STATISTICAL ANALYSIS PLAN

MINDFULNESS, EDUCATION, AND EXERCISE FOR COGNITIVE FUNCTION (MEDEX)

Statistical Analysis Plan
(SAP)

Version: 11

Authors: Eric Lenze, Professor of Psychiatry, Washington University SOM
J. Philip Miller, Professor of Biostatistics, Washington University SOM
Michael Yingling, Sr Statistical Data Analyst, Washington University SOM

Date: May 2022

New Edits:

- Updated Section 6 to include how extreme peak cortisol values were handled.
- Included a section to indicate how randomization was conducted.

CONTENTS

This 2 x 2 factorial design randomized controlled trial (RCT) will test the benefits of two interventions for age-related cognitive decline:

Mindfulness Based Stress Reduction (MBSR) teaches mindfulness, or the focusing of attention and awareness, through various meditation techniques.

Exercise – specifically, intense, multi-component exercise that seeks to achieve or exceed 70% of age-predicted heart rate; also appears to affect brain structure and function and improve cognitive performance.

Our main hypothesis is that MBSR and exercise each improve cognitive function in older adults, including memory and executive function. We also hypothesize that greater cognitive benefits result from the combination because of the complementary nature of their mechanisms, such as decreased cortisol with MBSR and improved insulin sensitivity with exercise. Finally, we hypothesize that enhanced neuroplasticity, as shown through changes in structural and functional neuroimaging, explain the cognitive improvements.

Aim 1. Examine effects of MBSR, exercise, and their combination for remediation of memory and cognitive control, with secondary outcomes of everyday cognition, functional performance, and social participation and engagement.

H1: MBSR and exercise will each produce benefits in healthy older adults' cognitive performance, and combined MBSR + exercise will show additive cognitive improvements from both interventions.

Aim 2. Examine mechanistic changes that underlie cognitive remediation from MBSR and exercise.

H2: (a) Decrease in peak cortisol accounts for cognitive improvements with MBSR. (b) Increased insulin sensitivity, aerobic fitness, and brain-derived neurotrophic levels (BDNF) levels account for improvements with exercise.

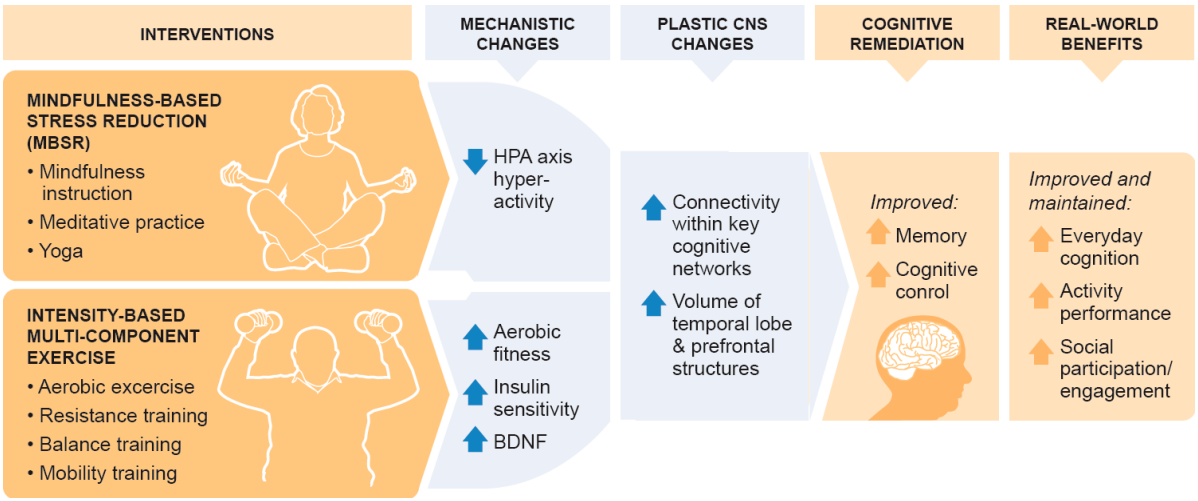
Aim 3. Examine whether the interventions cause changes in brain structure and functional connectivity and whether these changes help explain the cognitive improvements.

