analyses. The actual values and unadjusted changes from baseline to each visit in primary endpoint will be summarized using descriptive statistics. An exploratory analysis of the primary outcome (VPM) may be conducted as described above for the primary analysis, but including the treatment group assignment as a continuous variable (0, 3, 4.5, 6 weeks). The imputation method of multiple imputation may be explored as a sensitivity analysis.

- Tertiary/Exploratory Endpoints:
  - Communicative Effectiveness Index (CETI)
  - Apraxia of Speech Rating Scale (ASRS)
  - WAB-R

The adjusted mean change from baseline of each SET group will be compared to control with 95% confidence intervals. The statistical model will be a repeated measures model (SAS® MIXED procedure with REPEATED subcommand) including the following fixed effects: categorical visit (1 week, 3, and 6 months post treatment period) by treatment group (class variable) interaction, site, baseline aphasia severity score (centrally scored), baseline value of the exploratory endpoint, and age. In addition, actual values and changes from baseline to each visit in exploratory endpoints will be summarized using descriptive statistics. Since this study is not powered for direct comparison with control, no p-values will be reported.

16. EVALUATION OF POTENTIAL CONFOUNDERS

As described above, all baseline and treatment video-taped discourse samples will be scored centrally at the Aphasia Lab at USC using AphasiaBank. Special care will be taken to evaluate the influence of potential confounders such as apraxia of speech (rated on the ASRS), speech repetition scores (WAB-R, PRT), single word (WAB-R) and sentence level comprehension[65]. As a sensitivity analysis, the primary analysis will be repeated while adjusting for these potential confounders to define determinants of the primary outcome. Since SET has not been systematically studied as a potential clinical treatment, the goal of this project is to determine if SET is associated with a robust effect size regardless of the underlying source of impairment, but it is very likely that cognitive and linguistic factors that vary across persons with non-fluent aphasia may influence SET outcome to different degrees. We will assess these variations and the overall benefit of SET by systematically assessing the effect size of SET and examining individual characteristics that relate to treatment response. This will be accomplished by assessing for interaction effects of potential confounders and treatment group.