Cover Page

Title: Real-World Durability of Weight Loss With a Breath Biofeedback mHealth Program: Two-Year Outcomes in Adults With Overweight or Obesity

NCT number: Pending

Document Date: February 4, 2026

Version: v1

PROTOCOL AND STATISTICAL ANALYSIS PLAN (SAP)

Real-World Durability of Weight Loss With a Breath Biofeedback mHealth Program: Two-Year Outcomes in Adults With Overweight or Obesity

Date: February 4, 2026

Background and aim:

This Statistical Analysis Plan (SAP) describes the planned analysis for the retrospective real-world evidence (RWE) study of the Key to Health digital weight loss program. The study utilizes a retrospective, observational cohort design using de-identified data from commercially active users of the Key to Health program stored in the company's backend database. Users initiated the program between November 2018 and April 2023 and the data collection window spans from November 2018 through April 2025. Given the observational nature of the study, results are descriptive of long-term outcomes among program users rather than representative of causal effects.

The aim of the study is to evaluate the outcomes and durability of the program over 104 weeks (2 years) in a commercial cohort of adults with overweight or obesity. Specifically, this analysis aims to characterize weight loss trajectories in a real-world setting.

ANALYSIS OBJECTIVES AND OUTCOMES

Primary objective:

The primary objective is to estimate the mean percentage self-reported or connected-scale total body weight loss (TBWL%) at 104 weeks (2 years) in the population of app users, with TBWL% defined as the following: TBWL% = (Post-baseline weight – Baseline weight) / Baseline weight × 100

Secondary objectives:

Secondary objectives include estimating the mean TBWL% at 52 weeks (1 year), 24 weeks, and 12 weeks. Additionally, the analysis will estimate the proportion of users achieving clinically significant weight loss (defined as ≥5% and ≥10% loss) at 52 and 104 weeks, and assess the absolute weight change (lbs) at all timepoints.

Exploratory objectives:

Exploratory objectives include evaluating the relationship between device utilization frequency (breath sensor usage), breath acetone level, and weight loss outcomes, as well as assessing outcomes across demographic subgroups including Age (< vs ≥65 years), body mass index (BMI), and sex.

DEFINITIONS AND STUDY POPULATION

Eligibility criteria:

The analysis population will consist of all users meeting the following criteria.

Inclusion criteria are: (1) valid baseline self-reported weight recorded during onboarding; (2) program start date between Nov 2018 and Apr 2023; (3) age ≥18 years at baseline; and (4) BMI >25.

Exclusion criteria are: (1) invalid or biologically implausible baseline weight; and (2) no weight data recorded after the baseline date.

Analysis populations:

The dataset for this analysis will be derived from all eligible users who have a valid baseline weight and at least one valid weight measurement post-baseline. Users with no post-baseline data will be excluded by design. This population will be used for the primary analysis. The completer set (observed) consists of the subset of users who have a valid weight measurement within each specific time window of interest.

Definition of baseline vs. post-baseline weights:

Baseline weight will be defined strictly as the self-reported weight recorded during the onboarding flow. Post-baseline weights will be defined as any weight logged by the user via the app or connected scale after the onboarding timestamp. Weights logged on "Day 0" represent a weight entry the day of onboarding and will be considered post-baseline but will not be used to replace the self-reported baseline. Implausible weight changes, defined as change of >15% within one week, will be flagged and excluded. Post-baseline weights meeting this criterion will be excluded from visit-window selection.

Endpoints:

The primary efficacy endpoint is the percentage total body weight loss (TBWL%) at week 104. Secondary efficacy endpoints include TBWL% at weeks 12, 24, and 52; absolute change in

weight (lbs) at weeks 12, 24, 52, and 104; and the proportion of participants with ≥5% and ≥10% weight loss at weeks 52 and 104 (observed-case summaries).

Time windows:

Outcomes will be categorized into time point windows of interest based on days from program start. The deterministic rule for selection of the endpoint (i.e., post-baseline weight entry at that time point) will be the weight entry closest to the target day within the window. If two weights are equidistant, the earlier weight will be selected. The windows will be defined as follows: week 12 (target day 84; window 69–99 days); week 24 (target day 168; window 138–198 days); week 52 (target day 364; window 334–394 days); and week 104 (target day 728; window 668–788 days).

The wider window at week 104 (±60 days vs. ±15–30 days at earlier visits) reflects reduced logging frequency at long-term follow-up while maintaining clinical relevance for assessing weight maintenance. Sensitivity check using a narrower window (±30 days) may be performed, if feasible.

