Testing a Digital Intervention for Adolescents
NCT 04567888

Statistical Analysis Plan (SAP)

Version: 2.0

[redacted]

Date: Sept 14, 2022

CONTENTS

1 INTRODUCTION

The following document outlines a statistical analysis plan (SAP) to support the analysis of data from the Happify for Teens RCT study (NCT04567888). These analyses will be used to support the writing of a manuscript and will likely lead to future conference submissions. The protocol for the RCT has been published previously (see Boucher et al., 2021). The overarching aim of this study is to test the effects of the Happify app on teens' perceived stress and rumination.

2 DATA SOURCE

The dataset consists of teen participants in the Teens RCT study. Assessment data was pulled from Qualtrics by [redacted], the research coordinator in charge of the Teens RCT at the time in March/April 2022. Data from each assessment were pulled separately, cleaned and then merged together in SPSS by [redacted], Senior Director of Research Strategy and PI on the Teens RCT. Happify usage data was pulled from Twill servers by [redacted] in July 2022 using user IDs and emails from RCT participants assigned to the Happify for Teens condition, and then merged with the larger dataset by [redacted] to prepare for the final analysis.

3 ANALYSIS OBJECTIVES

The primary objective of this study is as follows:

- Comparing the Happify for Teens condition and the waitlist control groups, examine changes in perceived stress and rumination (brooding) across baseline, 4-weeks, 8-weeks (post-intervention), and 12-weeks.

The secondary objective of this study is as follows:

- Comparing the Happify for Teens condition and the waitlist control groups, examine changes in loneliness, optimism, and sleep disturbance across baseline, 4-weeks, 8-weeks (post-intervention), and 12-weeks.

The exploratory objectives of this study are follows:

- Determine the extent to which brooding mediates the relation between treatment condition (Happify for Teens vs. control) and perceived stress.
- Among those in the Happify for Teens treatment, we will explore the extent to which app usage (i.e., number of activities completed, number of active days) associates with each of the five outcome variables from the primary and secondary objectives.

4 ANALYSIS SETS/POPULATIONS/SUBGROUPS

To be eligible for this study, the following criteria needed to be met:

- range in age from 13-17 years of age
- perceived stress scale (PSS) score greater than 14
- rumination-brooding subscale score of at least 10

We will create a subgroup of the treatment condition, stratified by those who did and did not average two activities completed per week. This threshold has been established in prior studies of the Happify app as a minimum needed for adults to benefit significantly from the program (e.g., Parks et al., 2018). We will examine descriptive statistics for these subgroups.

5 ENDPOINTS AND COVARIATES

The primary criterion of interest will be perceived stress scores, self-reported at baseline, 4-week midpoint, 8-week post, and 12-week follow-up. The secondary criterion of interest will be brooding scores at the same time points. Additionally, three exploratory outcomes will be examined at the same time points: optimism, sleep disturbance, and loneliness. The effect of the Happify treatment will be examined for each outcome.

In each model, we will covary the effects of participant age, gender, and race. Gender will be coded into two binary variables (0 = man or other, 1 = woman; 0 = woman or man, 1 = other). Race will be coded into a single binary variable (0 = White, 1 = all other races). Given that individuals who identify as White comprise almost 75% of the sample, this coding scheme prevents saturating each model with a large number of dummy codes while still preserving much of the racial distribution of the sample.

6 HANDLING OF MISSING DATA AND OTHER DATA CONVENTIONS

Missingness will be evaluated using standard procedures when analyzing repeated measures data (van Buuren, 2018). First, missingness codes will be created that correspond with the variables in which data points are missing. A series of variables will be created that indicate missingness in each outcome at each measurement time point. We will then explore the extent to which various demographic markers and other variables in the dataset associate with missingness through a series of correlation tests. This practice will ultimately help us determine auxiliary variables for our models.

When working with missing data, there are two “gold standard” approaches to mitigating bias: multiple imputation and full information maximum likelihood (ML) estimation. However, as will be detailed below, multilevel modeling benefits from the fact that the approach does not assume that each group or cluster is balanced. In other words, even if individual A only has a baseline, but individual B has all four observations, no assumptions are violated, and the model will be unbiased (Raudenbush & Bryk, 2002). Thus, multilevel modeling allows us to avoid using a traditional missing data approach by simply harnessing the model’s extant capabilities. Given that we do not anticipate any significant missing data in the predictors or covariates, we plan to use restricted maximum likelihood (REML) estimation in the multilevel framework to account for missing outcome variable data. Final models will be run with identified auxiliary variables included as covariates.