H3: Improved functional connectivity within and across specific cognitive networks and increased volume of hippocampal and lateral prefrontal regions account for improved cognitive function with the interventions.

Aim 4. Examine predictors of variability in response to the interventions.

H4: Baseline cortisol and insulin sensitivity will predict degree of cognitive remediation from MBSR and exercise; that is, high baseline cortisol will predict greater improvement from MBSR, while low insulin sensitivity will predict greater improvement from exercise.

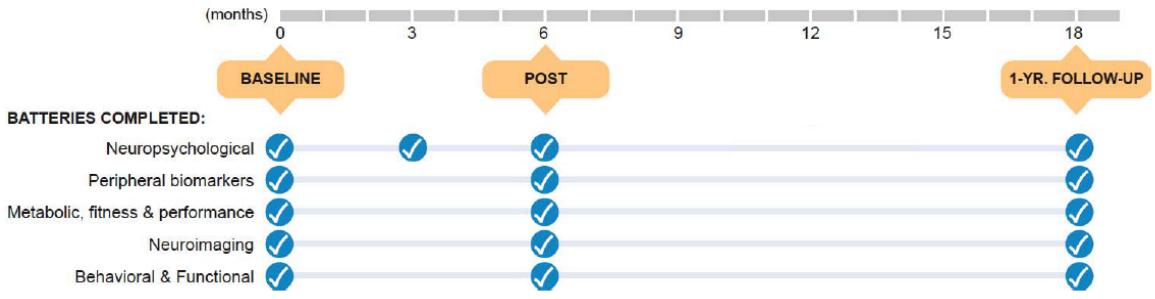
Conceptual model of study hypotheses



3 ANALYSIS SETS/ POPULATIONS/SUBGROUPS

We will randomize 580 non-demented healthy adults aged 65-84, in a 2x2 factorial design: MBSR alone, exercise alone, MBSR + exercise, or a health education control condition.

Timeline of Assessments



4 RANDOMIZATION

Groups of 10 to 15 participants were randomized to one of the four treatment conditions. The only stratum was site. The modified intent-to-treat analyses included all participants that attended at least one MBSR and/or Exercise class.

5 DATA SOURCE

Assessments include cognitive tests, biomarkers, neuroimaging assessments, and functional assessments. The NIH Toolbox for Assessment of Neurological and Behavioral Function and NIH Patient Reported Outcomes Measurement Information System (PROMIS) will be used. Using Ecological Momentary Assessment (EMA), we will assess anxiety, depression, cognition, as well as mindfulness via the Cognitive & Affective Mindfulness Scale (CAMS-R). Biomarkers include salivary cortisol, serum BDNF, and (exploratory) a multiplex peripheral immune panel. Metabolic and physical assessments include estimated VO2 max, oral glucose tolerance test (OGTT), body composition scan, standing balance tests, and the short physical performance battery. The neuroimaging battery includes structural and functional MRI. Data for observed tasks of daily living (OTDL), satisfaction and participation in social roles, and measures of insomnia will be collected.

6 ENDPOINTS AND COVARIATES

The study will consist of a six month acute intervention phase followed by a twelve month maintenance phase with monthly classes and other prompts to maintain intervention behaviors. Pre-specified covariates for the primary and secondary outcome models include site, gender, and age.

7 HANDLING OF MISSING VALUES AND OTHER DATA CONVENTIONS

Multiple imputation is to be used where the statistical model does not already provide appropriate protection for data missing at random. There are many instances where missing data is handled differently. These include:

- At UCSD, a number of individuals were tested with the wrong version of the SART. As a result, these individuals' SART component was not included in the cognitive control composite.
- For the behavioral assessments, as well as the CAMS-R, if the scale consists of 6-10 items and a participant is missing one item, then the average of the other items will be used to fill in the missing item. If more than 1 item is missing, then that participant's data will not be used for that scale. For a 1-5 item scale, any missing data will exclude that participant's data for that scale. For scales with greater than 10 items, a 10% rule will be employed to handle missing data.

