April 26, 2022

Statistical Analysis Plan for "Home-Based Computerized Cognitive Training for Mild Cognitive Impairment: 78-week randomized blinded trial of crossword puzzles versus games"

1. Introduction.

The aim of the study was to evaluate the efficacy of computerized cognitive training (suite of videogames, denoted as "Games") as compared to computerized crossword puzzle training (denoted as "Crosswords") in the two-site, 78-week, COG-IT randomized clinical trial for study participants with mild cognitive impairment (MCI). This statistical analysis plan provides more detailed descriptions of the statistical analyses reported in the paper.

2. Study Design

One hundred and seven participants aged 55-95 years with clinical diagnosis of MCI were randomized to Games or Crosswords at 1:1 ratio. The randomization was stratified by site (Columbia or Duke), MCI type (early or late MCI), and age (70 years and below or 71 and above). Block randomization of random size (2,4) was used to prevent clinicians from guessing what the next participant's treatment might be. Participants were evaluated at five scheduled visits (Weeks 0, 12, 32, 52, and 78) and research staff conducted three scheduled phone calls (Weeks 20, 42, and 64). Home-based computerized training for Games or Crosswords comprised four 30-minute training sessions weekly for 12 weeks. The six booster sessions each consisted of four 30-minute sessions. At weeks 32, 52, and 78, participants completed three booster sessions at home and the fourth in-clinic. At weeks 20, 42, and 64, participants completed all four booster sessions at home.

3. Hypotheses and Outcome Measures.

Hypothesis 1. Participants with MCI randomized to Games will show more improvement in cognitive outcomes as compared to Crosswords.

Primary outcome: ADAS-Cog 11 total score measured at baseline, 12 weeks, 52 weeks and 78 weeks.

Secondary outcome: Neuropsychological Testing Composite z-score measured at baseline, 12 weeks, 52 weeks and 78 weeks. Another secondary outcome, NCPT online cognitive performance test (measured at baseline and 78 weeks), was not analyzed because only 54 of 107 participants provided week 78 data due to software discontinuation.

Hypothesis 2. Participants with MCI randomized to Games will show more improvement in functional outcomes as compared to Crosswords.

There were two functional outcomes: UPSA and FAQ.

Main functional outcome: UPSA total score measured at baseline, 32 weeks and 78 weeks.

Additional functional outcome: FAQ total score measured at baseline, 12 weeks, 20 weeks, 32 weeks, 52 weeks and 78 weeks.

Hypothesis 3. The proportion of participants converting to dementia will be lower in the Games group as compared to Crosswords.

Outcome: Binary indicator of transitioning from MCI to dementia at 78 weeks.

Hypothesis 4: Alzheimer's disease risk variables (apolipoprotein E e4 genotype, baseline hippocampal volume, baseline UPSIT score**) will moderate the relationship between treatment group and change in ADAS-Cog11, the primary outcome.**

Exploratory Hypothesis: Participants with MCI randomized to Games will show less decline in brain MRI outcomes as compared to Crosswords.

Outcomes: change in hippocampal volume from baseline to 78 weeks; change in cortical thickness from baseline to 78 weeks.

4. Sample Size Calculation.

The sample size was calculated based on the primary outcome: change in ADASCog11 from baseline to end-point week 78. Power analysis used the RMASS program for longitudinal studies. A sample size of 100 participants was projected to detect an effect size of d=.58 with 80% power at 5% level of significance, assuming dropout uniformly distributed over time to reach 20% by 78 weeks. The projected sample size increased to 110 to balance site demographics based on DSMB recommendation but ended at 107 in March 2020 because of pandemic restrictions.

5. Statistical analysis

The analyses were conducted on the Intent-to-treat (ITT) sample, i.e., all randomized participants according to the treatment that they were assigned. Missing data on outcome variables were dealt with by using (longitudinal) linear mixed effects models which do not require complete measurements under the "missing at random" assumption. All hypotheses were tested at level of significance of 5%. There was no adjustment for multiple statistical comparisons in this pilot trial. All analyses were conducted using SAS 9.4.

Before testing the hypotheses, we first examined patients' baseline characteristics to make sure covariates were balanced between the two treatment arms. Nonparametric Wilcoxon's rank sum test was used for continuous and ordinal variables and Fisher's exact test was used for categorical variables.

Linear mixed effects models were used to evaluate the efficacy of Games as compared to Crosswords on cognitive and functional outcomes (**Hypotheses 1 and 2**). Specifically, for each outcome measure, we considered the following model

$$\Delta Y_{it} = \beta_0 + \beta_1 * Group_i + \boldsymbol{\alpha} * Time_{it} + \boldsymbol{\gamma} * Group_i * Time_{it} + \delta Y_{i0} + b_i + \epsilon_{it}, \qquad (1)$$

where ΔY_{it} is change of the outcome measure (baseline minus t weeks) for participant i at timepoint t, $Group_i$ is the treatment group indicator for participant i (1= Games and 0=crosswords), $Time_{it}$ is the visit time point (treated as a categorical variable), Y_{i0} is baseline value of the outcome measure for participant i, b_i is a participant-specific random intercept, and ϵ_{it} is the unexplained residual error term. $(\beta_0, \beta_1, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \delta)$ are fixed effects parameters. SAS PROC MIXED procedure was used to estimate the parameters using REML and random

intercept model. The efficacy of Games versus Crosswords at each time point was tested by forming contrasts of the fitted model. This model was then further adjusted for the three stratification variables (site, age group and MCI type).

We also conducted stratified analysis for each stratification factor separately. For example, for MCI type, we modify model (1) to incorporate interaction with MCI type, namely,

$$\Delta Y_{it} = \beta_0 + \beta_1 * Group_i + \boldsymbol{\alpha} * Time_{it} + \boldsymbol{\gamma} * Group_i * Time_{it} + \delta Y_{i0} + (\beta_2 + \beta_3 * Group_i + \boldsymbol{\alpha_1} * Time_{it} + \boldsymbol{\gamma_1} * Group_i * Time_{it} + \delta_1 Y_{i0}) * MCItype_i + b_i + \epsilon_{it}, \quad (2)$$

The efficacy of Games versus Crosswords for early MCI and late MCI at each time point was tested by forming contrasts of the fitted model.

For **Hypothesis 3**, Fisher's exact test was used to test the association between the binary outcome of dementia and treatment group. Since the number of participants transitioning to dementia was very small, we didn't further adjust for covariates.

For **Hypothesis 4**, linear mixed effects models were used to examine the moderator effect of each Alzheimer's disease risk variable. The model is similar to (2) with MCI type replaced by the Alzheimer's disease variable. The moderator effect was examined by testing the moderator by treatment interaction at each time point.

For the **exploratory hypothesis**, linear regression model was used to compare the effect of treatment on change in MRI outcome (baseline minus 78 weeks), adjusting for baseline values of the MRI measure. The model was then further adjusted for the three stratification variables. Parameters were estimated using SAS PROC GENMOD procedure with normal distribution and identity link.

For safety and compliance, we examined adverse events and number of completed sessions. For each type of adverse event, number and proportion of participants who experienced the adverse event was reported and compared by treatment arms using Fisher's exact test. Number of completed sessions was compared by treatment groups, and separately by site, using Wilcoxon's rank sum test. The proportion that had treatment interruption due to Covid-19 was reported and compared by treatment groups using Fisher's exact test.