7 STATISTICAL METHODS

7.1 INITIAL DATA EXAMINATION

For each variable of primary and secondary interest (i.e., perceived stress, brooding, optimism, sleep disturbance, and loneliness) and at each time point, we will create histograms and boxplots to examine each variable for normality and outliers. We will then calculate descriptive statistics for each criterion (i.e., mean, median, standard deviation, IQR, range) as well as appropriate frequency statistics for demographic covariates. As a sensitivity analysis, we will repeat the primary statistical models after excluding participants who failed at least three attention checks, as outlined in the protocol paper.

7.2 PRIMARY STATISTICAL MODELS AND ASSUMPTIONS

Repeated measures data, particularly with relatively high amounts of attrition (i.e., missing data in the criterion), are effectively analyzed via multilevel modeling (MLM)[1]. MLM has several advantages over traditional approaches, such as repeated measures (rm) ANOVA, the first of which is that rmANOVA assumes constant variance across treatment condition. Secondly, rmANOVA assumes sphericity (the within subject association between individual scores is equivalent across treatment conditions).

The MLM will be computed in R statistical software (R Core Team, 2022) and the *lme4* package (Bates et al., 2015) using REML estimation. Upon running the full model (see below), we will test the key MLM assumptions by examining residuals for linearity, normality, and homogeneity (Raudenbush & Bryk, 2002). We will also examine correlations between random and fixed effects. A significant correlation would violate the assumption of endogeneity, meaning that we would need to consider alternative approaches (McNeish & Kelley, 2019). The full MLM is depicted mathematically as follows (excluding age, gender, and race covariates),

$$\mathrm{PSS}_{\mathrm{ti}} = \gamma_{00} + \gamma_{01}\mathrm{Condition}_{\mathrm{i}} + \gamma_{10}\mathrm{Time}_{\mathrm{ti}} + u_{0\mathrm{i}} + u_{1\mathrm{i}} + r_{\mathrm{ti}}$$

where t denotes repeated observations, i denotes unique individuals, γ_{00} denotes the grand mean of the outcome (fixed intercept), γ_{01} denotes the fixed effect of treatment condition (a Level-2 variable), γ_{10} denotes the fixed effect of time (a Level-1 variable), $u_{0\mathrm{i}}$ denotes a random intercept, $u_{1\mathrm{i}}$ denotes a random slope for time (allowing each individual to have a unique trajectory), and r_{ti} denotes residual error.

The first MLM will include a random intercept and no random slope. We will then compute the model with the random slope included (see the above equation) and conduct a likelihood ratio test using deviance statistics. If the difference in deviance is significant, we will conclude that the model with the random slope fits the data best. We will also create spaghetti plots to depict random effects in each criterion variable.

7.3 ADDITIONAL STATISTICAL MODELS

---

[1] MLM goes by many aliases in the social and behavioral science literature, including but not limited to: hierarchical linear modeling (HLM), linear mixed effects modeling, mixed modeling, and random effects modeling. Though some of these terms denote slight distinctions by subfield, each is more or less the same. I will use multilevel modeling, or MLM, throughout my documentation.

As noted in the published protocol paper, we will also examine brooding as a mediator of the relation between treatment condition and stress. We will harness a path modeling approach. In this path model, we will lag the effect such that treatment associates with 4-week stress, mediated by baseline brooding, and so on (see Figure 1). Path modeling will be completed using the *lavaan* package in R (Rosseel, 2012). Mediation effects will be calculated using bootstrapped 95% confidence intervals.