- For participants that have only 1 day of usable cortisol data (out of the 3-day collection), that data would be used as the peak cortisol value; for those that have 2 days of usable cortisol data, take the median of the 2 days; for those that have 3 days of usable cortisol data, take the median of the 3 days. Values were excluded that were >3 standard deviations higher than the average peak value for the entire sample at baseline.

8 STATISTICAL PROCEDURES

8.1 Primary Outcome Analyses

8.1.1 **Aim 1. Examine effects of MBSR, exercise, and their combination for remediation of memory and cognitive control, with secondary outcomes of everyday cognition, functional performance, and social participation and engagement.**

H1: MBSR and exercise will each produce benefits in healthy older adults' cognitive performance, and combined MBSR + exercise will show additive cognitive improvements from both interventions.

The neuropsychological battery consists of several measures including list and paragraph (paper and pencil) recall measures, NIH Toolbox cognition (computerized) measures, and cognitive control (computerized eprime) measures. The battery takes approximately 90 minutes to conduct, and it is conducted at 0, 3, 6, and 18 months (with a practice battery done at the time of consent). From the neuropsychological battery, we will create <u>two co-primary outcome variables</u>:

1) A <u>memory composite variable</u> using the list recall, paragraph recall, and picture sequence memory tasks. Specifically, the variables included in the memory composite are:

Memory Composite Variable		
Variable Label	REDCap Name	Definition/Justification
Paragraph Recall Story1+Story2 Immediate	s_im_pos	Used in previous studies, tests immediate recall for stories
Paragraph Recall Story1+Story2 Delayed	s_dl_pos	Used in previous studies, tests for delayed recall for stories
Word List Learning (Immediate) Total Score	wll_totscr	Used in previous studies, tests immediate recall for lists of words
Word List Learning (Delay) Total Score	wlr_totscr	Used in previous studies, tests delayed recall for list of words

Picture Sequence Memory Test Uncorrected Score	------	From NIH Toolbox, tests immediate recall for sequences. The uncorrected score compares the score of the test-taker to those in the entire NIH Toolbox nationally representative normative sample, regardless of age or any other variable.

Each of the above memory variables will be used in the composite. A z-score is computed for each participant [(participant score - mean)/standard deviation], using the mean and standard deviation of that variable computed on all randomized participants at baseline. The composite memory variable is then created by averaging the z-scores of all available memory variables. This composite variable will be computed at baseline and 3, 6, and 18 month visits.

2) A cognitive control composite variable will include scores from the Consonant Vowel Odd Even (CVOE), Sustained Attention Response Task (SART), Color Word Interference, Flanker Inhibitory Control, List-Sorting Working Memory tasks and Dimensional Change Card Sort Test (DCCS). The cognitive control composite variable will be computed in a parallel fashion to the memory composite, using the variables listed below:

Cognitive Control Composite Variable			
Variable Label	REDCap Name	Definition/Justification	Use in Composite
Stroop Color Naming Accuracy	------	Percentage correct across all Stroop trials.	Average of these 2 z-scored variables will be used in composite
Stroop Color Naming Reaction Time	------	Mean response latency across all correct Stroop trials, excluding outliers. An outlier is defined as a reaction time that falls outside of +/- 3 SDs of the mean.	
SART No-Go Accuracy	------	Measures accuracy for No-Go trials. Percentage of times a participant was able to withhold response to a No-Go (i.e., a 3) trial).	Average of these 2 z-scored variables will be used in the composite
SART Go Trial CoV	------	Go Trial Coefficient of Variation (CV; unit-free; higher is more variable; estimate of within-participant RT variability).	
CVOE Switch Trial Accuracy	------	Used as a measure of accuracy for CVOE. Percentage correct-enter in decimal form; percentage of trials in which a participant accurately classified a stimulus, on the subset of trials in which participants switch from consonant-	

