Statistical Analysis Plan (SAP)

Principal Investigator	Mara Mather, PhD Professor of Gerontology, Psychology and Biomedical Engineering University of Southern California Los Angeles, California, USA
Protocol identification number	UP-17-00219
ClinicalTrials.gov identifier	NCT03458910
Author	Kaoru Nashiro, PhD Research Associate Professor University of Southern California Los Angeles, California, USA
Version	Draft 0.2
Date	09/09/2024

1. Objectives and hypotheses
   The main objective of the study was to test the hypothesis that daily biofeedback sessions stimulating heart rate oscillatory activity in baroreflex frequencies affect the function of brain networks involved in emotion regulation. Our main outcome measure was pre-to-post intervention changes in resting-state right amygdala functional connectivity with a medial prefrontal cortex region associated with heart rate variability (HRV). Our secondary outcome measures were pre-to-post intervention changes in up- and down-regulation of amygdala activity and self-reported emotion regulation effectiveness during viewing emotional pictures, as well as changes in ratings of emotional well-being. Other secondary outcome measures included HRV during rest, measures of cerebral blood flow, decision making, stress responsivity and cognitive performance. As a post-hoc outcome measure, we examined pre-to-post intervention changes in resting-state left amygdala functional connectivity with a medial prefrontal cortex region.

2. Sample Size & Power consideration
   121 younger adults and 72 older adults participated in the study.
   No prior studies had examined effects of these interventions on brain function so we were unable to estimate effect sizes based on prior neuroimaging data. We elected to power our study to detect medium or larger effect sizes. Our main planned statistical comparisons were repeated-measures ANOVAs with within-between interactions. For these, a total sample size of 46 would give 90% power to detect moderate effect sizes of f=.25 with α = .05, given an assumed correlation among the repeated measures of .5 (Faul et al., 2007). We also planned to examine within-subject change within each of the conditions. A sample size of 44 in each condition (HRV-increase, HRV-decrease) would give 90% power to detect within-group change effect sizes of d=.5 in a two-tailed t-test with α = .05 (Faul et al., 2007). Thus, we aimed for an N=200 completion rate across the two conditions and two age groups to be able to accommodate potential exclusions for movement during imaging or other data quality issues. Due to the COVID-19 pandemic leading to increased risk for participants completing lab visits, we terminated the study before achieving our initial goal of recruiting 100 older participants.

   Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* **39**, 175-191 (2007)

3. Outcomes
   For most outcome measures, we examined group difference (HRV-increase vs. HRV-decrease) in pre-post intervention change. The average time between pre and post intervention assessment was 5 weeks.

   Primary outcome measures
   - Group difference in pre-post intervention change in resting state functional connectivity between the right amygdala and medial prefrontal cortex (mPFC)

Secondary outcome measures

- Group difference in pre-post intervention change in:
  - Emotion regulation ability measured by self-reported effectiveness when instructed to regulate emotion during viewing emotional pictures
  - The ability to up- and down-regulate amygdala activity when instructed to regulate emotion during viewing emotional pictures
  - Emotional well-being measured by the Profile of Mood States (POMS), the State Anxiety Inventory (SAI), the Trait Anxiety Inventory (TAI) and the Center for Epidemiologic Studies (CES-D)
  - HRV measured by high frequency power (HF)
  - HRV measured by low frequency power (LF)
  - HRV measured by the root mean square of successive differences (RMSSD)
  - Inflammatory markers including C-reactive protein, IL-1b, IL-6, IL-8 and TNF-a
  - Stress recovery
  - Stress reactivity
- Group difference in cerebral blood flow during pre-intervention resting state and post-intervention paced-breathing
- Group difference in decision-making ability measured by multiple-choice responses during a computer-based decision-making task at the post-intervention MRI
- Group difference in the ability to up- and down-regulate task-relevant brain regions during a computer-based decision-making task at the post-intervention MRI

Other outcome measures

- Group difference in pre-post intervention change in:
  - Working memory measured by NIH Toolbox List Sorting Working Memory Test (LSWM)
  - Processing speed measured by NIH Toolbox Pattern Comparison Processing Speed Test (PCPS)
  - Inhibitory control and attention measured by NIH Toolbox Flanker Inhibitory Control and Attention Test (Flanker)
  - Sustained attention measured by Sustained Attention to Response Test (SART)
  - Stress measured by cortisol levels
- Group difference in recognition memory after about three and half weeks of training
- Group difference in recall memory after about three and half weeks of training

Safety outcomes

No adverse events were observed.

