STATISTICAL ANALYSIS PLAN

PHASE II STUDY
FUTILITY STUDY OF tDCS IN STROKE PATIENTS WITH APHASIA

Prepared by:
Jordan J. Elm, Ph.D.
Medical University of South Carolina
Division of Biostatistics & Epidemiology
135 Cannon St., Suite 303
Charleston, SC 29425

August 29, 2012
Version 1

TABLE OF CONTENTS

1. LIST OF ACRONYMS

- A- tDCS Anodal transcranial direct current stimulation
- ABA Apraxia Battery for Adults
- AE Adverse Event
- AQ Aphasia Quotent
- ASRS Apraxia of Speech Rating Scale
- BNT Boston Naming Test
- DCU Data Coordination Unit (at MUSC)
- DSMB Data Safety Monitoring Board
- ITT Intent-to-Treat
- PNT Philadelphia Naming Test
- PPTT Pyramids and Palm Trees Test
- S- tDCS Sham transcranial direct current stimulation
- SAE Serious Adverse Event
- SAP Statistical Analysis Plan
- SLP Speech Language Pathologist
- tDCS transcranial Direct Current Simulation
- trt Treatment Group
- tx Treatment
- WAB-R Western Aphasia Battery- Revised
- WAIS-III Wechsler Adult Intelligence Scale III Score
- WebDCU™ Web-based Clinical Trial Management System

2. STATISTICAL ANALYSIS PLAN AND STATISTICAL REPORTS

This document provides the details of the statistical analyses planned for the trial. Prior to locking the database and breaking the code, this Statistical Analysis Plan (SAP) will be updated. The SAP will define all "pre-specified, pre-planned analyses."

The DCU Biostatistics Core under the direction of Jordan Elm, PhD will generate closed and open DSMB reports semi-annually or more frequently, as determined by the DSMB. Each DSMB report provides cumulative summary statistics on enrollment; subject status in the study (e.g., number completed study, drop outs, etc); baseline characteristics; safety data, including AEs and SAEs by MEDdra code, severity, and relatedness to the study intervention; and data management/quality information (e.g., timeliness and completeness of data entry by the clinical centers via the WebDCU™; number of data queries generated and resolved). The statistics for the closed DSMB Reports are provided by intervention group displayed as A or B. The open report contains aggregated statistics only, i.e., not by intervention group.

3. SYNOPSIS OF THE STUDY

A Randomized, Multi-Center, Double-Blind, Pilot Study of tDCS for stroke patients with aphasia

4. STUDY DESIGN AND OBJECTIVES

At total of 74 subjects from 2 sites will be randomized 1:1 to tDCS (n=37) or sham (n=37). This study design will allow us to assess whether it is futile to proceed with further study of tDCS in a larger Phase III trial.

The primary objective of the study is to compare tDCS versus sham-control in stroke patients with aphasia in order to assess whether it is futile to proceed with further study of tDCS.

Planned enrollment duration: approximately 4 years (0.77 patients /site/month)

Duration of follow-up in Primary Study Phase: 3 weeks
Duration of follow-up in Secondary Study Phase: 7 weeks from baseline
Duration of follow-up in Post-testing phase: 27 months from baseline

5. DEFINITION OF TARGET POPULATION

5.1 Target Population

Stroke patients with aphasia receiving a computerized treatment for anomia.

5.2 Intent-to-Treat Sample

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

5.3 Safety Analysis Sample

All randomized participants who receive at least one session of the intervention are included in the safety analyses, regardless of the duration of intervention administered. Patients will be classified according to the treatment group that was actually received.

5.4 Per Protocol Sample

A secondary analysis of the primary outcome will be done, using the subset of patients defined by the Per Protocol Sample. This will be defined as having completed at least 80% of the expected study sessions on schedule. Patients will the classified according to the treatment group that was actually received. Patients for whom there were major protocol deviations will be excluded.

6. RANDOMIZATION AND BLINDING

The study is to be conducted in a double-blind manner. The subjects, the site investigators, and the USC and MUSC clinical staff involved in this study will not know the intervention assignment. Select members of the Data Coordination Unit (DCU) will be partially blinded, i.e., they will know the intervention group assignment as A or B, but not whether the patient receives active tDCS or sham. The study statistician and the DSMB will have a sealed envelope with the intervention group identifiers. This envelope would only be opened if the study statistician is directed to open it by the DSMB or at the end of the futility study.