STATISTICAL ANALYSIS

Estimand:

The primary estimand is the mean TBWL% at 104 weeks in the full analysis (ITT) population (eligible users with baseline weight and ≥1 post-baseline weight) under a Missing At Random (MAR) assumption, estimated using a Mixed Model for Repeated Measures (MMRM). This estimand reflects the average outcome among enrolled users conditional on the observed data and covariates.

General principles:

All statistical tests will be 2-sided with a significance level of alpha = 0.05. Continuous variables will be summarized using means, standard deviations (SD), and 95% confidence intervals (CI). Categorical variables will be summarized using counts and percentages.

Primary analysis (MMRM):

The primary endpoint will be analyzed using a Mixed Model for Repeated Measures (MMRM). The dependent variable is TBWL% at each visit.

Fixed effects include: Visit (Categorical: 12w, 24w, 52w, 104w); baseline weight (Continuous); age (Continuous); sex (Categorical); and cohort year (Categorical).

The primary covariance structure is Unstructured (UN). If convergence fails, simpler structures (e.g., autoregressive or compound symmetry) will be explored. Degrees of freedom will be estimated using the Satterthwaite approximation. The main output of interest will be the Least-Squares Mean (Estimated Marginal Mean) for week 104 with corresponding 95% CI.

Model specification will be assessed visually using normal probability plots and residuals vs fitted values plots. If departures from model assumptions are observed, transformations (e.g., log-transformation) will be explored.

Handling of missing data:

The primary strategy utilizes MMRM under the Missing At Random (MAR) assumption. No statistical imputation (e.g., last observation carried forward, mean imputation) will be performed. The MMRM approach provides a principled method for addressing missing data under MAR by including all participants with at least one post-baseline observation.

SENSITIVITY ANALYSES

To assess robustness of the primary week 104 TBWL% estimate to missing-data assumptions, the following sensitivity analyses will be performed:

MNAR sensitivity (Tipping Point / Pattern-Mixture with δ-adjustment):

A pattern-mixture framework will be used to evaluate departures from the primary MAR assumption. For participants missing the Week-104 outcome, Week-104 TBWL% will be imputed under increasingly conservative Missing Not At Random (MNAR) scenarios by applying a prespecified δ-shift (worsening penalty) to the Week-104 prediction from the primary MMRM. Scenarios will include "Return to Baseline" (assuming all missing data regressed to 0% loss).

IPW / weighted analysis for attrition:

Inverse-probability-of-observation weights (IPW) will be used to account for differential observation at follow-up visits. The probability of having an observed Week-104 outcome will be modeled as a function of baseline covariates (Baseline weight, Age, Sex, Cohort Year). Weighted estimation of Week-104 mean TBWL% will be compared to the primary MMRM result. If feasible, an expanded model may additionally include earlier observed weight-change summaries.

Sensitivity analyses will be conducted as feasible based on data availability and implementation resources; any deviations from the prespecified plan will be documented.

Observed-case (completer set) summaries:

Descriptive statistics will be reported using observed data only among participants with an in-window weight measurement at each visit (weeks 12, 24, 52, and 104), including mean TBWL%, mean absolute weight change (lbs), and responder rates (≥5% and ≥10% loss) at weeks 52 and 104.

SECONDARY ANALYSES

Secondary continuous endpoints (TBWL% at Weeks 12, 24, and 52; and absolute weight change at all visits) will be estimated using the MMRM framework. For TBWL%, estimates at earlier visits are extracted directly from the primary longitudinal model. For absolute weight change, an analogous MMRM will be fitted with weight change (lbs) as the dependent variable. Binary responder endpoints (≥5% and ≥10% weight loss) will be summarized descriptively by visit using counts and percentages within the Observed (Completer) population.

EXPLORATORY ANALYSES

Subgroup analyses will be conducted if feasible to evaluate the consistency of outcomes across key demographic categories, including Age (stratified as <65 vs. ≥65 years), Sex, Weight input type (manual entry vs connected scale), and Baseline weight categories. Additionally, the associations between weight loss magnitude and (1) program engagement (defined by app and/or breath sensor utilization frequency) and (2) breath acetone levels will be assessed using mixed models, adjusting for baseline covariates.

PROGRAMMING PLAN

Statistical analysis will be performed using R (Version 4.0 or later) or SAS. The longitudinal data analysis will be performed using relevant packages for MMRM (e.g., mmrm in R, nlme, or SAS PROC MIXED).