Statistical Analysis Plan (SAP)

Efficacy of a Fully Automated Electronic Screening and Brief Intervention to Reduce Harmful Alcohol Consumption: A Randomized Clinical Trial

Version 1

March, 2026

Table of Contents

1. Introduction and Overview

This document describes the statistical analysis plan for a single-center, open-label, randomized controlled trial comparing a fully automated electronic screening and brief intervention (e-SBI) delivered via a WhatsApp-based chatbot with a traditional in-person SBI. The primary objective is to evaluate the effectiveness of e-SBI in reducing harmful alcohol consumption (HAC) compared to in-person SBI.

This SAP was finalized prior to database lock and unblinding. Any post-hoc analyses will be clearly identified as such in all publications and reports.

2. Study Objectives

Objective	Description
Primary	To compare the effectiveness of e-SBI vs. in-person SBI in reducing harmful alcohol consumption, measured by change in AUDIT risk zone classification and continuous AUDIT scores.
Secondary	To compare adherence rates, referral rates to specialized services, and lifestyle behavior changes between the two intervention arms.

3. Study Endpoints

3.1 Primary Endpoints

Endpoint	Definition	Measurement Time
Change in AUDIT risk zone	Reclassification from higher risk zone (II, III, IV) to lower zone (I, II, III)	Baseline to 6 months
Change in continuous AUDIT score	Absolute and relative change in total AUDIT score (0-40)	Baseline to 6 months
Adherence to e-SBI	Completion of all e-SBI components (proportion)	Immediately post-intervention
Adherence to in-person SBI	Attendance and completion of scheduled in-person BI (proportion)	Immediately post-intervention
Referral rate	Proportion of Zone IV participants successfully referred to CAPS-AD	Within 30 days post-intervention

3.2 Secondary Endpoints

Endpoint	Definition	Measurement Time
Binge drinking episodes	Number of episodes (≥5 drinks for men, ≥4 for women) in past 30 days	Baseline, 3 months, 6 months
Lifestyle behavior composite score	Changes in smoking, physical activity, diet	Baseline and 6 months
Participant satisfaction	Likert-scale satisfaction score (1-5)	Post-intervention
Retention rate	Proportion of participants completing follow-up assessments	6 months

4. Sample Size and Power

Parameter	Value
Total sample size	516 participants
Per group	258 participants
Estimated attrition rate	20-25%
Power (1-β)	80%
Significance level (α)	0.05 (two-sided)
Minimum detectable effect size	Cohen's d = 0.25 for continuous outcomes; 10-15% absolute difference for proportions

Justification: Based on pilot data showing 38% reduction in binge drinking (Baldin et al.), the sample size is sufficient to detect a clinically meaningful difference between groups.

5. Analysis Populations

Population	Definition	Primary Use
Intention-to-Treat (ITT)	All randomized participants, analyzed according to original group assignment regardless of adherence	Primary analysis
Per-Protocol (PP)	Participants who completed the intervention as assigned without major protocol violations	Sensitivity analysis
Complier Average Causal Effect (CACE)	Participants who complied with their assigned intervention (defined a priori)	Secondary/Sensitivity analysis
Safety Population	All participants who received any intervention component	Safety/adverse event reporting

6. Descriptive Analysis

6.1 Baseline Characteristics

For each randomized group (e-SBI vs. in-person SBI), the following baseline characteristics will be described:

Variable Type	Variables	Descriptive Statistics
Demographic	Age, sex, race/ethnicity, education, income	Mean (SD) or median (IQR); n (%)
Clinical	AUDIT zone (I-IV), AUDIT continuous score	n (%); mean (SD) or median (IQR)
Lifestyle	Smoking status, physical activity level, diet quality	n (%)
Alcohol-related	Family history of AUD, prior treatment for alcohol use	n (%)
Technology access	Smartphone ownership, WhatsApp usage, internet access	n (%)

6.2 Flow of Participants (CONSORT)

A CONSORT diagram will be presented showing:

- Number screened
- Number eligible
- Number randomized (per group)
- Number receiving allocated intervention
- Number lost to follow-up (with reasons)
- Number analyzed (ITT and PP)

6.3 Missing Data Summary

For each key variable, the proportion and pattern of missing data will be reported, stratified by randomized group.