The randomization will take place centrally via the trial Website (WebDCU). Subjects will be randomized 1:1 (A-tDCS: S-tDCS). The objective of subject randomization is to protect the randomness of treatment allocation and to prevent serious imbalances in site, baseline age, aphasia type, and severity (as measured by the WAB-R). The block urn method (Zhao et al. 2011) and the minimal sufficient balancing method (Zhao et al. 2012) will be used. The computer program developed at the DCU makes the treatment assignment based on the current status of treatment group distribution within each stratum as well as overall balance of treatment assignment. The randomization scheme will never be deterministic.

A “Real-Time” randomization procedure is implemented via the Trial Website on the WebDCU™ System where the clinical center staff enters the basic baseline (e.g., age, aphasia type, and AQ severity) and eligibility information of a subject prior to enrollment. If the subject’s eligibility status is confirmed, the computer program on the WebDCU™ server will evaluate the treatment arm distribution and generate a treatment assignment based on the randomization scheme. The clinician (speech-language pathologist – SLP) enrolling the patient will not see the treatment assignment, only a numeric patient number.

7. BASELINE INTERVENTION GROUP COMPARABILITY

During the course of the study as a part of the DSMB reports, summary statistics for the following baseline variables will be computed and compared between intervention groups. At study conclusion, the statistical tests for comparison will be two-sample *t*-test or Wilcoxon rank sum test for continuous scale variables and chi-square or exact test for categorical variables.

Demographic:
- Mean age
- Gender
- Mean years of education

Clinical characteristics:
- WAB-R summary scores
- BNT Score Summary
- ABA
- PPTT
- PNT
- Cinderella Story
- ASRS
- WAIS-III

8. PRIMARY ANALYSIS

8.1 Primary Outcome

The primary outcome will be the change in the number of correctly names items on the PNT (plus a portion [N=80] of the trained items) (where change from pre-treatment to immediate post-testing). To assess change in naming ability, the primary outcome in this study will be administered twice (and averaged to reduce variability) on two consecutive days immediately before treatment starts and twice after treatment is completed. The change will be computed as the difference in the number of correctly named items between the average of the two pre-treatment assessments to the average of the two post-treatment sessions. The PNT is a computer based assessment of naming in persons with aphasia and includes 175 pictures representing mid- and high-frequency nouns from a word frequency list compiled by Francis and Kucera (1982).

8.2 Statistical Hypotheses

The primary hypothesis is:

$$\mathbf{H_0:}\ \mu_T - \mu_S \geq \delta$$
$$\mathbf{vs}$$
$$\mathbf{H_A:}\ \mu_T - \mu_S < \delta$$

where μ_T is the expected change (pre-treatment and immediate post-testing) in the number of correctly named items for the treatment arm, μ_S is the expected change in the sham control group, and δ is the minimum improvement sufficient to warrant further evaluation of tDCS treatment. We define that minimum improvement to be 1.5 points more than the observed change in the sham control group. If we reject the null hypothesis, then tDCS is considered clearly ineffective (in improving anomia in stroke patients) and will not be considered for further study. If we fail to reject the null hypothesis, we would consider undertaking a Phase III study of tDCS.

8.3 Sample Size Determination

In preliminary studies, the mean difference in change from baseline in the active and sham groups in naming accuracy was ~2.5 points (Baker, Rorden, & Fridriksson, 2010, Fridriksson et al., submitted). In one study, the S-tDCS (sham) group improved ~1.2 point from baseline to 1 week (Baker, Rorden, &

Fridriksson, 2010). In another study, the S-tDCS group improved ~3.9 points from baseline to 1 week (Fridriksson et al., submitted). For power calculations, we assumed the mean change from baseline to immediate (1 week) post-testing for the A-tDCS group under the null hypothesis of non-futility is equal to $\mu_s+\delta=4+1.5=5.5$. When the sample size is 33 per group, a two sample t-test with a 0.10 one-sided significance level will have 85% power to reject the null hypothesis that the A-tDCS treatment is 1.5 points better than S-tDCS (sham) and declare futility when the A-tDCS treatment comes from a distribution with mean change of 4 (assuming the pooled SD is 2.6). As with most clinical studies, a certain amount of drop-outs (including subject withdrawal or lost-to-follow-up) can be expected. In preliminary studies less than 5% of visits were missing. Conservatively assuming the drop-out rate to be no more than 5% over 5 weeks, the required sample size was inflated from 33 to 37 per intervention arm to account for the effect of the drop-outs in the intent-to-treat analysis using an inflation factor (Friedman, et al, 1985).

8.4 Multiplicity

Since this is a phase II study, the false positive error rate has been relaxed (Schoenfeld 1980). For a futility design the type I and type II error rates are reversed as compared to a traditional, superiority hypothesis. For the primary analysis, the probability of incorrectly declaring a drug futile is 0.1 (or 10%, type I error is the false negative rate). Given that the active tDCS comes from a distribution with a mean change of the control group (e.g. the treatment is like the sham change), the probability of incorrectly moving to a Phase III trial is 0.15 (or 15%, type II error is the false positive rate).

