

Statistical Analysis Plan

Version 1.0

ActiveCBT for Depression: Transforming Treatment Through Exercise Priming

Clinicaltrials.gov identifier: NCT06001346

Version	Date	Who	Comments
1.0	24 Aug 2023	TM	First draft focusing on R61 phase

Contents

1	Introduction	3
1.1	Objective of the Statistical Analysis Plan	3
1.2	Description of the Study Design.....	3
2	Interim DSMB Reviews: Goals and Format	5
3	Enrollment and Eligibility	7
4	Baseline Characteristics	7
5	Pre-CBT Condition Characteristics	7
6	Safety Analyses	8
7	Efficacy Analyses	8
7.1	Primary Outcomes	8
7.2	Secondary Outcomes	8
7.3	Subgroup Analyses.....	9
8	Interim Monitoring	9
9	Sample Size and Power.....	10
10	Data Completeness and Study Conduct.....	10
11	Unblinding of Treatment Comparisons	11

1 Introduction

1.1 Objective of the Statistical Analysis Plan

The objective of this statistical analysis plan (SAP) is to describe the general analytic strategy and the statistical methods used to analyze the data generated by the CBT+ trial. The CBT+ trial is funded by an R61/R33 award and will include an R61 phase, and assuming the R61 milestone is achieved, it will also include an R33 phase.

The study statisticians will update this document in coordination with the clinical investigators prior to unblinding. It may also be updated based on protocol amendments.

1.2 Description of the Study Design

Current treatments for depression, including cognitive behavior therapy (CBT) and medication, are not effective for many people, leading to a high global burden of depression. This study investigates the effects of a novel intervention approach, intentionally sequencing aerobic exercise immediately before CBT sessions to take advantage of the short-term biological (e.g., increased brain-derived neurotrophic factor) and psychological (e.g., lower anhedonia) effects of exercise to increase the potency of the psychotherapy using established techniques to measure both specific (i.e., behavioral activation) and common (i.e., working alliance) factors underlying the antidepressant effect of CBT. This study will determine the potential of exercise priming to enhance key mechanisms of CBT and improve depression.

Research Objectives

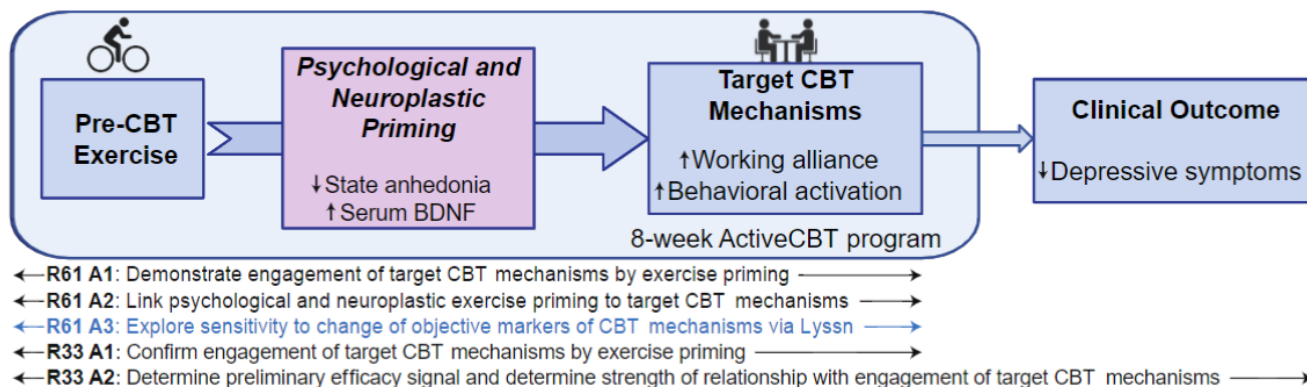
The overall goals of this project are to examine the efficacy of exercise priming (i.e., intentional sequencing of aerobic exercise immediately before cognitive behavioral therapy ['ActiveCBT']) to treat depression. Specific aims are to 1) demonstrate the effects of exercise priming on mechanisms associated with therapy success (i.e., working alliance, behavioral activation), 2) examine how psychological (e.g., anhedonia) and neuroplastic (e.g., brain-derived neurotrophic factor) mechanisms relate to CBT mechanisms, and 3) determine the efficacy of exercise priming for treatment of depression.

