Analytical Plan for Efficacy of an ethnic sisterhood group in reducing depression in African American women: randomized controlled trial

DOCUMENT: SAP-2023-035-MB-v01

From: Felipe Figueiredo **To:** Montielle Brandman

2023-12-15

TABLE OF CONTENTS

Analytical Plan for Efficacy of an ethnic sisterhood group in reducing depression in African American women: randomized controlled trial

Document version

Version	Alterations
01	Initial version

1 ABBREVIATIONS

- CI: confidence interval
- SD: standard deviation

2 CONTEXT

2.1 Objectives

To determine the efficacy of a social intervention of an ethnic sisterhood group in reducing depression in African American women.

2.2 Hypotheses

Depression will change at different rates in control and ethnic sisterhood groups.

3 DATA

3.1 Raw data

Upon study start the raw data will be collected in a raw table, that will be processed before analysis. It is recommended that collection is performed in separate tables, for each purpose. Demographic data will be collected only once and can be done with a single instrument. Each row represents all information collected from each study participant, and each participant included will require a unique study ID.

Depression scores and other repeated measures can use the same instrument at each assessment point, provided that the study ID is recorded for each new measurement. Tables can be merged for the analysis by matching on participant ID.

The outcome should be recorded as a continuous variable (see section 4.4). The experimental group (control or intervention) must be recorded as a binary variable (see section 4.3).

All other variables (section 4.5) should be recorded as continuous (eg. number of children) or binary variables, whenever possible. Marital status, religious affiliation, income and others will be recorded as categorical variables. It is recommended that the number of classes be as small as possible, to avoid loss of precision in the analysis estimates, without the need to increase the sample size. Broader categories should be preferred, whenever possible.

3.2 Analytical dataset

After the cleaning and merging processes 12 variables will be included in the analysis. The total number of observations excluded due to incompleteness and exclusion criteria will be reported in the analysis. Table 1 shows the structure of the analytical dataset. N represents the total number of observations, which two or three per individual, depending on the decision for the final design (see section 4.1).

Table 1 *Analytical dataset structure after variable selection and cleaning.*

id	exposure	outcome	assessment	age	marital	income	educ	religion	insurance	occupation	children
1											
2											
3											
…											
N											

All variables in the analytical set were labeled according to the raw data provided and values were labeled according to the data dictionary for the preparation of production-quality results tables and figures.

4 STUDY PARAMETERS

4.1 Study design

This is a randomized controlled trial. All participants will receive daily motivational messages, predetermined in the study protocol (control treatment). In addition to the control treatment they will be randomized to participate in the sisterhood intervention on the weekends.

Depression scores will be assessed at baseline before the intervention and again at 6 weeks when the study period ends. A third intermediary measurement is still being decided on and, if approved, the study sample will be comprised of three observations per individual.

The measure of efficacy is defined as the change in average depression score between time points.

4.2 Inclusion and exclusion criteria

N/A

4.3 Exposures

Sisterhood ethnic group on the weekends.

4.4 Outcomes

Specification of outcome measures (Zarin, 2011):

1. (Domain) Depression
2. (Specific measurement) Beck Depression Inventory-II
3. (Specific metric) Change from baseline
4. (Method of aggregation) Mean

Primary outcome

Average depression score.

4.5 Covariates

- Age
- Marital status
- Total annual income
- Highest educational level completed
- Religious affiliation
- Healthcare coverage status
- Occupation

- Number of children

5 STATISTICAL METHODS

5.1 Statistical analyses

5.1.1 Descriptive analyses

The epidemiological profile of the study participants will be described. Demographic (sex and age) and clinical variables will be described as mean (SD) or as counts and proportions (%), as appropriate. The distributions of participants' characteristics will be summarized in tables and visualized in exploratory plots.

5.1.2 Inferential analyses

All inferential analyses will be performed in the statistical models (described in the next section).

5.1.3 Statistical modeling

In order to account for the longitudinal dimension of the design a linear mixed model will be fit to the data, with random intercepts for the participant ID. This modeling approach is able to account for individual changes between scores at baseline and end of study, as well as determine the effect of the intervention while adjusting for other covariates.

Two approaches will be attempted to include the temporal dimension explicitly in the model. A random slope for the session indicator will be added to the participant ID random intercept together with a fixed effect for the session. If the first approach does not produce a convergent model, a second approach will be attempted by including a fixed-effect interaction term between the session indicator and the exposure indicator. Both approaches allow for the rates of changes to vary independently over time, and the resulting selected model (if any) will be presented as a complementary analysis.

It is intended to include as many covariates in the model as the sample power allows. If a model with all variables fails to converge then an approach of incremental inclusion will be used until the model fails to converge. If attempts at reaching a convergent model become too cumbersome due to collinearity between covariates, then a forward stepwise selection process will be applied with the minimal model that includes only the structure of the design (assessment, exposure and ID).

5.1.4 Missing data

No missing data imputation will be performed. All evaluations will be performed as complete case analyses. Missing data counts and proportions will be reported in tables.

5.2 Significance and Confidence Intervals

All analyses will be performed using the significance level of 5%. All significance hypothesis tests and confidence intervals computed will be two-tailed.

5.3 Study size and Power

N/A

5.4 Statistical packages

This analysis will be performed using statistical software R version 4.3.0.

6 OBSERVATIONS AND LIMITATIONS

Recommended reporting guideline

The adoption of the EQUATOR network (http://www.equator-network.org/) reporting guidelines have seen increasing adoption by scientific journals. All clinical trials are recommended to be reported following the CONSORT guideline (Schulz K F, Altman D G, Moher D., 2010).

7 REFERENCES

- **SAR-2023-035-MB-v01** – Efficacy of an ethnic sisterhood group in reducing depression in African American women: randomized controlled trial
- Zarin DA, et al. The ClinicalTrials.gov results database – update and key issues. N Engl J Med 2011;364:852-60 (https://doi.org/10.1056/NEJMsa1012065).
- Gamble C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343 (https://doi.org/10.1001/jama.2017.18556).
- Schulz K F, Altman D G, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials BMJ 2010; 340 :c332 (https://doi.org/10.1136/bmj.c332).

8 APPENDIX

This document was elaborated following recommendations on the structure for Statistical Analysis Plans (Gamble, 2017) for better transparency and clarity.

8.1 Availability

All documents from this consultation were included in the consultant's Portfolio.

The portfolio is available at:

https://philsf-biostat.github.io/SAR-2023-035-MB/