STUDY STATISTICAL ANALYSIS PLAN

Digital Health Platform (DHP) to Deliver Mindfulness as a Stress Management Intervention Leveraging Electronic (SMILE) Health Records for Racial and Ethnic Populations During the COVID-19 Pandemic: Clinical Trial

NCT number NCT06242080
Document Date 01/23/2024

STATISTICAL CONSIDERATIONS

Statistical Hypotheses

Brief Summary:
The goal of this clinical trial is to evaluate the SMILE app, a Digital Health Platform (DHP), that will deliver a mindfulness intervention, designed to mitigate COVID related stress. Additionally, the SMILE app will remotely collect self-reported psychological and physiological metrics of mental health and autonomic regulation. Study participants are adults who self-identify as African American, Black and/or Latino, and who have clinically significant levels of anxiety.

The study aims are:

- Aim 1: Establish the effectiveness and durability of an 8-week Mindfulness DHP intervention. The investigators will focus on two constructs important to mental health and hypothesize that: A) Anxiety, self-report stress and quality-of-life measures will significantly improve when comparing: A.1) Pre-to-post intervention, and; A.2) Control vs. intervention groups over 8 weeks and at 1-month follow-up. B) Arousal, autonomic indices of HRV (reflecting parasympathetic activation) will significantly improve, when comparing: B.1) Pre-to-post intervention, and; B.2) Control vs. intervention groups over 8 weeks and at 1-month follow-up.

- Aim 2: Establish the sustainability of two Mindfulness DHP interventions utilizing retention, usage (frequency), and participant satisfaction. The investigators hypothesize that MTIA will be more sustainable than MAPP because the interactive, instructor-guided group model will create an environment of community and accountability.

- Aim 3: Exploratory: Examine associations between COVID-19 related stress, mental health outcomes, and HRV. Examine the extent to which COVID-19 related stress and mental health symptoms are linked to HRV at baseline and how that relationship changes over time.

Participants will be assigned to 1 of 3 arms of the study: MTIA intervention, MAPP intervention, or wait-list control. All participants will be mailed a device with the SMILE app installed, and equipment for recording cardiac data in the home. All participants will complete the baseline psychometrics measures and physiological stress test using the instructions provided on the SMILE app. Those assigned to the MTIA or MAPP intervention groups will then participate in their assigned intervention over the subsequent 8 weeks. During these 8 weeks, psychometric and physiological data will be completed biweekly for all participants. 3 months following the initial baseline, all participants will complete a final psychometric/physiological evaluation.

- Primary Endpoint(s): The primary endpoint is anxiety as measured by the GAD-7 scale, a commonly used clinical anxiety measure. We hypothesize that both intervention groups will show significant improvements compared to baseline at 8 weeks and 12 weeks. We further hypothesize that, compared to the wait-list control group, the mean anxiety for participants in the MTIA group and the MAPP intervention groups will show a significantly greater

reduction in anxiety at the end of the intervention and at 12 weeks. Our null hypothesis is that neither intervention group will show a significant reduction in anxiety either compared to baseline or compared to the control group.

- Secondary Endpoint(s): For each of the secondary endpoints listed in Section 3.0 (See table), we hypothesize that, compared to the control group, the mean (or median value as appropriate) for each measure will show improvement at the end of the intervention period.

Sample Size Determination

Power analysis was primarily informed by psychological variables (Aim 1.A), which are more common outcomes in published studies, then expected power was applied to the HRV outcome (Aim 1.B). To conduct the power analysis for Aim 1.A, recent studies of app-based mindfulness mediation were surveyed for effect sizes for treatment effects on stress, irritability, and psychological wellbeing measures. Effect size were calculated and adjusted for correlations of pre and post scores where possible. Between-group standardized mean differences varied depending on outcome, but small-to-medium effect size of Cohen's d =. was observed for outcomes compared to psychoeducation, waitlist, and assessment-only control groups(32,52). Aiming for 80% power, an alpha level of 0.05, and expected effect size of between subjects Cohen's d=0.4, we calculated that 100 participants will be needed per condition at the final time point to detect an effect of this size using a 2-sided independent-samples t-test. We expected that there would be a correlation among outcomes for participants in the same class within the MTIA intervention arm, as observed in our previous study (ICC=0.021), which can influence power to detect effects between groups(53). Using that intra-class correlation estimate and an average of 9 subjects per group, we calculated that we would need a sample of 117 subjects at the final time point for the MTIA arm to meet the previously defined power criteria for Aim 1.A. For Aim 1.B, the target sample size would also provide more than 97% power to detect HRV changes equal to those found in a prior study (Cohen's d = .55138); even with expected 10% physiological data loss due to mechanical error or signal noise as is common in physiological studies. Given an expected pre-post retention rate of 85% for MTIA group and 75% for MAPP and WLC groups, we estimate we would require recruitment of 138 participants in the MTIA group and 133 participants in each of the MAPP and WLC groups. We do not expect to have 80% power for each of the exploratory analyses.

