LabyrinthVR Study Protocol and Statistical Analysis Plan

Document date: August 22, 2019

Study Protocol

1. Participant arrives for their pre-training cognitive assessment session and is first oriented to the study details. Informed consent is taken.

2. Cognitive Assessments at Pre-training session: MDT, WALK, 4MT, TOVA
During his/her first visit, a participant will be oriented to their respective training task and procedures, followed by the expectancy questionnaire.
LabyrinthVR or placebo iPad games orientation.
Blood draw may be taken before cognitive testing.
fMRI/MRI may be collected in association with some of the tasks.
Time on tasks, including orientation to the games for the participant's assigned condition, is approximately 3 hours.

3. Labyrinth Training Session Procedures (3 sessions per week, 4 to five weeks):
Warm up (~3 minutes)-
Training-
"The purpose of this level is to provide your first introduction to this City so that you begin learning a new neighborhood as well any potential errands you might be assigned to complete later on. Please find all of the errands, indicated by three visual cues, and then exit through the Park gates. While walking speed is not important, your time to walk around will run out at ten minutes. If you require more time than that to learn the neighborhood, we will ask you to repeat this training level."
Participants should move through as many levels as times allows the first session. Participants will take the goggles off and break every three trials, or more often if requested. They should stay in the OMNI platform to protect from falling while wearing the VR headset.
At the end of the training session, study member will
review session feedback with the participant on their progress for that day. If the participant indicates any feeling or other symptoms of nausea or dizziness, participant will not continue with the study. If the participant reports feeling fine, then proceed to the next training session.

4. Cognitive Assessments at Post-training session: MDT, WALK, 4MT, TOVA
LabyrinthVR or placebo iPad games orientation
Blood draw may be taken before cognitive testing.
fMRI/MRI may be collected in association with some of the tasks.
Time on tasks, including orientation to the games for the participant's assigned condition, is approximately 3 hours.

5. Blood Draws:
Up to 200 mls of blood will be collected from each draw, which provides sufficient numbers of cells for the in vitro studies of the factors listed below.
We will assay blood (in batches) for chemistries and cell counts (i.e., CBC , electrolytes) and most of the following factors identified by neuronally-derived exosomes (NDE): Neurosteroids, BDNF, synaptotagmin, neurogranin, synaptopdodin, synaptophysin, and GAP43.

6. Follow-up Cognitive Assessments at 90-days after Post-training assessments:
MDT, WALK, TOVA

Statistical Analysis Plan

Evidence of target engagement achieved by the LabyrinthVR training regimen will come through the results of cognitive outcome assessments with associated neuroimaging data. These results will directly address whether the training delivered by the adaptive wayfinding game (i.e., the intervention) is linked with improved performance in specific, untrained cognitive capabilities, relative to placebo controls. Because the cognitive and neuroimaging data are all compared as repeated measures, results will show changes in functional outcomes for each participant or patient. Those individual change scores will then be compared between factors in order to identify the effects of manipulations revealed by the experimental design.

Each of the independent cognitive measures will be analyzed with three-factor repeated-measures ANOVA as 2 Times (T 1 | T2) X 3 Conditions (Game manipulation 1 | Game manipulation 2 | Placebo Control. Results showing are greater for either training condition than for the placebo control condition, will show target engagement and the beneficial effect from the training intervention on cognitive performance for participants. Statistical comparisons in each ANOVA will use the threshold < 0.05 for reliable differences between factors. Each outcome measure is independent of the others and could be considered a sign of training effects in its own right. As follow-up comparisons, the analyses will also be based on ANCOVA, which may better account for variability in the groups' baseline performance that can dilute the statistical significance of training-induced gains evident in change scores T2-T1.

Follow-Up Assessments T3: The stability of training effects will be probed approximately 90 days after post-training assessments, when the cognitive outcome measures of high-fidelity LTM retrieval will be assessed again during the final visit for Labyrinth training participants (i.e., T 3). Repeated-measures ANOVA will compare each of the independent outcome measures across performance assessed at T1, T2 and T3.

Power Calculation for Sample Size-
In the Labyrinth Pilot Results, we observed a difference between the change scores for the LabyrinthVR versus the Placebo Control subgroups (i.e., T2-T1 = LDI), such that mean LDI = 0.11. Considering this observed training effect and variability observed from a larger sample in the MDT task (i.e., LDI = 0.11, SD = 0.12, respectively), Cohen's d ~ 0.9. Although such an effect size can be estimated to achieve 0.80 power to detect true positives with only 21 samples per condition (i.e., n = arm), we propose a more conservative approach to collect sample size of 27 per arm to achieve 0.80 power assuming effect size ~ 0.8.

MRI Prescriptions and Data Analysis-
Scanner images will be acquired on the UCSF Neuroscape Siemens 3T Magnetom Prisma with a 64-channel head coil. A high-resolution anatomical T1-MPRAGE scan will be collected in 160 volumes, 1mm isotropic voxels and 240 mm2 FoV. For functional scans, 48 T2*-weighted gradient-echo slices (no skip, voxel size = 2.5mm isotropic, TR = 1500ms, TE = 28ms, flip angle 80°, and 211 mm2 FoV) will be acquired using a multiband prescription in an axial oblique orientation parallel to the longitudinal axis of the hippocampus. Diffusion weighted imaging (DTI) will be acquired with 64 directions and five b-zero volumes.
Functional data will be modeled using a general linear model (GLM) in Analysis for Functional NeuroImages (AFNI_2017). To begin the univariate analyses, each participant's GLM will be calculated with task-related regressors and six nuisance variables for motion parameters. Volumes associated with trials showing head movement >1.0mm will be excluded from fMRI analysis.
Task-related regressors will model the amplitude of the impulse response for the 15 seconds (10 TR's) following stimulus onset. Fit coefficient maps for critical task-related regressors will be extracted from each GLM (i.e., univariate analysis). Taking account of concerns about serial correlations in functional data collected in rapid event-related designs with Multiband EPI sequences, a 1.5s TR will be used to avoid the temporal resolution threshold at which additional corrections on statistical inference have been suggested.

Mapped differences in fMRI connectivity associated with an interaction of Time and Condition will identify functional networks associated with target engagement and, critically, suggest the localization of key changes in cortical plasticity arising from LabyrinthVR training.

A trial-wise beta-series correlation method will be used to analyze whole-brain functional connectivity, based on seed regions of interest (ROIs) obtained in the univariate analyses. For each participant, a new GLM design matrix will be constructed to model each trial with a unique covariate. Whole-brain maps of beta–series correlation coefficients will be estimated from each seed ROI to each gray matter voxel in each task-related category. Similar to the univariate analysis, beta-series maps will be compared using 2 X 3 ANOVA.

Statistical maps for pair-wise comparisons in all fMRI analyses will have a voxel-wise threshold of $p<0.005$, and then correction to parameters determined by AFNI 3dFWHMx (Dec 2017 version) to hold reliable cluster-extent, $p\text{-corrected} < 0.05$.