7. Normality Testing

Prior to parametric analyses, the following variables will be tested for normality:

- AUDIT continuous score (baseline and follow-up)
- Change in AUDIT score
- Age
- Lifestyle behavior composite score

7.1 Normality Tests

Test	Application
Shapiro-Wilk test	Primary test for normality (preferred for n < 5000)
Kolmogorov-Smirnov test	Secondary test (with Lilliefors correction)
Visual inspection	Q-Q plots, histograms, boxplots

7.2 Decision Rule

Normality Test Result	Statistical Approach
p > 0.05 (normal distribution)	Parametric tests (t-test, ANOVA)
p ≤ 0.05 (non-normal distribution)	Non-parametric tests (Mann-Whitney U, Wilcoxon signed-rank)
Mixed distribution	Bootstrap methods or transformation (e.g., log, square root)

8. Collinearity Analysis

To assess potential multicollinearity in multivariable models, correlations among candidate covariates will be evaluated prior to model fitting and the following procedure will be applied:

8.1 Variables to be Assessed for Collinearity

- Age
- AUDIT baseline score and AUDIT risk zone classification
- Socioeconomic status (education, income)
- Lifestyle behaviors related variables (smoking, physical activity, sleep quality…)

8.2 Collinearity Diagnostics

Measure	Threshold for Concern	Action if Exceeded
Pearson/Spearman correlation (r)	$\|r\| > 0.70$	Consider variable removal or combination
Variance Inflation Factor (VIF)	VIF > 5	Remove or combine collinear variables

8.3 Handling of Collinearity

When collinearity is identified, variable selection will prioritize:

- Clinical relevance
- Parsimony of the final model
- Preservation of interpretability

To achieve these goals, the following strategies may be applied:

- Hierarchical variable selection guided by clinical judgment (preserves interpretability and relevance).
- Principal component analysis (PCA) for dimensionality reduction — used only when interpretability of original variables is not essential.
- Ridge regression if collinearity persists after variable removal and PCA is not appropriate (e.g., when original variable interpretation must be kept).

Variable removal alone will not be the primary solution unless justified by clinical or parsimony criteria.

9. Between-Group Comparisons

9.1 Primary Endpoint Analysis

Endpoint	Statistical Test	Covariates (if applicable)
Change in AUDIT zone (categorical)	Chi-square test or Fisher's exact test	Stratified by baseline zone
Change in continuous AUDIT score	Student's t-test (normal) or Mann-Whitney U (non-normal)	Baseline AUDIT score
Adherence rate	Chi-square test	None
Referral rate	Chi-square test or Fisher's exact test	None

9.2 Secondary Endpoint Analysis

Endpoint	Statistical Test	Covariates
Binge drinking episodes	Negative binomial regression (count data)	Baseline episodes, group
Lifestyle composite score	ANCOVA (normal) or quantile regression (non-normal)	Baseline score, age, sex
Participant satisfaction	Mann-Whitney U test	None
Retention rate	Chi-square test	None

9.3 Longitudinal Analysis (Repeated Measures)

For outcomes measured at multiple time points (baseline, 3 months, 6 months):

Primary model: Linear mixed-effects model (LMM) with:

- Fixed effects: group, time, group × time interaction
- Random effects: participant-level intercept (and slope if appropriate)
- Covariates: baseline AUDIT score, age, sex

Alternative (non-normal): Generalized estimating equations (GEE) with appropriate link function (logit for binary, log for count, identity for continuous).

10. Intention-to-Treat (ITT) Analysis

10.1 Primary ITT Approach

All participants will be analyzed according to their original randomized group, regardless of:

- Adherence to the intervention
- Withdrawal from the study (if outcome data available)
- Protocol violations

10.2 Handling of Post-Randomization Exclusions

Participants will **not** be excluded from the ITT analysis for any post-randomization reason except:

- Withdrawal of consent for data use (data will be excluded entirely)
- Loss to follow-up (handled via imputation – see Section 11)

11. Missing Data Handling – Multiple Imputation (MICE)

11.1 Assumptions

Missing data will be assumed **missing at random (MAR)** after accounting for observed covariates. Sensitivity analyses will test the robustness under missing not at random (MNAR) assumptions.

11.2 Variables with Expected Missingness

Variable	Expected Missing Rate	Imputation Method
AUDIT follow-up (6 months)	20-25%	MICE
Lifestyle behaviors	15-20%	MICE
Satisfaction survey	10-15%	MICE
Baseline demographics	<5%	Complete case or mode imputation

11.3 MICE Procedure

Parameter	Specification
Software	R (mice package) or SPSS (multiple imputation)
Number of imputations	20 (rule of thumb: ≥ percentage of missing data)
Maximum iterations	50
Imputation model variables	All analysis variables + auxiliary variables (e.g., weekday of visit, recruitment site)

11.4 Imputation Models by Variable Type

Variable Type	Imputation Method
Continuous (normal)	Predictive mean matching (PMM)
Continuous (skewed)	PMM with 5 nearest neighbors
Binary (e.g., sex)	Logistic regression
Categorical (≥3 levels)	Polytomous logistic regression
Count (e.g., binge episodes)	Poisson or negative binomial regression

11.5 Pooling of Results

Rubin's rules will be used to combine estimates and standard errors across imputed datasets.