Design

The CBT+ trial will compare 30 minutes of exercise versus 30 minutes of quiet rest prior to cognitive behavioral therapy (CBT) for major depressive disorder (MDD). Participants will undergo eight weeks of CBT therapy, with each CBT session preceded by their assigned condition, and during the R33 phase, participants will continue follow-up for depressive symptoms for one year. The R61 phase of the CBT+ trial aims to show modulation of the co-primary mechanistic endpoints Working Alliance Inventory-Short Revised (WAI) and Behavioral Activation for Depression Scale (BADs). In contrast, the R33 phase of the CBT+ trial aims to re-confirm target engagement via WAI and BADs and evaluate effects on depressive symptoms as measured by the Hamilton Rating Scale for Depression (HAM-D).

'Go criteria' to move to the R33 will require 80% of participants to attend ≥ 6 of 8 weekly sessions and an average effect size $d \geq 0.55$ for either of the two target CBT mechanism measures (WAI or BADs) across 8 weeks favoring ActiveCBT or both at $d \geq 0.35$. The average effect size will reflect the average of 8 weekly estimates of Cohen's d , i.e., the mean

difference divided by the pooled standard deviation. Effect sizes of $d = 0.55$ and $d = 0.35$ are anticipated to reflect 13 and 9 points for BADS and 7 and 5 points for WAI, respectively.



Population

The target population for this trial is adults between 18 and 65 with MDD and at least mild depressive symptoms who are CBT-naïve and are willing and able to safely perform exercise. Additional eligibility criteria are described in the study protocol.

Duration

Participants will be followed for at least 8 weeks of CBT therapy and for 1 year after completing therapy.

Randomization

During the R61 (R33) phase, participants will be randomized 1:1 (2:1) to either exercise or quite rest. Randomization will be stratified by mild versus moderate to severe depression (i.e., HAM-D ≤ 17 versus >17). Allocation schedules will be pre-generated using permuted blocks with randomly varying block sizes. The R61 phase will use block sizes of 2, 4, and 6, and the R33 phase will use 3, 6, and 9 block sizes.

Sample size

The R61 phase of the trial will enroll 40 participants, and the R33 phase will enroll 100 participants.

The R61 sample size of 40 participants (20 in each condition) was selected to ensure high probability for achieving the R61 'Go' criteria for target engagement under the expected effects of Cohen's $d \geq 0.50$ for both WAI and BADS (i.e., target engagement), and a low probability for achieving the R61 'Go' criteria the when $d = 0$ for both WAI and BADS (i.e., no target engagement). Table 2 contains probabilities of achieving the R61 'Go' criteria for different effects on WAI and BADS, and correlations between WAI and BADS.

Table 2: Probability of Meeting R61 'Go' Criteria				
True Effect (Cohen's d)	WAI BADS	Correlation between WAI & BADS		
		0.1	0.4	0.7
	0.00 0.00	0.004	0.007	0.012
	0.50 0.00	0.395	0.390	0.389
	0.50 0.50	0.759	0.737	0.742
	0.75 0.00	0.872	0.871	0.871
	0.75 0.75	0.992	0.985	0.982

Assuming 10% drop-out, the R33 phase enrollment of 100 participants (67 in the exercise condition and 33 in the quiet rest condition) will provide 90% power to detect standardized regression coefficients (Pearson's correlations) between each average target CBT mechanism measure and HAM-D of $r = 0.33$. Allocation among completers of 60:30 will provide 90% power to detect a $d = 0.73$ difference between arms on each target CBT mechanism at each of the 8 assessments. Combining the minimum detectable r and d implies that these data will facilitate detecting a preliminary effect on HAM-D of anything greater than $0.24 \text{ SD} = 0.33 \times 0.73$ units, which is likely ~ 2 points and the smallest clinically meaningful difference suggested for HAM-D. Secondly, these data will provide 90% power to detect a between-group effect on HAM-D at 8 weeks and 12 weeks later of $d = 0.73$, or 6 points, which would constitute a large meaningful effect; however, the goal of the R33 phase is not to establish this direct effect, rather it is to provide preliminary evidence for a direct effect through engagement of the target mechanism of higher working alliance and behavioral activation. Moreover, the reported minimum detectable r and d values are conservative as models will be adjusted for baseline covariates and incorporate partial data from drop-outs.

Monitoring

An independent DSMB will review interim data regularly for safety. If the trial is stopped early, further enrollment will be terminated if applicable, and the trial data will be unblinded. The extent of unblinding will be guided by DSMB recommendations. It is expected that all participants will continue to be followed through the completion of data collection.

2 Interim DSMB Reviews: Goals and Format

Analysis cohort: Safety and efficacy analyses (closed reports, described below) will be conducted in the intention-to-treat (ITT) population.

