A Priori Data Analysis Plan

RESPONSE Trial: REmediation of SPatial NEglect Trial

Protocol Number: PSC-1004-13

NCT01965951

Document Date: 6/6/18 3:33 PM

1. GENERAL STATISTICAL PLAN

The data analysis plan *a priori* defines a primary intent-to-treat (ITT) population, a set of secondary evaluation populations, a primary outcome measure, a set of secondary outcome measures, a single primary evaluation time point, a secondary evaluation time point, a primary statistical analysis methodology, a criterion for statistical significance, and guidance for interpretation of results.

The **primary ITT population** is defined as all participants who complete a training set-up visit. Note that this includes all randomized participants except those who drop/withdraw post-randomization and pre-setup visit.

The **primary outcome measure**, per determination of an independent external review committee comprised of experts in the diagnosis and evaluation of the trial target population (i.e., individuals with hemispatial neglect following acquired brain injury), is the Posner cueing task; a spatial attention measure that captures the visual *Field Bias* score (average reaction time to Left trials – average reaction time to Right trials), as reported here:

> Rengachary, J., d'Avossa, G., Sapir, A., Shulman, G. L., & Corbetta, M. (2009). Is the posner reaction time test more accurate than clinical tests in detecting left neglect in acute and chronic stroke?. *Archives of physical medicine and rehabilitation*, *90*(12), 2081-2088.

To capture a valid assessment of participants' performance in the Posner cueing task (*Field Bias* score) independent of criterion shifts or speed-accuracy trade-offs [Lange-Malecki and Treue, 2012; Spence et al., 2001; Teufel et al., 2010], we will employ an inverse efficiency scoring approach (IES; Townsend and Ashby, 1983) to account for potential differences in accuracy across Left and Right trial types (e.g., Valid and Invalidly cued trials). Mean reaction times (RTs) will be divided by the proportion of correct trials responded to in each specific condition.

The **secondary outcome measures** will evaluate training exercise-based progress, as well as performance in several measures identified by domain, including: spatial attention, spatial cognition, cognitive performance, functional ability, quality of life, and quality of sleep.

Spatial Cognition

- Greyscales Task (raw scores)
- Posner Cueing Task (raw scores)
- Spatial Working Memory Task (raw scores)

Cognitive Performance

- DKEFS Verbal Fluency Letter Fluency (raw scores)
- DKEFS Verbal Fluency Category Fluency total (raw score)
- DKEFS Verbal Fluency Category Switching total (raw score)
- DKEFS Verbal Fluency Category Switching accuracy total (raw score)
- WAIS-IV Digit Span total (sum of forward, backward, sequencing raw scores)
- WAIS-IV Digit Span Forward total (raw scores)
- WAIS-IV Digit Span Backward total (raw scores)

- WAIS-IV Digit Span Sequencing total (raw scores)
- Continuous Performance Task (raw scores)

Functional Ability

- Catherine Bergego Scale total (raw score)
- Barthel Index total (raw score)

Quality of Life

- SF-12 Physical Composite Score
- SF-12 Mental Composite Score

Quality of Sleep

- Pittsburgh Sleep Quality Index total (sum of component scores)
- Pittsburgh Sleep Quality Index Sleep Efficiency (raw score)

The **primary analysis time point** is V2, the post-training assessment. The **secondary analysis time point** is V3, the follow-up assessment.

The **primary statistical analysis methodology** is a linear mixed-effects model approach. We will first compare treatment and active control groups in the ITT population to determine if any differences in baseline demographic, characterization, outcomes variables, or total program use time remain after the randomization process. Any such factors that show trend level significant differences ($p < 0.1$) will be noted and used as covariates in the linear mixed model analysis. We will examine the data from each outcome measure using a linear mixed model with group and time as fixed factors, site as a random factor, and co-variates as necessary from the baseline analysis. Missing data will be handled with an iterative maximum likelihood procedure to optimally estimate model parameters. The key value for significance will be the group-by-time interaction factor for the model. This modeling will be conducted with a Type I error set at 0.05 for each model.

Secondary, exploratory analysis will test for effects in exercise-based progress as well as by domain (e.g., working memory, attention, processing speed) to determine the unique benefit in each area of attention, cognition and independent function. First, to determine if the treatment effect varies based on participants' clinical severity of neglect (Mesulam total misses) or symptom subtype (e.g., sensory loss vs. intact sensory systems), we will correlate neglect severity (or subtype(s)) with gain scores on composite spatial cognition or executive performance and composite functional performance (as well as individual assessment measures). Additionally, we will measure whether improvements on certain components of the training program (e.g., improvements at inhibiting one's response to target stimuli) are associated with composite attention, executive or functional performance scores (as well as individual assessment measures). This could indicate whether certain patterns of training improvements are related more to improvements in attention, executive or functional performance. These analyses will be evaluated after correcting for multiple comparisons (i.e., Bonferroni).

