Statistical Analysis Plan

WASABI: Wiring Adolescents with Social Anxiety Via Behavioral Interventions

Protocol Number: PSC-1016-19

NCT03990870

Document Date: July 8, 2019

(1) ANALYSIS POPULATIONS

There are three *a priori* defined analysis populations, including a primary analysis population (i), a secondary analysis population designed to compare effect sizes in populations with no missing data (ii) and a population who completed all training visits (iii)

i. *Intent to Treat (ITT) population:* This is the *a priori* primary analysis population, defined as including all randomized participants who attended at least one group video call.
ii. *Intent to Treat (ITT) Fully-Evaluable (FE) population:* This is a secondary analysis population, defined as including all members of the ITT population that complete a post-intervention visit. Note that a participant may complete a specific visit but have missing data for a test in which case the participant is in the overall FE population but does not contribute data to the FE population for that visit, e.g., the number of evaluable cases for a specific test on a specific visit may be smaller than the FE population for that visit because of missing data.
iii. *Intent to Treat (ITT) Completers (C) population:* This is a secondary analysis population, defined as including all members of the ITT-FE population who complete all intervention sessions. Note that the C populations are strict subsets of the FE populations; a person who completes the target number of videocalls but does not complete the evaluation visit is not a member of the C population.

(2) STATISTICAL PLAN

Feasibility

The Phase I data analysis plan *a priori* defines a primary intent-to-treat (ITT) population, a set of secondary evaluation populations, a set of primary outcome measures, a set of secondary outcome measures, a set of potential mediators/moderators, a single primary evaluation time point, a criterion for statistical significance, and a statistical analysis methodology for secondary outcomes.

- The primary ITT population is defined as all participants who attend one *dCBGT* video call. Note that this includes all randomized participants except those who drop/withdraw post-randomization and pre-training.

- As the main goal of this Phase I trial is to evaluate the feasibility, acceptability and usability of the WASABI app. To this purpose, we will conduct an analysis of the following primary outcome measures in participants randomized to the WASABI+dCBGT arm:

  1. Assessment adherence (percentage of assessment sessions attended). Efforts will be made to assess all participants who have completed the minimum required intervention activities (attended at least 2 group calls, completed EMAs at least five times). *Based on our previous studies (CLIMB and MOODIFY) and pilot, we expect successful data collection of baseline and post-intervention outcome measures for ≥90% of participants.*

  2. Engagement with the digitally delivered Cognitive Behavioral Group Therapy (*dCBGT*) and with group Instant Messaging (IM): number of group sessions attended, number of messages sent per week on the group chat. *Based on our previous studies (CLIMB),* we *hypothesize that participants randomized to WASABI+ dCBGT will participate in 80% of the* dCBGT *sessions, and send at least 8 messages/week in the group chat*.

  3. EMA completion rate. *Based on our previous studies and pilot, we hypothesized that participants randomized to WASABI+ dCBGT would respond to 70% of the EMA prompts.*

  4. SPAI-B completion rate. *Based on our previous studies and pilot, we hypothesized that participants randomized to WASABI+ dCBGT would complete the weekly SPAI-B questionnaire 70% of the time.*

5. Overall program completion rate. *Based on findings from our previous findings (CLIMB) and from the pilot in anxious youth, we hypothesized full program completion in ≥79% (≥8 of 12) study participants in the treatment arm.*

6. Usability ratings were obtained from all study participants post-intervention via a 7-point Likert-scale exit survey. This is a brief and embedded post-study questionnaire on program satisfaction, clarity, enjoyment, perceived benefits, and ease of fit into schedule. Participants rate each sentence on the following 7-point Likert scale: 1 = Completely Agree; 2 = Mostly Agree; 3 = Somewhat Agree; 4 = Undecided; 5 = Somewhat Disagree; 6 = Mostly Disagree; 7 = Completely Disagree. Clinical trials run by Posit Science software using this survey have shown high internal consistency with a Chronbach's alpha of .917. *Based on our previous findings, we hypothesize exit survey ratings of at least* ≥4.5 ±1.5 *on the 7-point Likert scale items.*

7. Reported side effects (raw score). *Based on our previous findings, we expected 0 adverse events due to program use).*

The secondary outcome measures were collected by trained personnel blind to treatment assignment at baseline and after the intervention:

- clinician ratings of anxiety and social anxiety symptoms through the Liebowitz Social Anxiety Scale for Children and Adolescents (LSAS-CA
- self-report inventories of social anxiety symptoms through the Social Phobia and Anxiety Inventory-Brief (SPAI-B)– at baseline, weekly during the intervention, and after the intervention;
- a broad measure anxiety and potential generalization of treatment to other anxiety domains, the Multidimensional Anxiety Scale for Children 2nd Edition (MASC-2);
- social skills by administering the Social Skills Questionnaire (SSQ) ;
- social functioning using the Adolescent Social Self-Efficacy Scale (ASSES).

The criterion for statistical significance is $p < 0.05$. Results with $p < 0.1$ will be described as trends.

The primary analysis time point is the post-intervention assessment.

For feasibility, we relied on descriptive statistics. For preliminary efficacy, in smaller sample sizes like ours, there is a greater likelihood that there will be larger differences at baseline between groups simply due to sampling error. Groups were compared on all baseline measures using Mann-Whitney U tests (for ordinal and continuous variables) or Pearson chi-square/Fisher's exact test (for categorical variables). To test whether groups differed in change over time, data were analyzed via linear mixed-effects models. Mixed-effects models are ideal for the analysis of longitudinal data, due to the fact that it tolerates missing observations–assuming that it is missing at random (MAR). First, simple growth models (change from baseline to post) were conducted to examine change over time (for all participants). Because complete case analyses have been shown to provide biased estimates in the presence of missing data (Enders, 2010), data were then modeled using the intent-to-treat (ITT) framework, in which all randomized subjects are included in the analysis, and missing data were handled via Full Information Maximum Likelihood (FIML). This method for handling missing data is considered a gold standard approach, as it gives relatively unbiased estimates (Enders, 2010), and is the standard for the analysis of randomized trials. For all models, fixed effects of time, group, and group x time interaction were included; a random intercept of subject was used. Kenward-Roger degrees of freedom were used due to the small sample size; statistical significance was evaluated at a two-sided alpha of $p = 0.05$; however, null hypothesis testing was de-emphasized given the small sample size of the control group (n = 8). We emphasize the effect size (Cohen's *d*) instead. Although various estimates of effect size exist for LMM, we chose a simplified version, defined as:

$$\frac{(\Delta_{(dCBT+WASABI)} - \Delta_{(dCBT)})}{\sigma_{pre}}$$

Where Δ refers to slope (i.e., change from baseline to post-intervention) and σ_{pre} refers to the pooled baseline standard deviation. Thus, values of *d* refer to change in standard deviation units from baseline.