Status: Release
Version: 2
Release Date: 02/08/2019 01:24:10 PM CST
NCT #: NCT03421145

STATISTICAL ANALYSIS PLAN

A prospective, randomized, blinded, clinical study to evaluate effects of a 3M oral rinse on plaque and gingivitis.

Clinical Study Number EM-11-050003
Center for Dental Research,
Loma Linda University School of Dentistry

Prepared by
Shelley-Ann Walters
3M Biostatistician

Statistical Analysis Plan

Approvers
Do not change the table below as it is automatically populated by the system when the document is released.

Signer	Role	Date Signed
us328761:Zehrer Cindy L	Clinical	February 08, 2019 01:24:08 PM CST
US314980:Meeter Carol A	Clinical	February 08, 2019 12:50:05 PM CST

Contents

1. INTRODUCTION

The following Statistical Analysis Plan (SAP) was written and approved prior to the study enrollment. The text in this SAP reflects the analysis methods planned as of the date of the SAP. It includes more details to the analyses mentioned in the protocol and may be amended to reflect any additional analyses planned prior to unblinding the data.

2. STUDY OBJECTIVES

2.1 Primary Objective

To investigate the effects of the 3M Oral Rinse relative to the placebo vehicle control on plaque and gingivitis levels, at 6 months, based on the Modified Quigley-Hein Plaque Index (PI) and the Löe-Silness Gingival Index (GI) scores.

2.2 Secondary Objectives

· To investigate the effects of the 3M Oral Rinse vs. Placebo Vehicle Control on plaque and gingivitis, at 3 months, based on the Modified Quigley-Hein Plaque Index (PI) and the Löe-Silness Gingival Index (GI) scores.

· To investigate the effects of the 3M Oral Rinse vs. Placebo Vehicle Control on Eastman Interdental Bleeding Index (BI) scores at 3 and 6 months

· To investigate the relative efficacy, using pairwise comparisons, of the four treatment groups, namely 3M Oral rinse, vehicle control, PerioShield (predicate) and water control, with respect to the PI and GI scores at 6 months.

· To assess the safety of the 3M oral rinse by tracking abnormal changes to soft and hard tissues, reported side effects and adverse events during the study.

3. INVESTIGATIONAL PLAN

3.1 Study Design

This is a prospective, single site, randomized, parallel, blinded clinical trial. Subjects will be randomly assigned to one of four treatment groups. The treatment groups are presented below:

Treatment Groups

Treatment Code	Study Product	Description
A	3M Oral Rinse	The volume of 15 mL will be swished for 30 seconds followed by expectoration, twice a day (morning and evening), after brushing teeth, for the 6-month duration of the study.
B	3M Vehicle Control	The volume of 15 mL will be swished for 30 seconds followed by expectoration, twice a day (morning and evening), after brushing teeth, for the 6-month duration of the study.
C	PerioShield (Predicate)	PerioShield™ Oral Health Rinse (0.2% delmopinol hydrochloride) is marketed in the US and will be purchased for use in the study. The subjects will follow the product instruction for use using 10 mL for 30 seconds followed by expectoration, twice a day (morning and evening), after brushing teeth, for the 6-month duration of the study.
D	Water	The water control will be purchased sterile water. The volume of 15 mL will be swished for 30 seconds followed by expectoration, twice a day (morning and evening), after brushing teeth, for the 6-month duration of the study.

3.2 Randomization

The treatment groups consisting of four different oral rinses will be assigned to the letter labels A, B, C, D.

The randomization was created by a biostatistician who will not be involved in the analysis of the study data, keeping the study biostatistician blinded to the randomization schedule. The randomization schedule created treatment assignments for 216 subjects, 16 extra over the protocol specified sample size to account for contingencies. The randomization ratio was 1:1:1:1 among the four treatment groups.

Details of the randomization scheme followed, and the protocol deviations related to randomization errors will be documented in the final study report.

3.3 Blinding and Blinding Assessment

Before data lock and final study data analysis, the biostatistician will remain blinded to treatment assignment. The final data review and data classification meeting will be conducted before data lock and statistical programming will occur using dummy treatment assignments.

The Principal Investigator and clinical examiners will be blinded to the treatment products. No clinical examiners will see the allocated rinse assigned per the randomization schedule. Due to the rinse packaging differences and to maintain the blind, all study products will be placed in a non-transparent container. Subjects will be told not to open the container until they return to their home and to bring study supplies back to their next scheduled appointment in the same container. Only the study coordinator at the site will know the randomization assignment.

Each subject will be asked to indicate to what treatment he or she thinks they have been assigned. This assessment will be done immediately after the first dose and at the end-of-study visit. The subject will indicate if he or she thinks the treatment assigned is (a) the new 3M Oral Rinse, (b) one of the control rinses, or (c) does not know or cannot make an educated guess. The responses to the blinding assessment questionnaire at both time points will be summarized by the randomized treatment group.

4. Efficacy Endpoints

4.1 Löe-Silness Gingival Index (GI)

0=	Absence of inflammation.
1=	Mild inflammation: slight change in color and little change in texture.
2=	Moderate inflammation: moderate glazing, redness, edema, hypertrophy. Tendency to bleed upon probing.
3=	Severe inflammation: marked redness and hypertrophy. Tendency to spontaneous bleeding.