Goals of the interim reviews:

- Review the conduct of the trial.
- Protect the safety of study participants.
- Advise on modifying/stopping the trial for evidence of harm or poor trial conduct.

Timing of the reviews: The DSMB will conduct reviews twice per year and approximately every 6 months.

The DSMB may request more frequent reports focused on safety at any time.

Review meetings will typically consist of an open session, closed session, optional executive session (closed), and optional debrief (open).

Open report to the DSMB

The open report will contain:

- A synopsis of the trial design and current status of the trial
- Responses of the study team to DSMB requests
- A summary prepared by the study leadership
- Data summaries for all enrolled participants pooled across randomization groups (blinded):

- Enrollment
- Baseline characteristics
- Eligibility violations and protocol deviations.
- Summary reports for data completeness and study conduct
- Intervention fidelity (unblinded)

Unless noted otherwise, all data summaries in the open report will be pooled across the randomization groups.

While the study is ongoing, unless noted otherwise, summaries by randomization group and comparisons of randomization groups will be restricted to the confidential closed report. Additionally, all summaries of follow-up data other than adherence, data completeness, intervention fidelity, and study conduct (pooled across randomization groups) will be restricted to the confidential closed report.

Closed report to the DSMB

The closed report will contain the following data summaries by (masked) randomization group:

- Data summaries in the open report, by randomization group (enrollment, baseline characteristics, eligibility violations).
- Data summaries on data completeness and study conduct.
- Data summaries to assess safety, including event listings for incident AEs and SAEs
- Other data summaries requested by the DSMB. During the course of the trial the DSMB may request additional analyses be prepared by the study statistician. If such requests respond to an emerging safety trend, they should be kept confidential.

Cut-date for interim reviews: Analysis data sets will be locked several weeks prior to the review date to allow the study statistician time to prepare a consistent report and the DSMB adequate time to review the report prior to the scheduled meeting.

3 Enrollment and Eligibility

For the open report, the following enrollment and eligibility summaries will be provided:

- Enrollment over calendar time: plot by day or week, cumulative and increments.
- Eligibility: number (%) and reasons for in-eligibility

For the closed report, enrollment and reasons for in-eligibility will be summarized by randomization group.

4 Baseline Characteristics

Unless noted otherwise, categorical variables will be summarized with the number and percentage (N, %) of participants in each category, and continuous variables will be summarized with the median and interquartile range (IQR); in the open report, in addition, the mean (SD) and range may be provided. The data summaries will be provided overall (i.e. pooled across randomization group) in the open report, and by (masked) randomization group in the closed report.

The following baseline characteristics will be reported:

- Demographic Information
 - Age
 - Sex at birth and gender
 - Race / Ethnicity
 - Education level
 - Income
 - Marital status
 - Employment
- Medical Information
 - BMI
 - Average SBP, DBP, HR
 - Current cardiovascular medication
 - Current tobacco use
 - Past and current medical conditions
 - Baseline health assessments (e.g. SF-36, GAD-7, NCI, ACE, VAS and DARS Anhedonia, PHQ-9, BADS, ATQ, Nature Behavior, IPAQ-SF)
- Depression Symptoms
 - SCID
 - HAM-D

5 Pre-CBT Condition Characteristics

In this section, the following data summaries are described:

- Number and percent of expected sessions attended
- Average heart rate and RPE

For those in the exercise condition

- Time on bike

- Average RPM and Load (Watts)

6 Safety Analyses

Safety analyses will include the following information about safety outcomes:

- Detailed listing of unanticipated problems and serious adverse events
- Longitudinal summaries of PHQ-9 Total Score, including incident suicidal and self-injury behavior

7 Efficacy Analyses

7.1 Primary Outcomes

The primary outcomes for the R61 phase of this study are WAI and BADS. WAI is assessed after each CBT session 1 through 8 and reflects that particular session. BADS is assessed prior to each CBT session, at the intake visit, and at the final visit immediately following completion of therapy. BADS reflects the previous week such that the intake and session 1 assessments are used to establish a baseline. The primary estimand for WAI will be the mean WAI across sessions 1 through 8 with the difference between the active and calm conditions serving as the primary comparative estimand. These will be on-treatment estimands that include all persons assigned to the corresponding condition who attend the corresponding CBT session. In particular, all available data from participants will be included, but data from participants who withdraw and stop attending CBT sessions will not be imputed. The primary estimand for BADS will be defined similarly but reflect the mean change from baseline (assessed at session 1) across the assessments at sessions 2 through 8 and the final visit. The difference in mean WAI and change in BADS will be standardized by their respective pooled standard deviation to form a standardized effect measure. These effects will be estimated using a generalized estimating equation (GEE) assuming an AR(1) working correlation. The linear predictor will include intervention by session number effects as a factor. The GEE for change in BADS will also adjust for the baseline BADS value.