		vowel to odd-even judgments (or vice-versa).	
CVOE Standardized Global Switch Cost	------	Measures RT difference during switch to non-switch, normalized to each individual's RT. This variable must only be computed when all data are acquired for an individual. Unit-free; z-scored milliseconds; z-scored difference score that documents how much a participant slows their response on the mixed trials (e.g., having to change from consonant-vowel judgment to odd-even judgments), relative to the pure block (e.g., maintaining the consonant-vowel judgment for the entirety of a block).	Average of these 2 z-scored variables will be used in composite
Flanker Uncorrected Score	------	From Toolbox, measures response inhibition. The uncorrected score compares the score of the test-taker to those in the entire NIH Toolbox nationally representative normative sample, regardless of age or any other variable.	This variable will be used in composite.
List Sorting Working Uncorrected Score	------	From Toolbox, measures working memory. The uncorrected score compares the score of the test-taker to those in the entire NIH Toolbox nationally representative normative sample, regardless of age or any other variable.	This variable will be used in composite.
Dimensional Change Card Sort Uncorrected Score	------	From Toolbox, measures cognitive flexibility. The uncorrected score compares the score of the test-taker to those in the entire NIH Toolbox nationally representative normative sample, regardless of age or any other variable.	This variable will be used in composite.

We will compute an intraclass correlation coefficient (ICC) to determine whether it is necessary to correct the models for group randomization.

We will use marginal models for our time course analyses. Proc Mixed in SAS includes all data, i.e., if a participant has baseline data but no post data, that participant's baseline data is still included in the analysis.

The model will include the between subject main effects of MBSR and exercise, their interaction, as well as the two-way and three-way interactions between time and the between subject effects. In addition, a model that includes site will also be generated to test for site differences. Time (Baseline, 3, 6, and 18 months) will be a within subjects effect with an unspecified covariance matrix because of the uneven time intervals between visits.

The primary test of the effectiveness of MBSR will be the change in the composite scores from baseline to 6 months in the MBSR "yes" participants versus the change in the "no" participants as computed with the appropriate contrast. A similar contrast will be constructed that will involve the change in composite scores from baseline to 18 months

Similarly the test for the exercise intervention will be the difference in change in the composite variables between those randomized to receive the exercise intervention and those who were not.

The interaction test (which will determine whether the two interventions are additive [no significant interaction], or synergistic or interfering) will be tested in an analogous fashion.

All randomized individuals will be included in this primary analysis in order to adhere to the modified Intent to Treat Principle (i.e., all included individuals obtained at least one dose of the intervention). All available data will be included, independent of any noncompliance, dropout or protocol deviations. A 2-tailed significance level of .05 will be used for each of the two primary outcomes. We will compute and report effect sizes with 95% confidence intervals.

8.1.2 **Aim 2. Examine mechanistic changes that underlie cognitive remediation from MBSR and exercise.**

H2: (a) We hypothesize MBSR will indirectly provide a positive influence on cognitive performance by decreasing cortisol. (b) We predict exercise will indirectly provide a positive influence on cognitive remediation by increasing insulin sensitivity, aerobic fitness, and BDNF levels.

We will first test whether any of these potential mediators are affected by the interventions using the models of H1. If they show effects of the interventions, then appropriate structural equation models will be constructed to assess their mediating effects.

Mechanistic marker	Definition
Cortisol Levels	Median of three peak values at each time point. Peak values will reflect the greater of wake or wake+30 on a given day. Area Under the Curve (AUC) will also be computed.
Insulin Sensitivity	HOMA-IR and OGIS values modeled from the OGTT glucose and insulin values at the 0, 90, and 120 timepoints.
Aerobic Fitness	• Exact time (seconds) to reach 85% of age-predicted maximum heart rate • METs at 85% of age-predicted maximum heart rate • VO2 at 85% of age-predicted maximum heart rate
BDNF Level	Not assayed at this point but may be in the future.

8.1.3 **Aim 3. Examine whether the interventions cause changes in brain structure and functional connectivity and whether these plastic changes help explain cognitive improvements.**

H3: We hypothesize the interventions will indirectly provide a positive valuable influence on cognitive performance through its effect on plasticity changes in the central nervous system (CNS). Specifically we expect the interventions will increase hippocampal and ventral/dorsal pre-frontal volumes as well as provide beneficial improvements to resting state networks (RSNs), which in turn will explain cognitive improvement.