For secondary outcomes and safety analyses, no adjustment of Type I error probability will be considered, since they will be treated as exploratory.

8.5 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. For the primary futility analysis we will impute missing data using multiple imputation [Rubin, 1987] assuming a monotone missing mechanism and missing is at random (MAR). Similar methods will be employed for secondary analyses and secondary outcomes. For safety data (AEs) no data imputation will be done.

8.6 Interim Analysis

No formal interim analyses are planned.

8.7 Adjusting for Covariates

The primary analysis will be adjusted for site and baseline stroke severity (AQ).

9. SECONDARY ANALYSES

9.1 Secondary Efficacy Outcome measures

In addition to the primary outcome, several secondary analyses will be conducted. We will examine changes in types of naming errors (defined by the PNT) by tDCS treatment group. The Cinderella story will be analyzed by comparing lexical diversity (VOCD) for nouns, verbs and adjectives; number and types of errors; length and patterns of pauses by treatment group. Pre and post comparisons with both non-aphasic and aphasic speakers from the AphasiaBank database who share a number of demographic features (e.g., type and severity of aphasia, age, etc) will also be made.

At the end of the study, for the interval scale variables, mean change from baseline to immediate post-testing in secondary outcome measures will be reported by treatment group along with the 95%

confidence intervals. Treatment comparisons will be made with a paired *t*-test. For binary variables, the proportion of subjects immediately post-testing will be reported by treatment group along with the 95% confidence intervals.

9.2 Exploratory Analyses of the Primary Outcome measures

Spearman's correlation coefficients will be computed for pre-post change in anomia versus aphasia type first using all patients and then by tDCS group. This association will be further explored using the longitudinal data in a GLMM model described in section 3.14 (post-testing phase). Since there will be little statistical power to detect an interaction of aphasia type and treatment, we will examine the main effect of aphasia type (adjusting for the other covariates) by tDCS group.

9.3 Post-Testing Phase

The longer follow-up post-testing phase will provide exploratory information on whether the immediate post-testing improvement after 3 weeks of treatment can be sustained. The mean (95% CI) changes from baseline to immediate (within 1 week) post-testing, 4 weeks posting-testing and 6 months post-testing will be reported. Box and whisker plots will be produced to show the distribution of naming accuracy over time by group.

To explore the longitudinal data, a Repeated Measures model (SAS® MIXED procedure with REPEATED sub-command) will be constructed by tDCS group. The model will include the following fixed effects: visit (as a classification variable), aphasia type, severity, and clinical center as independent variables. Various error structures will be examined (e.g., compound symmetry, autoregressive, unstructured), and the final structure will be selected based on the model Akaike Information Criteria. A similar modeling approach may be applied to other secondary outcomes (e.g., number of naming errors).

9.4 Exploratory Analyses of BDNF polymorphism

A two group t-test with a 0.05 two-sided significance level will have 80% power to detect a difference in mean reaction time of 2.6, assuming that the common standard deviation is 2.6, when the sample sizes in the two allele groups are 12 having at least 1 MET allele and 25 without the MET allele, respectively (a total sample size of 37 patients in the A-tDCS group assuming 33% have at least 1 MET allele).

10. SAFETY ANALYSES

All adverse experiences will be summarized in terms of frequency, severity and relatedness to the study intervention using the MedDRA code. All subjects who received tDCS will be included in the safety analysis. At the end of the study, the cumulative incidences of adverse events are compared between the two treatment groups using Fisher's exact test at the two-sided alpha level of 0.05.

The repeated measures of the FACES pain rating scale will be compared by treatment group by fitting a repeated measures proportional odds model.

References

Friedman L, Furberg C, DeMets D. Fundamentals of Clinical Trials. 2nd ed. Littleton, MA: PSG Publishing Company, Inc., 1985.

Pocock SJ, Simon R. (1975) Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial, Biometrics 31:102-115,1975.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys,* New York: John Wiley & Sons, Inc. p.166-167.
Schoenfeld D. Statistical considerations for pilot studies. Int J Radiat Oncol Biol Phys. 1980 Mar;6(3):371-4.

Zhao W, Weng Y., Block urn design - a new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. *Contemp Clin Trials*. 2011 Nov;32(6):953-61. Epub 2011 Aug 22. PMID: 21893215

Zhao W, Hill MD, Palesch Y., Minimal sufficient balance--a new strategy to balance baseline covariates and preserve randomness of treatment allocation. *Stat Methods Med Res*. 2012 Jan 26. [Epub ahead of print], PMID: 22287602