11.6 Sensitivity Analysis for Departures from MAR (Assessing MNAR)

Inverse probability-of-censoring weighting (IPCW) will be performed using predictors of dropout to reduce potential attrition-related bias under MAR. To assess the robustness of findings to missing not at random (MNAR) mechanisms, the following sensitivity analyses will be conducted:

Analysis	Description
Inverse probability-of-censoring weighting (IPCW)	Weight complete cases by inverse probability of dropout to correct for selection bias under MAR
Pattern-mixture models	Assess effect of MNAR by stratifying on missingness pattern
Delta adjustment	Add a fixed offset to imputed values (e.g., +2 AUDIT points for missing)
Complete case analysis	Compare with MICE results; major discrepancies will be reported

These analyses will help evaluate whether conclusions are robust to violations of the MAR assumption.

12. Complier Average Causal Effect (CACE) Analysis

12.1 Definition of Compliance

Compliance is defined based on engagement with the intervention prior to the 6-month follow-up assessment. Completion of follow-up is **not** required for compliance classification.

Arm	Definition of Compliance
e-SBI arm	Completed at least 80% of chatbot interactions and completed 6-month follow-up
In-person SBI arm	Attended scheduled in-person intervention session and completed 6-month follow-up

12.2 Rationale for CACE

A secondary Complier Average Causal Effect (CACE) analysis will be conducted to estimate the intervention effect among participants who adhered to their assigned intervention. In the presence of non-adherence, intention-to-treat (ITT) estimates may be biased toward the null. CACE provides a complementary estimate of the causal effect among compliers.

12.3 CACE Model

CACE analyses will be performed using instrumental variable (IV) methods, with randomized group assignment as the instrumental variable. Continuous outcomes (e.g., change in AUDIT score) will be analyzed using two-stage least squares (2SLS) with baseline covariates (age, sex, baseline AUDIT).

Model specification

Stage 1 (Compliance prediction):

$$\text{Complier}_i = \pi_0 + \pi_1 \text{RandomizedArm}_i + \pi_2 X_i + u_i$$

Stage 2 (Outcome model among compliers):

$$Y_i = \beta_0 + \beta_1 \text{Complier}_i + \beta_2 X_i + \epsilon_i$$

(estimated via 2SLS with corrected standard errors)

Where:

- Y_i = outcome (e.g., change in AUDIT score)
- $\mathrm{RandomizedArm}_i$ = instrumental variable
- X_i = baseline covariates (age, sex, baseline AUDIT)
- $\mathrm{Complier}_i$ = predicted probability of compliance from Stage 1

12.4 Assumptions for CACE

Assumption	Verification
Exclusion restriction	Randomization affects outcome only through intervention receipt (plausible by design)
Relevance	Strong association between randomization and compliance (check F-statistic > 10)
Independence	Randomization ensures independence of potential outcomes (plausible by design)
Monotonicity	No "defiers" (those who do opposite of assignment) – plausible in this context

Results from the CACE analysis will be considered supportive and exploratory.

12.5 CACE Estimation Methods

Method	Application
Two-stage least squares (2SLS)	For continuous outcomes (AUDIT change)
Bivariate probit	For binary outcomes (AUDIT reclassification)
Structural equation modeling (SEM)	For sensitivity analysis

Losses in one arm, inverse probability weighting may partially adjust for attritionunder strong assumptions, whereas CACE mainly addresses noncompliance and generally does not solve the core missing-data problem.

12.6 Relationship Between CACE and Missing Data

CACE analysis primarily addresses non-adherence to assigned intervention. It does **not** solve the core missing-data problem (attrition). Missing outcome data will be addressed separately using the methods described in Section 11 (Multiple Imputation and sensitivity analyses). The combination of ITT with multiple imputation, CACE for non-adherence, and MNAR sensitivity analyses provides a comprehensive approach to handling both missingness and non-compliance.

13. Predictor Analysis for HAC Reduction

These analyses will be considered exploratory and hypothesis-generating.

13.1 Objective

To identify baseline factors associated with successful reduction in harmful alcohol consumption (defined as ≥2-point reduction in AUDIT score or downward zone transition).