For this aim, we will repeat the same marginal model employed in the primary outcome analysis under H1 using structural MRI measures from FreeSurfer (left/right hippocampal volume and left/right prefrontal cortex volume, surface area, or cortical thickness).

The methodology for the resting state is outside the scope of this SAP.

8.1.4 **Aim 4. Examine predictors of variability in response to the interventions.**

H4: We hypothesize two of the mechanistic variables measured at baseline -- cortisol and insulin sensitivity -- will explain variability in cognitive remediation within conditions. Specifically, we predict greater cognitive improvement among participants in the MBSR condition who had high peak cortisol as opposed to lower peak cortisol at baseline. Likewise, we hypothesize improvement in cognition for the exercise intervention will be higher for participants with low insulin sensitivity at baseline as opposed to higher insulin sensitivity at baseline.

To examine which individuals are most likely to benefit from the intervention, we will conduct moderation analyses. These models will be constructed at a later date and will not be included in the primary manuscript.

8.2 Secondary Analyses

Neuropsychological battery:
Similar contrasts as used in the primary outcome analysis will be used to determine the persistence of the effects from 6 months to 18 months.

We will repeat the same marginal model as used for the primary outcome analysis for each of the individual tests used in the composites. We will also look at the correlations between tests within each composite at baseline and test whether that correlational structure changes over time.

Subsequent post-hoc analyses will include per-protocol analyses (tests of primary and MRI outcomes in those who were highly adherent to the interventions). Prior to conducting these analyses, we will remove any participants with intervention contaminations (e.g., those randomized to the MBSR intervention engaging in exercise and vice versa). In addition, we will create subgroups of participants who changed the most (top tertile for putatively beneficial change), vs. those who changed the least (bottom tertile) for several physiological and performance markers.

We will report both session attendance and home practice as measures of adherence as well as whether participants took breaks from participating, which was collected during the maintenance phase.

Other outcome variables:

Models similar to those constructed under H1 will be constructed.

1) OTDL (performance-based task of daily living)
2) Self-reported outcomes via standard questionnaires.
    a. Neuro-QoL—Cognitive Function
    b. Neuro-QoL—Positive Affect
    c. PROMIS—Ability to Participate in Social Roles
    d. PROMIS—Anxiety
    e. PROMIS—Depression
    f. PROMIS—Satisfaction with Participation in Social Roles
    g. PROMIS—Sleep Disturbance
3) Self-reported outcomes via EMA (for outcomes with more than one item, a sum was computed).
    a. Negative Affect
        i. I feel depressed…
        ii. I feel worried…
    b. Positive Affect
        i. I feel cheerful…
        ii. I am able to enjoy life…
    c. Cognitive Problems
        i. I am having trouble remembering things…
        ii. My thinking is slow…
        iii. I am having trouble concentrating…
    d. Mindfulness
        i. I am easily distracted… (reverse scored)
        ii. I accept the thoughts and feelings I have.
        iii. I am focused on the present moment.
    e. Sleep Quality
        i. My sleep quality was…
    f. Sleep Quantity
        i. I slept _____ hours….

Manipulation Checks:
Marginal models similar to those employed in the primary outcome analysis under H1 will be used to test the effectiveness of MBSR and Exercise in several process variables. These include GXT (Time to 85% Age-Predicted Max HR, Mets at 85% Age-Predicted Max HR, and VO2 at 85% Age-Predicted Max HR), CAMS-R, modified-SPPB total score, DEXA total fat mass, and DEXA total lean mass, cortisol AUC, and insulin sensitivity (HOMA-IR and OGIS).

9 QUALITY CONTROL PLANS

Data quality will be assessed either through the REDCap double data entry module, or through periodic chart reviews. If the latter, in the event that data entry errors are discovered, then additional charts will be randomly selected for review. Range and consistency checks will also be employed.

10 PROGRAMMING PLANS

Statistical software (SAS and R) will be used to produce results and raw data will be available for accompanying papers to be submitted for publication.