As a sensitivity analysis for both WAI and BADS, we may adjust for baseline HAM-D and additional predictive patient characteristics collected at baseline with observed imbalances across groups (e.g., $p < 0.2$). We may also adjust for therapist. As a sensitivity analysis, we will multiply impute missing WAI and BADS outcomes and estimate the primary estimands using Rubin's rules. In addition, as an exploratory analysis we will include the BADS follow-up assessments in the GEE to estimate the longer-term effects on change in BADS.

Go/No-Go Criteria: The 'Go criteria' to move from the R61 phase to the R33 phase of this study requires 80% of participants to attend ≥ 6 of 8 weekly sessions and an average standardized effect size (d) ≥ 0.55 for *either* WAI or change in BADS; OR both standardized effect sizes ≥ 0.35 . Standardized effects of $d=0.55$ and 0.35 are respectively anticipated to reflect 13 and 9 points for change in BADS, and 7 and 5 points for WAI.

7.2 Secondary Outcomes

Secondary outcomes include the exercise priming measures DARS, VAS, and Serum BDNF. DARS and VAS are measured at the intake visit, each CBT session, and each follow-up visit. At each CBT session, these are measured pre-condition, post-condition, and post-CBT. Serum BDNF is measured similarly at CBT sessions 1, 4 and 8. DARS, VAS and Serum

BDNF will be evaluated using repeated measures correlation analyses of change in pre-versus post-exercise. Specifically, GEEs will be implemented to facilitate verifying and evaluating the extent to which moderate exercise causes acute effects on serum BDNF and state anhedonia. Exploratory analyses will also include post-CBT and follow-up assessments. Mediation analyses will also be carried out to assess the extent to which changes in the exercise priming measures mediate the effect of exercise on WAI and change in BADS. We will evaluate their mediating effects on WAI and change in BADS by estimating the direct and indirect effects using GEEs that adjust for the exercise priming measures.

Lyssn session metrics (collaboration and empathy) will be evaluated similarly to the primary outcomes. Each session will be recorded, and Lyssn will generate session-by-session metrics for collaboration and empathy with the primary statistical tests comparing mean and time point-specific session ratings between the two groups.

Secondary clinical outcomes include HAM-D, SCID, PHQ-9, and C-SSRS. These measures are collected at intake, and follow-up visits, with PHQ-9 and C-SSRS, also collected at each CBT session. These outcomes will be analyzed similarly to change in BADS using a GEE. All other secondary measures collected longitudinally will be analyzed similarly.

In general, longitudinally measured outcomes will be analyzed to estimate the condition-specific trajectories and the differences between the two condition trajectories. Spaghetti plots will be used to visualize the raw data, and line plots with 95% pointwise confidence intervals will be used to visualize the condition-specific trajectories and their differences.

7.3 Subgroup Analyses

Subgroup analyses will aim to assess for consistency of a condition effect. Any subgroup-specific effect estimates will be interpreted cautiously due to multiple comparisons and low within-subgroup power. Subgroup analyses of primary outcomes and select secondary outcomes will be performed by the following baseline factors:

- Sex at birth
- Race / Ethnicity
- Current tobacco use
- Baseline nature behavior
- Baseline physical activity (IPAQ)
- Baseline depression symptom severity (HAM-D; mild vs. moderate-to-severe)

As an exploratory analysis, we will also conduct a subgroup analysis based on number of CBT sessions completed. For each subgroup analysis, we will estimate the subgroup specific condition effect and 95% confidence interval, and report the p-value for the interaction effect (i.e., a statistical test for condition effect heterogeneity across the subgrouping variable).

8 Interim Monitoring

The DSMB will be asked to recommend early termination or modification only when there is clear and substantial evidence of harm. At the discretion of the DSMB, the frequency and content of safety reports will be determined and may be modified throughout the trial. The DSMB may also request additional data summaries.

9 Sample Size and Power

The R61 sample size of 40 participants is chosen to ensure high probability for achieving the Go criteria for target engagement under the expected effects of $d \geq 0.50$ and a low probability for achieving the Go criteria when $d = 0$. These analyses conservatively assume 32 of the 40 participants provide complete data (80% retention), though we expect to retain more than 32 participants (given >90% retention in the pilot) and we will incorporate partial endpoint data in these analyses from participants who withdraw.