13.2 Predictor Variables to be Tested

Domain	Candidate Predictors
Demographic	Age, sex, education, income, employment status
Clinical	Baseline AUDIT zone, AUDIT continuous score, comorbidities
Lifestyle	Smoking status, physical activity, diet
Psychosocial	Social support, perceived stress, motivation to change
Intervention-related	Assigned group (e-SBI vs. in-person), adherence level
Technological	Digital literacy, smartphone ownership, prior WhatsApp use

13.3 Modeling Strategy

Step 1 – Univariate screening

- Each predictor tested individually using t-test, chi-square, or logistic regression
- Predictors with $p < 0.20$ enter multivariable model

Step 2 – Multivariable logistic regression (primary)

- Dependent variable: HAC reduction (binary: yes/no)
- Independent variables: all predictors with $p < 0.20$ from Step 1
- Variable selection: backward elimination ($p < 0.05$ to retain) or LASSO regularization

Step 3 – Multivariable linear regression (secondary)

- Dependent variable: absolute change in AUDIT score (continuous)
- Same predictor set and selection procedure

13.4 Model Diagnostics

Diagnostic	Criterion
Hosmer-Lemeshow goodness-of-fit	$p > 0.05$ indicates adequate fit
Area under ROC curve (AUC)	> 0.70 acceptable, > 0.80 good
Calibration plot	Visual inspection
Residual analysis	For linear regression (normality, homoscedasticity)

13.5 Interaction Testing

Pre-specified interactions to be tested (with $p < 0.10$ considered significant):

- Group (e-SBI vs. in-person) × baseline AUDIT zone
- Group × age
- Group × digital literacy

14. Sensitivity Analyses

Analysis	Description	Purpose
Per-protocol (PP)	Analyze only participants who completed intervention as assigned	Assess robustness of ITT findings
Completer analysis	Analyze only participants with complete follow-up data	Compare with MICE results
Bootstrap	1,000 resamples for confidence intervals	Assess sampling variability
Subgroup analyses	By baseline AUDIT zone, age group, sex	Explore potential effect modification

All sensitivity analyses will be interpreted as supportive and not confirmatory of the primary ITT findings.

15. Subgroup Analyses

Pre-specified subgroups (exploratory; results will be interpreted cautiously):

Subgroup	Definition	Analysis Method
Baseline AUDIT risk	Zones I, II, III, IV	Interaction test in logistic regression
Age	<40, 40-60, >60 years	Stratified analysis
Sex	Male, female	Interaction test
Digital literacy	High vs. low (based on median split)	Stratified analysis

Note: No adjustment for multiple comparisons in exploratory subgroup analyses. Findings will be reported as hypothesis-generating.

16. Software

Software	Version	Purpose
R	≥ 4.2	Primary analysis (mice, lme4, AER, ggplot2)
SPSS	≥ 28	Secondary/descriptive analysis
SAS	≥ 9.4	Sensitivity (if needed)
G*Power	3.1	Post-hoc power calculations

17. Reporting Standards

All analyses will follow:

- **CONSORT 2010** statement for randomized trials
- **CONSORT extension** for pragmatic trials
- **ICH E9** statistical principles for clinical trials

18. Tables and Figures Planned

18.1 Tables

1. Table 1: Baseline characteristics by randomized group (ITT population)
2. Table 2: Primary outcomes by group (ITT, PP, CACE)
3. Table 3: Secondary outcomes by group
4. Table 4: Missing data summary and imputation diagnostics
5. Table 5: CACE analysis results
6. Table 6: Multivariable predictors of HAC reduction
7. Table 7: Subgroup analyses (forest plot format)
8. Table 8: Adverse events (if collected)

18.2 Figures

1. Figure 1: CONSORT flow diagram
2. Figure 2: AUDIT score change over time (line plot, group × time)
3. Figure 3: Forest plot of subgroup effects
4. Figure 4: Histograms of missing data patterns
5. Figure 5: Calibration plot for prediction model
6. Figure 6: ROC curve for HAC reduction prediction

19. Signatures and Version Control

Version	Date	Author	Changes
1.0	March, 30 2026	Caio Vitale Spaggiari	Original SAP

Approval:

- Principal Investigator: Martino Martinelli-Filho Signer ID: SCQJYLR118... Date: March, 30 2026

- Lead Statistician: Caio V. Signer ID: KAHLUTR418... Date: March, 30 2026

20. References

1. Little RJA, Rubin DB. Statistical Analysis with Missing Data. 3rd ed. Wiley; 2019.
2. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-399.
3. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. J Am Stat Assoc. 1996;91(434):444-455.
4. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c869.
5. ICH E9 Expert Working Group. Statistical Principles for Clinical Trials. Stat Med. 1999;18(15):1905-1942.