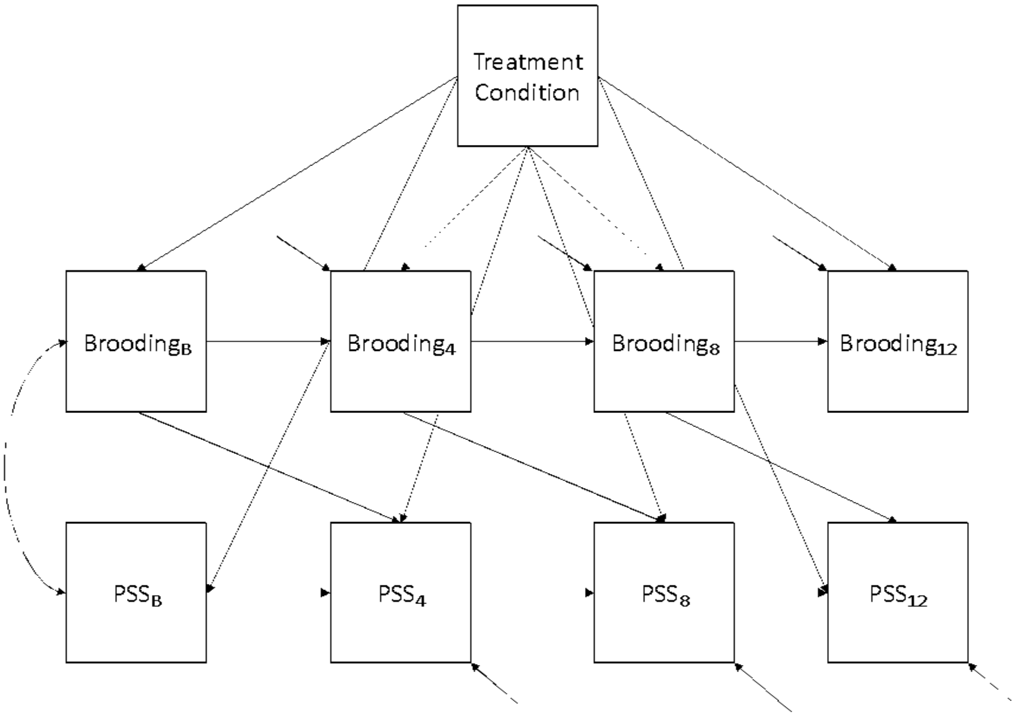

Figure 1. The proposed path model that will allow us to explore the extent to which brooding mediates the relation between treatment and perceived stress (PSS). Subscripts indicate each time point (B = Baseline; 4, 8, 12 = 4-, 8-, and 12-weeks, respectively). Single-headed arrows connecting two variables indicate regression pathways. Two-headed arrows indicate covariance. Single-headed arrows on a single box denote unmeasured endogenous variance.

We will also explore the relation between app usage and each outcome through a series of MLMs. The MLM parameters will be similar to the primary model (described previously). One MLM will be run for each primary and secondary outcome (brooding, perceived stress, loneliness, optimism, sleep disturbance), predicted by app usage. We will define the usage variable as: (Number of activities completed) / (Total number of days active on the app). This approach accounts for how involved participants were with the treatment while controlling for continued usage. We will parse usage into between- and within-person effects (see Carpenter et al., 2016). Between-person usage will be a Level-2 variable, indicating each participant's usage over the entire number of active days in the study, whereas within-person usage will be at Level-1, indicating the amount of app usage (relative to days active) between assessments.

Finally, we plan to create a dose response curve (DRC) model to explore optimal teen usage of the Happify platform. Prior research on optimal usage with Happify has come from adult populations. A dose response curve, calculated using overall app usage and overall change in well-being indices would help Twill gain a better grasp of how the app functions for teens. We will create the DRC using the *drc* package in R. We intend to create a nonlinear DRC in which each of the five outcome variables is regressed onto each participant's total number of activities completed / total number of weeks enrolled in the study.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(i01).

Boucher, E. M., Ward, H. E., Stafford, J. L., & Parks, A. C. (2021). Effects of a digital mental health program on perceived stress in adolescents aged 13-17 Years: Protocol for a randomized controlled trial. *JMIR Research Protocols, 10*(4), e25545.

Carpenter, J., Crutchley, P., Zilca, R. D., Schwartz, H. A., Smith, L. K., Cobb, A. M., & Parks, A. C. (2016). Seeing the "big" picture: Big data methods for exploring relationships between usage, language, and outcome in internet intervention data. *JMIR, 18*(8), e5725.

McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods, 24*(1), 20-35.

Parks, A. C., Williams, A. L., Tugade, M. M., Hokes, K. E., Honomichl, R. D., & Zilca, R. D. (2018). Testing a scalable web and smartphone based intervention to improve depression, anxiety, and resilience: A randomized controlled trial. *International Journal of Wellbeing, 8*(2), 22-67.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Rosseel Y (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

van Buuren, S. (2018). *Flexible imputation of missing data*. CRC press.