The **criterion for statistical significance** is $p<0.05$. Results greater than $p = 0.05$ and less than $p = 0.11$ will be described as trends.

To evaluate the effect of the experimental software treatment program (incorporating TAPAT) on spatial attention) we will conduct the analysis based on the pre-training (baseline) and post-training data.

Finding a significance level of $p < 0.05$ on the primary outcome measure will support the statement that the intervention improves spatial attention in the population.

To evaluate the endurance of effects following completion of the experimental software treatment program (incorporating TAPAT) we will use the same modeling approach based on the pre-training assessment (baseline) and end of study (follow-up visit) assessment data using a mixed-effects model parallel to those used to meet the primary and secondary aims. Finding a significance level of $p < 0.05$ on the primary measure will support the hypothesis that the training drives persistent improvement in spatial attention above and beyond an active control.

To identify specific populations of treatment responders and non-responders) we will conduct exploratory data analyses post-hoc to examine predictors of treatment gain and predictors of lack of treatment gain based on baseline participant demographic, cognitive, and functional measures, as well as on learning rate and plateau performance measures derived over the course of the experimental software treatment program (incorporating TAPAT) use. We will initially employ single regression models to identify variables that predict change in cognitive and functional assessments, and then use generalized linear modeling to explore combinations of predictors of treatment gain or lack of gain.

2. CALCULATIONS FROM CRFS

The following calculations must be made from CRF data to create the required data tables. These calculations are performed automatically by scripts that operate on the tables from the EDCS and produce processed data in new tables.

- Posner Cueing Task
    1. Reaction Time Difference for Left - Right trails of *Field Bias* (Left Valid and Invalid RT – Right Valid and Invalid RT)
    2. Reaction Time Difference for Invalid trails or *disengagement* effect ((Left Invalid-Valid RT) – (Right Invalid-Valid RT))
    3. Reaction Times, Averaged Across All Trials (Left Invalid RT, Right Invalid RT, Left Valid RT, Right Valid RT)
    4. Accuracy, Averaged Within and Across All Trials (Left Invalid RT, Right Invalid RT, Left Valid RT, Right Valid RT)
- Pittsburg Sleep Quality Index
    1. PSQI total (sum of the component scores, DURAT + DISTB + LATEN + DAYDYS + HSE + SLPQUAL + MEDS)

    PSQI Component Scores

    2. Sleep Efficiency is calculated in a two-step process
    Step 1: Calculate Sleep Efficiency
    Sleep Efficiency = [Duration of Sleep/(Difference between GNT and GMT)]*100

Step 2: Recode Sleep Efficiency

i. IF Sleep Efficiency ≥ 85, THEN set value to 0
ii. IF Sleep Efficiency < 85 and ≥ 75, THEN set value to 1
iii. IF Sleep Efficiency < 75 and ≥ 65, THEN set value to 2
iv. IF Sleep Efficiency < 65, THEN set value to 3

3. Sleep Disturbance value is calculated based on the sum of (Q5b + Q5c + Q5d + Q5e + Q5f + Q5g + Q5h + Q5i + Q5j)
   i. If sum = 0, Sleep Disturbance = 0
   ii. If sum ≥ 1 and ≤ 9, Sleep Disturbance = 1
   iii. If sum > 9 and ≤ 18, Sleep Disturbance = 2
   iv. If sum > 18, Sleep Disturbance = 3
4. Duration of Sleep value is recoded based on hours of sleep indicated
   i. IF Q4 ≥ 7, Duration of Sleep = 0
   ii. IF Q4 < 7 and ≥ 6, Duration of Sleep = 1
   iii. IF Q4 < 6 and ≥ 5, Duration of Sleep = 2
   iv. IF Q4 < 5, Duration of Sleep = 3
5. Sleep Latency value is calculated in a two-step process
   Step 1: Recode Q2 into Q2new
   i. IF Q2 ≥ 0 and ≤ 15, Q2new = 0
   ii. IF Q2 > 15 and ≤ 30, Q2new = 1
   iii. IF Q2 > 30 and ≤ 60, Q2new = 2
   iv. IF Q2 > 60, Q2new = 3
   Step 2: Calculate and Recode
   v. IF Q5a + Q2new = 0, Sleep Latency = 0
   vi. IF Q5a + Q2new ≥ 1 and ≤ 2, Sleep Latency = 1
   vii. IF Q5a + Q2new ≥ 3 and ≤ 4, Sleep Latency = 2
   viii. IF Q5a + Q2new ≥ 5 and ≤ 6, Sleep Latency = 3
6. Day Dysfunction Due to Sleepiness is calculated and re-coded as follows
   i. IF Q8 + Q9 = 0, Day Dysfunction Due to Sleepiness = 0
   ii. IF Q8 + Q9 ≥ 1 and ≤ 2, Day Dysfunction Due to Sleepiness = 1
   iii. IF Q8 + Q9 ≥ 3 and ≤ 4, Day Dysfunction Due to Sleepiness = 2
   iv. IF Q8 + Q9 ≥ 5 and ≤ 6, Day Dysfunction Due to Sleepiness = 3