Table 1: Probability of Meeting Proposed Criteria				
True Effect (d) WAI BADS		Correlation between WAI and BADS		
		0.1	0.4	0.7
0.00	0.00	0.004	0.007	0.012
0.50	0.00	0.395	0.390	0.389
0.50	0.50	0.759	0.737	0.742
0.75	0.00	0.872	0.871	0.871
0.75	0.75	0.992	0.985	0.982

The probability of achieving the Go criteria of observing $d \geq 0.35$ on both endpoints or $d \geq 0.55$ on either endpoint under different true effects and correlations is in **Table 1**. As intended, when true target engagement is absent on both endpoints ($d=0$) the Go criteria are unlikely to be achieved (2% chance or lower), and conversely when true target engagement is moderate on both endpoints ($d \geq 0.5$) or strong ($d \geq 0.75$) on one endpoint the Go criteria are very likely to be achieved (74% chance or higher).

With 32 participants contributing 8 paired measurements for DARS and 3 for Serum BDNF, these data provide 80% power to detect a correlation of 0.19 and 0.34 between DARS and either WAI or change in BADS. WAI and change in the BADS are each predictive of overall treatment success, indicating a moderate effect size difference in WAI and change in BADS across time as clinically relevant. A 1-point WAI change corresponds to a 0.5-to-1-point change in HAM-D, indicating that a 6-point difference in WAI would be associated with 3-6 points on the HAM-D, indicating we are putatively targeting a clinically meaningful effect.

10 Data Completeness and Study Conduct

Data completeness and study conduct reports will be provided by randomization group (for the closed report) and pooled across randomization groups (for the open report).

The following data summaries will be provided to assess data completeness and study conduct:

- Number and percent of participants with protocol deviations, and type of protocol deviation
- Expected and observed number (% of expected) of participants who completed visits.
- Length of follow-up: Median, IQR, range
- Number and percent of participants who withdrew consent or were (potentially) lost to follow-up.
- Listing of participants who withdrew consent, date of withdrawal, and reason for withdrawal.

11 Unblinding of Treatment Comparisons

Trial results will be unblinded when all participants have completed follow-up. Upon the recommendation of the DSMB, if the sponsor and study leadership concur, results may be unblinded earlier.

While the trial is ongoing, unless noted otherwise, access to any data summaries by randomization group will be restricted to the members of the DSMB and the unblinded statistician. Both the DSMB and statistician will remain masked to the randomization group encoding until the DSMB requests otherwise. At the completion of each phase of the trial, the study investigators, DSMB and statisticians will be unblinded to the data from that trial phase.

Table 1. Timing of Assessments Across the 8-week Intervention and 52-Week Study Involvement

	Intake Visit	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8	Final Assessment	20 Week Follow Up	52 Week Follow Up
ENROLLMENT												
Informed Consent	X											
Eligibility Check	X											
Randomization		X										
INTERVENTION												
Active CBT		X	X	X	X	X	X	X	X			
Calm CBT		X	X	X	X	X	X	X	X			
ASSESSMENTS												
Primary Outcomes												
WAI		X	X	X	X	X	X	X	X			
BADS	X	X	X	X	X	X	X	X	X	X	X	X
Lyssn		X	X	X	X	X	X	X	X			
Secondary Outcomes												
State Anhedonia	X	X*	X*	X*	X*	X*	X*	X*	X*	X	X	X
BDNF		X*			X*				X*			
Clinical Outcomes												
HAM-D	X									X	X	X
SCID	X									X	X	X
PHQ-9	X	X	X	X	X	X	X	X	X	X	X	X
C-SSRS	X	X	X	X	X	X	X	X	X	X	X	X
Other Secondary Outcomes												
Demographics	X											
Health History	X											
Treatment List	X									X	X	X
Anthropometric Assessments	X											
ACES	X											
DSM5	X											
WHODAS	X											
IPAQ	X				X					X	X	X
activPAL	X				X				X			
HRQOL	X				X					X	X	X
GAD-7	X				X					X	X	X
ATQ	X	X	X	X	X	X	X	X	X	X	X	X
SEQ		X	X	X	X	X	X	X	X		X	X
NCI	X	X*	X*	X*	X*				X*			
Nature Behavior	X	X*	X*	X*	X*				X*			
Satisfaction Survey										X		
Notes: Anthropometric assessments = height, weight, heart rate, blood pressure; * Measures taken 3 times (pre-condition, post-condition, and after therapy); activPAL issued at the visit and worn for 7 days before being returned at the next visit or provided 7 days before Final Assessment (indicated by x ———).												