4. Populations and subgroups to be analyzed

Population analyzed is composed by all randomization subjects who completed assessments and whose data quality was sufficient for data analysis.

5. Analyses

Primary outcome measures

- Resting state functional connectivity between the right amygdala and medial prefrontal cortex (mPFC): We performed a 2 (time point: pre, post) × 2 (condition: HRV-increase, HRV-decrease) mixed ANOVA. The effect of interest was a time x condition interaction, and P values less than 0.05 were considered statistically significant.

Secondary outcome measures

- Emotion regulation (behavioral analysis):
  We conducted a three-way ANOVA to examine whether emotional intensity rated by participants during emotion regulation changed after HRV biofeedback intervention for younger and older adults separately. The 2 x 2 x 3 ANOVA model included three factors: time-point (2 levels: pre- and post-intervention), intervention condition (2 levels: HRV-Increase, HRV-Decrease) and regulation type (3 levels: down-regulation, viewing and up-regulation). P values less than 0.05 were considered statistically significant.

- Emotion regulation (MRI analysis):
  We conducted a three-way ANOVA to examine whether BOLD activity in the left and right amygdala changed during emotion regulation after HRV biofeedback intervention for younger and older adults. The 2 x 2 x 3 ANOVA model for each amygdala included three factors: time-point (2 levels: pre- and post-intervention), intervention condition (2 levels: HRV-Increase, HRV-Decrease) and regulation type (3 levels: down-regulation, viewing and up-regulation). P values less than 0.05 were considered statistically significant.

- Emotional well-being (pre/post):
  To test whether emotional well-being changed during the course of the study and whether this depended on training condition, we performed a series of two-way mixed ANOVAs, one for each measure. In these ANOVAs, time was a within-subjects factor (2 levels: pre- and post-intervention) and condition was a between-subjects factor (2 levels: HRV-Increase, HRV-Decrease). For each ANOVA, the effect of interest was a time x condition interaction, and only data from participants with values for both time points were included. P values less than 0.05 were considered statistically significant. ANOVAs were performed with R (Version 4.0.3) and the R package 'ez'.

- Emotional well-being (all weeks):
  To test whether emotional well-being changed during the course of the study and whether this depended on training condition, we fit a series of linear mixed effects models, one for each measure. For these models, fixed effects were time (1 level per week), condition (2 levels: HRV-Increase, HRV-Decrease), and a time x condition interaction. The models included random intercepts for participants. For each model, data from all available participants at each time point was used. To assess parameter significance, F- and p-values were determined using Satterthwaite's method. P values less than 0.05 were considered statistically significant. These

analyses were performed with R (Version 4.0.3) and the R packages 'lmer4' and 'lmerTest'.

- HRV measured by high frequency (HF) HRV, low frequency (LF) HRV and the root mean square of successive differences (RMSSD):
  To test whether HRV changed during the course of the study and whether this depended on training condition, we performed a series of two-way mixed ANOVAs, one for each measure. In these ANOVAs, time was a within-subjects factor (2 levels: pre- and post-intervention) and condition was a between-subjects factor (2 levels: HRV-Increase, HRV-Decrease). For each ANOVA, the effect of interest was a time x condition interaction, and only data from participants with values for both time points were included. P values less than 0.05 were considered statistically significant. ANOVAs were performed with IBM SPSS statistics (Version 27) and R (Version 4.0.3) with R package 'ez'.

- Inflammatory markers:
  For each inflammatory marker (IL-1b, IL-6, IL-8, TNF-a, CRP), we conducted a two-way ANOVA to examine whether its levels changed after HRV biofeedback intervention for younger and older adults separately. The 2 x 2 ANOVA model included two factors: time-point (2 levels: pre- and post-intervention) and intervention condition (2 levels: HRV-Increase, HRV-Decrease). P values less than 0.05 were considered statistically significant.

- Recovery from stress:
  To test whether stress recovery changed during the course of the study and whether this depended on training condition, we performed a series of two-way mixed ANOVAs, one for each measure of stress recovery. In these ANOVAs, time was a within-subjects factor (2 levels: pre- and post-intervention) and condition was a between-subjects factor (2 levels: HRV-Increase, HRV-Decrease). For each ANOVA, the effect of interest was a time x condition interaction, and only data from participants with values for both time points were included. P values less than 0.05 were considered statistically significant. ANOVAs were performed with R (Version 4.0.3) and the R package 'ez'.