3. DATA NOT FROM CRFS

The Optum™ SF-12v2 assessment is an online questionnaire used to measure physical and mental health. Composite scores for these two measures are calculated by Optum, stored on Optum servers, and can be extracted as needed. Data for the continuous performance task included in the EDC does not reflect all the raw data available for analysis. Raw data for this online assessment is captured and stored on Amazon servers and can be extracted for analysis.

- SF-12 Physical Composite Score
- SF-12 Mental Composite Score
- Continuous Performance Task

- Posner Task
- Spatial Working Memory
- Greyscales

4. CALCULATION OF DOMAIN COMPOSITE SCORES

The composite is composed of the individual measures described in section 1 (above), combined as follows:

i. For each included variable convert the distribution of raw scores in the ITT population at baseline to z-scores (with a mean of 0 and a standard deviation of 1. Note that the raw score distribution may be non-normal (with skew or kurtosis) and the goal of this step is to transform that distribution into a normal distribution.
ii. Calculate the composite score at V_1, V_2, and V_3 for each participant based on the baseline data set.

Five a priori domain specific composites are defined as follows and constructed as above
Spatial Cognition: Greyscales Task (raw scores), Spatial Working Memory Task (raw scores)

Cognitive Performance: DKEFS Verbal Fluency (Letter Fluency, raw scores; Category Fluency total, raw scores; Category Switching total, raw scores; Category Switching, switching Accuracy total, raw scores), WAIS-IV Digit Span total (sum of forward, backward, sequencing raw scores), WAIS-IV Digit Span (Forward total, raw scores; Backward total, raw scores; Sequencing total, raw scores), Continuous Performance Task (raw scores)

Functional Ability: Catherine Bergego Scale total (raw score), Barthel Index total (raw score)

Quality of Life: SF-12 Physical Composite Score, SF-12 Mental Composite Score

Quality of Sleep: Pittsburgh Sleep Quality Index total (sum of component scores), Pittsburgh Sleep Quality Index Sleep Efficiency (raw score), Component scores

ANALYSIS POPULATIONS

There are four *a priori* defined analysis populations, including a primary analysis population (i), three secondary analysis populations designed to compare effect sizes: populations with no missing data (ii), populations who completed all training visits (iii), and populations that demonstrated a bias in spatial attention at baseline (iv).

iii. *Intent to Treat (ITT) population:* This is the *a priori* primary analysis population, defined as including all randomized participants who completed the training set-up visit
iv. *Intent to Treat (ITT) Fully-Evaluable (FE) population:* This is a secondary analysis population, defined as including all members of the ITT population that complete a post-intervention or V2 visit (the ITT-FEV2 population), or all members of the ITT

population that completed a V3 or delayed follow up visit (the ITT-FEV3 population). Note that a participant may complete a specific visit but have missing data for a test in which case the participants is in the overall FE population but does not contribute data to the FE population for that visit, e.g., the number of evaluable cases for a specific test on a specific visit may be smaller than the FE population for that visit because of missing data.

v. *Intent to Treat (ITT) Fully-Trained (FT) population:* This is a secondary analysis population, defined as including all members of the ITT-FE population who complete a target number of training visits (32 visits). Note that the FT populations are strict subsets of the FE populations; a person who completes the target number of training visits but does not complete the evaluation visit is not a member of the FT population.

vi. *Intent to Treat (ITT) Spatial*-Bias *(SB) population*: This is a secondary analysis population, defined as including all members of the ITT population who demonstrate a bias in spatial attention that is numerically greater on the ipsilesional side of space on the primary outcome measure (Posner cueing task – field bias).