

SEHAT Trial Statistical Analysis Plan

Official Title:

Evaluating the Impact of Novel Oral Health Promotion Program “SEHAT” among Adolescents in Pakistan’s Resource-Constrained Settings – A Cluster Randomized Trial

Document Type:

Statistical Analysis Plan

Document Version / Date:

Version 1.0 – April 10, 2026

Principal Investigator:

Dr. Sadia Musharraf

BDS, MCPS, PhD (Candidate)

Assistant Professor, Shifa College of Dentistry, Shifa Tameer e Millat University, Pakistan

Confidentiality Statement:

This document contains confidential information related to the SEHAT trial. It is intended solely for regulatory review and research purposes. Names of participants are not included.

SEHAT Trial Statistical Analysis Plan

Descriptive statistics will be used to describe participants' demographic characteristics and variables of interest. Baseline comparisons of demographic characteristics between intervention and control groups will be done to assess effective randomization. Summary statistics will be reported as frequency and percentages N (%) for categorical variables including toothbrushing frequency, sugary food and drinks intake, baseline oral health status, and mean (standard deviations) for continuous variables including plaque scores, bleeding scores and oral health knowledge scores. Baseline data will be used to investigate the characteristics of any participants who discontinue or deviate from the trial.

Primary Analysis

- The primary analyses will use the intention-to-treat (ITT) principle, where each eligible schoolchildren at baseline, to participate in the study, will be considered in analysis according to the class randomization assignment regardless of whether the child leaves school or lost to follow-up after randomization or is not evaluated at the follow-up dental examinations, but still all data will be used in the analysis.
- Generalized linear mixed-effects model will be used for individual-level analysis to assess the associations between oral health status and individual-level factors (attitudes, knowledge). The multilevel random-effects model is used because of the outcomes for adolescents in one cluster are not independent and the data are organized at more than one level (i.e. nested data). The units of analysis are adolescents (at a lower level) who are nested within clusters, i.e classrooms (at a higher level) and parameters vary at more than one level.
- Models will be adjusted for baseline characteristics and potential confounders. P-value of ≤ 0.05 will be considered statistically significant.

Secondary Analyses

- A secondary analysis will be conducted on the per-protocol cohort using the adolescent that have completed follow-up dental examination.
 - Subgroup analyses gender, grade level, and school settings
 - Mediation analysis to explore mechanisms of intervention effects
 - Per-protocol analyses to assess the impact of intervention adherence

Missing Data

- Multiple imputation techniques will be used to handle missing data, assuming data are missing at random.
- Qualitative data from semi-structured interviews with adolescents and teachers will be analyzed thematically using NVivo or equivalent software. A combined inductive and deductive approach will be applied, with coding guided by COM-B Model and the Theoretical Domains Framework (TDF). Inter-coder reliability and iterative theme refinement will enhance rigor. Themes will focus on feasibility and acceptability, behavioral drivers and barriers, social and environmental Influence, and practical challenges.