- Reactivity to stress:
  To test whether stress reactivity changed during the course of the study and whether this depended on training condition, we performed a series of two-way mixed ANOVAs, one for each measure of stress reactivity. In these ANOVAs, time was a within-subjects factor (2 levels: pre- and post-intervention) and condition was a between-subjects factor (2 levels: HRV-Increase, HRV-Decrease). For each ANOVA, the effect of interest was a time x condition interaction, and only data from participants with values for both time points were included. P values less than 0.05 were considered statistically significant. ANOVAs were performed with R (Version 4.0.3) and the R package 'ez'.

- Cerebral blood flow measured by Arterial Spin Labeling (ASL):

We conducted a two-way ANOVA to examine whether whole-brain cerebral blood flow, as measured by arterial spin labeling MRI, changed over the course of the study as a result of training condition. Specifically, we examined cerebral blood flow at rest at the beginning of the study (i.e., before any intervention) and we examined cerebral blood flow during paced breathing at study completion (i.e., at the end of 5-week training). A 2x2 ANOVA was modeled with time as a within-subjects factor (2 levels: pre- and post-intervention) and training condition as a between-subjects factor (2 levels: HRV-Increase, HRV-Decrease). The effect of interest was the time x training condition interaction. Only participants with usable arterial spin labeling MRI scans at both time points were included in analyses. P values less than 0.05 were considered statistically significant. Statistical analyses were performed in IBM SPSS Statistics version 27.

- Decision-making (behavioral analysis):
  To compare the response to unfair and fair offers between groups, we conducted two Mann-Whitney tests. In these analyses, acceptance rates for unfair and fair offers were dependent variables and group (HRV-Increase vs. HRV-Decrease) was defined as between subject independent variable. P values less than 0.05 were considered statistically significant.

- Decision-making (MRI analysis):
  We conducted two independent t-test analyses to examine whether BOLD activity in the dorsal anterior cingulate cortex and anterior insula in response to linear effect of unfairness differs between group. Linear effect of unfairness was defined using a linear regression with offer values as independent variable and brain activity as the dependent variable. P values less than 0.05 were considered statistically significant.

Other outcome measures

- Cognition (Working memory, Processing Speed, Inhibitory Control and Attention):
  To test whether cognitive function changed during the course of the study and whether this depended on training condition, we performed a series of two-way mixed ANOVAs, one for each measure. In these ANOVAs, time was a within-subjects factor (2 levels: pre- and post-intervention) and condition was a between-subjects factor (2 levels: HRV-Increase, HRV-Decrease). For each ANOVA, the effect of interest was a time x condition interaction, and only data from participants with values for both time points were included. P values less than 0.05 were considered statistically significant. ANOVAs were performed with IBM SPSS statistics (Version 27).

- Sustained Attention to Response Task (SART):
  We conducted a three-way ANOVA to examine whether number of errors during the SART task changed after HRV biofeedback intervention for younger and older adults separately. The 2 x 2 x 2 ANOVA model included three factors: time-point (2 levels: (pre- and post-intervention), intervention condition (2 levels: HRV-Increase, HRV-Decrease), and error type (2 levels: commission error and omission error). The commission error occurs when participants pressed a button in trials where they are not supposed to press it. The omission error occurs when participants did not

press the button in trials where they are supposed to press it. P values less than 0.05 were considered statistically significant.

- Stress measured by cortisol levels at awakening and 30 min later: We conducted a three-way ANOVA to examine whether cortisol levels at awakening and 30 min later changed after HRV biofeedback intervention for younger adults. The 2 x 2 x 2 ANOVA model included three factors: time-point (2 levels: (pre- and post-intervention), intervention condition (2 levels: HRV-increase, HRV-decrease), and cortisol collection timing (2 levels: cortisol at awakening and 30 min later). P values less than 0.05 were considered statistically significant. To reduce participant burden for older adults, we did not collect saliva samples from them so could not assess this measure for them.

- Recognition memory: We conducted a two-way ANOVA on recognition memory measures to determine if emotional memory differed between conditions as a result of the HRV biofeedback intervention for younger and older adults separately. In the 3 x 2 ANOVA, valence was a within-subjects factor (3 levels: neutral, positive, negative) and condition was a between-subjects factor (2 levels: HRV-Increase, HRV-Decrease). P values less than 0.05 were considered statistically significant.

- Recall memory: We conducted a two-way ANOVA on free recall memory to determine if emotional memory differed between conditions as a result of the HRV biofeedback intervention for younger and older adults separately. In the 2 x 2 ANOVA, valence was a within-subjects factor (2 levels: positive, negative) and condition was a between-subjects factor (2 levels: HRV-Increase, HRV-Decrease). P values less than 0.05 were considered statistically significant.