Title: Acceptability and Efficacy of Zemedy App Versus Education and Relaxation Training App for IBS

NCT# 04665271

Document Date: 08-09-2024

Statistical Analysis Plan

Analyses: All analyses were carried out in SPSS V29.0.1.1. First, independent sample ttests were used to assess for differences in demographic and clinical variables between those who downloaded their assigned app and enrolled in the trial and those who did not, as well as between the two enrolled groups at baseline, and to compare the baseline scores of treatment completers versus individuals who dropped out of the study and failed to supply follow-up data at post-treatment.

Next, change over time both within group and between groups was examined. Analyses were completed using both per protocol (completer) and intention-to-treat (ITT) (sensitivity analysis) approaches using multiple imputation to estimate missing data. Per protocol analyses only included individuals who completed the post-treatment follow-up questionnaires. ITT (which accounts for that missing data) is generally considered the gold standard in clinical trials, especially longitudinal randomized controlled trial with missing data due to attrition [48], although it has also been criticized for being overly conservative, and for confounding the issues of acceptability and efficacy [49]. We conducted sensitivity analyses using multiple imputation to account for missing data at post-test, the only time point at which between group comparisons were possible. Because the data were missing not at random (MNAR) we used a monotone

imputation with linear regression and 5 iterations. We then analyzed the pooled data using multivariate and univariate analysis of covariance, as well as within group paired sample t-tests.

Initially, paired sample t-tests were used to examine within group change from baseline to immediate post-treatment for each group and to examine maintenance of gains from post CBT treatment to 3- and 6-months follow-up. Subsequently, multivariate and univariate general linear models were used to examine between group effects at post treatment (8 weeks), controlling for baseline levels of the dependent variables.

With respect to between group differences, we first conducted a conservative MANCOVA predicting all outcome measures (dependent variables GSRS, IBSQoL, VSI, GICog, FFQ, BDI) from the independent variable condition, controlling for all of those same measures at baseline. We then ran a series of ANCOVAs predicting each of the outcome measures individually at post-treatment from condition controlling for that measure's baseline value.

With respect to clinically significant change, we used Jacobsen and Truax's Criterion B (falling within 2 SD of the healthy mean), which is more conservative than Criterion A (falling 2 SD below the pathological mean) [50]. With respect to effect sizes, we report on partial eta squared (ηp^2), Cohen's *d,* and the pooled pretest standard deviation for weighting the differences of the pre-post-means (d_{ppc2}) [51]. The d_{ppc2} measure is designed for pre-post controlled trials where the sample sizes may be uneven across groups, as was the case in our data set.