Populations for Analyses

Our primary analysis will consist of a modified intention-to-treat analysis that will retain all participants randomized to the intervention groups who attended at least one intervention session (MTIA) or interacted with the MAPP at least once. In addition, among all randomized participants, this includes individuals with at least one post-baseline assessment. Our justification for this approach is that it is common when one or more arms include an individually randomized group because participants may not be available to attend the intervention after waiting for the cohort to begin.

We will also conduct an analysis using all randomized participants (standard Intention-to-treat) and a sensitivity analysis including only participants who attend 5 or more MTIA sessions, access the MAPP on 5 weeks, and complete the 8-week and 12-week assessments (per-protocol population).

Statistical Analyses

General Approach

The analysis will include descriptive statistics of the study population consisting of means and standard deviations for continuous variables and percentages for categorical variables. Variables will be checked for normality. Transformations common for the variable will be employed as indicated. All analyses will control for age, gender (woman vs. not), ethnicity (Hispanic/Latino vs. not), and baseline anxiety. Control for age and gender is indicated due to their strong associations with the endpoints. Analyses will use two-tailed tests with a p-value of 0.05 considered statistically significant.

Analysis of the Primary Endpoint(s)

The primary outcome, The GAD-7 scale, is measured at baseline, 2 weeks, 4 weeks, 6 weeks, 8 weeks and 12 weeks. The analysis will be conducted using mixed effects models. Modeling will examine within-group changes in outcome variables between intervention groups. The main effect of interest will be treatment (group) X time interaction effects. Within-individual correlations will be modeled using random intercepts. The addition of random slopes will be tested. Model comparisons will be conducted using likelihood ratio tests for nested models and Akaike and Bayes Information Criterion (AIC and BIC) for non-nested models. In the MTIA condition, random intercepts will be used to adjust for cohort differences. Fixed effects will include gender (woman vs. not), ethnicity group (Hispanic/Latino or not), and age.

Treatment X time differences will be assessed using an omnibus test (F-test) with a null hypothesis that there is no treatment X time effect among the three treatment arms. If the difference of treatment effect is significant at the 0.05 level, we will then examine individual group contrasts with the control group as the reference category.

The primary analysis, using the modified intention-to-treat, uses mixed effects models that are robust to missing data.

Analysis of the Secondary Endpoint(s)

Analysis of the psychological endpoints relies on scales and will be assessed using the same procedure as described above for the GAD-7. For variables with fewer assessment time points, a multiple regression model may be employed. A complete Statistical Analysis Plan will be developed prior to the database lock.

For HRV analyses] and mean heart period [HP]), we will include within-timepoint observations at the 15-second epoch level, with labels for each observation along three dimensions (time, visit, and subject). Multiple-level modeling will be used to test for differences in the dynamic behavior of HP and High frequency HRV (HF_HRV) across time. In the case that the variability of either HP or HF_HRV is too great at the 15-second epoch level, we will apply a triangular smoothing algorithm with a span of 3-5 epochs to stabilize the estimates of both HP and HF_HRV. Vagal efficiency (VE)

will be calculated over all epochs spanning the posture conditions. HRV analyses will include fixed effects for BMI and time-varying covariates for medication use, caffeine, tobacco, alcohol, and time of day.

The Aim 2 sustainability variables include1) retention (withdrew vs. not, measured by proportions); 2) usage (measured by a) attendance and b) homework completion in the MTIA group and by a) app interaction—opening pages—b) and homework completion in the MAPP group); and 3) satisfaction with the intervention as measured by a Likert scale question. Retention is analyzed with a chi square statistic (or logistic regression). The analysis of usage will depend on the distributions. Prior to the database lock, we will examine usage distributions to determine if a t-test is possible or if a different type of analysis (e.g., Poisson model) will be needed. Satisfaction will be compared with a t-test or nonparametric test, as indicated.

Safety Analyses

N/A

Baseline Descriptive Statistics

Baseline characteristics will be calculated for each intervention group. These include demographic variables and baseline levels for the study endpoints. In addition, we will examine groups defined by their adherence to the protocol: participants who withdraw after attending at least one class or interacting with the app at least once compared to those who have 2 or more interactions with the intervention. Also, we will compare characteristics of completers and non-completers.

Planned Interim Analyses

NA

Sub-Group Analyses

Subgroup analyses will depend on the distribution of the overall sample by age, gender, and ethnicity. Subgroup analyses will be addressed in the formal SAP.

Tabulation of Individual Participant Data

Individual participant data will be listed by measure and time point.

Exploratory Analyses

A goal of the study is to understand the relationships between the psychological and quality-of-life variables and the heart rate variability assessments. We will examine the correlations among the variables at baseline and over time. Models under consideration include structural equation models that can consider the correlations among the psychological variables.