Statistical Analysis Plan

Study Title: Evaluation of an Asynchronous Remote Communities Approach to Behavioral Activation for Depressed Adolescents

Version Date: 1/29/2024

Clinician Feasibility, Acceptability, Appropriateness, System Usability, & Burden (Aim 1): **Ha1**: BA+ARC clinicians will report better feasibility, acceptability, appropriateness, system usability, and less burden than clinicians in BA-Only. Due to the small sample size of clinicians (n = 10) and limited statistical power for significance testing, we will compute Cohen's d effect sizes and standard deviations for differences between conditions on acceptability, feasibility, usability, and burden. Cohen's d > .5 in favor of BA+ARC will be deemed supportive of the hypothesis. Small (d < .3) or negative effect sizes will be deemed indicative of a need for modifications of BA+ARC prior to a larger trial in order to improve these implementation outcomes. Modifications will be based on a qualitative review of usability problems and suggestions for improvement similar to analyses described in Aim 1.

Patient Feasibility, Acceptability, Appropriateness, System Usability, & Burden (Aim 2a), Target Mechanisms (Aim 2b) and Patient Outcomes (Aim 2c): **Ha**: BA+ARC adolescents will report better feasibility, acceptability, appropriateness, system usability, and less burden than clinicians in BA-Only. **Hb:** BA+ARC adolescents will demonstrate improvements in target mechanisms including working alliance, treatment timeliness and efficiency, social belongingness, mental health stigma, and engagement in care (i.e., greater homework completion, session attendance, and BA skill use) compared to BA-Only adolescents. **Hc:** Adolescents treated by BA+ARC clinicians will have improved self-reported and caregiver-reported outcomes than adolescents treated by BA-Only, including reduced depression symptoms, diagnosis and functional impairments. Basic data screening and descriptives will be conducted, including screening for errors, missing data, and distributions and time trends. We will test for between-group differences at initial timepoint for all demographic and outcome measures; any differences greater than d > .3 will be statistically controlled for by adding these variables as covariates in the model building below, along with psychotropic medication status. Missing data will be addressed via full information maximum likelihood estimation (FIML) in mixed effects model building. We employ the Benjamini-Hochberg false discovery rate[93] to avoid familywise error. Generalized mixed effects models, to account for partial nesting of patient within clinician within ARC will be used to test our primary aims. Random terms will be included as appropriate and feasible, generally in preparation for a larger trial, as the relatively small sample size is likely to forbid model convergence without Hessian errors. Regardless, via this analytic approach we will obtain intraclass correlation estimates that will permit a priori sample size determination for a larger trial. We will use link functions, as appropriate to distribution, to model skewed, zero-inflated, dichotomous, or other non-normal distributions. Because change may occur differently for the intervention and sustainment timeframes (e.g., initial measure to 12-week end of treatment, and from 12-week to 24-week follow-up), we will compute these as piecewise models, testing different rates of change during these times. Experimental effects will be tested in the following ways. Longitudinal outcomes will be tested by adding condition and a condition x time interaction terms for the piecewise time periods. We will center time at week 12 in order to test between-group differences at the end of treatment. Models will be selected by comparing the chi square, Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC) deviance of nested models, with smaller values of chi square, AIC, and BIC indicating better model fit. To test **Ha2** for patient-reported implementation outcomes, we will apply these models to longitudinal patient-reported data on feasibility, acceptability, appropriateness, usability, and burden of the overall experience. To test **Hb**, we will apply these models to the longitudinal outcomes provided by clinicians, patients, and/or caregivers, including working alliance[87], timeliness and efficiency[90], social belongingness, stigma, daily homework completion and BA skill use, and weekly session attendance. To test **Hc,** partially nested longitudinal mixed effects models as described above will test the outcomes of depression symptoms, any diagnosis on the K-SADS, and functional impairment, as reported by patients and caregivers.

Exploratory Aims: We will conduct preliminary analyses to determine 1) whether patient outcomes are mediated by measures collected via *Hb,* and 2) which BA+ARC features are linked to improved outcomes.

A traditional Path Analysis[94] extended to multilevel framework[95] will be used to identify mediation effects for all variables with an effect size of $d > .30$ in analyses conducted in Hb and Hc. Due to small sample size, these analyses will be considered exploratory and informative for a larger trial.

C2.6 Power. Aim 1: HCDE research suggests that 5 users sufficiently identifies >85% of usability problems[75]. Aim 2*:* The minimum detectable effect size for all patient outcomes in Aims 2a2, b, and c is $d = .623$. As appropriate for this pilot, exploratory study, we assumed 10 clinicians and 70 clients evenly assigned to 2 conditions, a clinician-level ICC of .05 for client outcomes[96], a one-sided test, and an alpha of .10.

C2.7. Implications of hypothesis testing: Positive findings will set the stage for larger scale efforts to more formally study a number of areas: whether ARC can be used across various EBPIs to improve adolescent mental health; determine if EBPI+ARC is superior to EBPI-Only; ability of successful target engagement to mediate subsequent reduction in depressive symptoms; whether EBPI+ARC can increase access to care in low-resourced communities; and whether we can successfully use EBPI+ARC in non-English speaking populations. Positive findings will also provide further support for mental health systems to invest effort in utilizing technology platforms that would facilitate ARC delivered EBPIs.