The gingival surface surrounding each tooth (all natural teeth) will be scored on six surfaces: 1) mesio-facial; 2) mid-facial; 3) disto-facial; 4) mesio-lingual; 5) mid-lingual; and 6) disto-lingual. Third molars and those teeth with cervical restorations or prosthetic crowns will be excluded from the scoring procedure. The GI score of the individual will be obtained by adding the scores of all surfaces and dividing by the number of surfaces examined.

4.2 Modified Quigley-Hein Plaque Index (PI)

0=	No plaque.
1=	Separate flecks of plaque at the cervical margin of the tooth.
2=	A thin, continuous band of plaque (up to one mm) at the cervical margin of the tooth.
3=	A band of plaque wider than 1 mm but covering less than 1/3 of the crown of the tooth.
4=	Plaque covering at least 1/3, but less than 2/3 of the crown of the tooth.
5=	Plaque covering 2/3 or more of the crown of the tooth.

Prior to examination, a plaque disclosing solution will be applied to all surfaces of teeth. Each tooth is scored for supragingival plaque on six surfaces: 1) mesio-facial; 2) midfacial; 3) disto-facial; 4) mesio-lingual; 5) mid-lingual; and 6) disto-lingual. Third molars and those teeth with cervical restorations or prosthetic crowns will be excluded from the scoring procedure. The PI score of the individual will be obtained by adding the scores of all surfaces and dividing by the number of surfaces examined.

4.3 Eastman Interdental Bleeding index (BI)

0=	Absence of bleeding.
1=	Presence of bleeding

The interdental cleaner is inserted between two teeth from the facial aspect, depressing the interdental tissues 1 to 2 mm. This is repeated four times and the presence or absence of bleeding within 15 seconds is recorded. The absence or presence of bleeding will be recorded for each examined site (between two teeth), and a score will be calculated as the number of bleeding sites out of the number of examined sites for each evaluation on a subject.

5. ANALYSIS POPULATIONS

5.1 Safety Population

All subjects will be included in the safety analyses.

5.2 Intent-To-Treat Population

The intent-to-treat (ITT) data set will be the primary efficacy data set and will include all subjects randomized to a treatment group. Subjects will be analyzed in the group to which they were randomly assigned. Subjects with missing post-baseline PI and/or GI data will be included in the ITT analysis using appropriate input data sets and imputation methods. The Expectation-Maximization (EM) algorithm will be used, and this procedure assumes the data will be multivariate normal and missing at random. If these assumptions prove to be violated, other methods will be considered.

5.3 Per-Protocol Population

A second per-protocol (PP) data set will be defined excluding any subjects with major deviations or lack of compliance with taking their assigned study treatment. The PP dataset will be used only for the primary efficacy analysis (3M oral rinse vs. vehicle control) and the secondary efficacy analysis with all four treatment groups.

The following major protocol deviations will lead to exclusion from the per protocol data set:

(1) Subject who is missing either one or both primary endpoints of GI and PI values at 6 months
(2) Subject who did not meet the inclusion criteria of baseline Gingival Index (GI) score of >= 1.0 and a baseline Plaque index (PI) score of >= 1.5.
(3) Subject with < 80% of planned dose (or alternatively >20% of no rinsing).
(4) Subject with >=3 consecutive days of no rinsing within 4 weeks prior to PI and GI assessments.
(5) Subject who indicated antibiotics, anti-inflammatory, anticoagulants, or immunosuppressant drug use, regardless of dose, within 6 weeks prior to PI and GI assessments.

All deviations not captured above that will result in exclusion from the PP analysis will be defined and justified in a blinded fashion. In addition, documentation will be made of these data classification decisions prior to data lock.

6. GENERAL STATISTICAL CONSIDERATIONS

Means and standard deviations will be calculated for all clinical measurements and assessments. Categorical variables will be summarized in frequency distributions. Analysis of variance (ANOVA) or (ANCOVA) will be used to test for treatment group differences between the subject groups over the visits. When appropriate, alternative statistical methods such as normalizing transformations or nonparametric tests may be used to analyze the data.

GI, PI, and BI scores will be calculated as the average tooth score across all measures and teeth surfaces as one score per patient, per visit. For example, a patient with 28 evaluable teeth will have 168 data points (6 per tooth) averaged per tooth and then averaged over all teeth to produce one score per patient per visit. The subject indices computed will range between 0-3 (GI) or 0-5 (PI) or 0-1 (BI).

All statistical analyses will be performed using SAS version 9.4 or above (SAS Institute, Cary, NC) or another statistical software package. Any changes to the methods proposed in this SAP will be documented in the "Changes to Planned Analyses" section of the final clinical study report.

6.1 Determination of Sample Size

The sample size was estimated based on GI and PI scores from preliminary clinical study data and the literature. It is estimated that 34 subjects per group will provide 90% power to detect a difference of 0.40 (standard deviation of 0.5) in the GI and PI scores using a 1-sided 0.025 alpha level. To account for possible dropouts, a sample size of 50 subjects per group will be recruited.

6.2 Handling of Dropouts and Missing Data

Subjects with missing PI and/or GI data will be included in the ITT analysis using appropriate input data sets and imputation methods. It is assumed that the data will be multivariate normal and missing at random, and Expectation-Maximization (EM) algorithm will be used. If this assumption proves to be unreliable, other methods will be considered. In addition, a sensitivity analysis using alternative imputation methods to handle the missing data, namely last observation carried forward and using group means, will be conducted to examine the robustness of the primary analysis.

If a subject becomes pregnant during the study, the subject will be discontinued, and assessments completed.

6.3 SAS Imputation Procedures

SAS PROC MI procedure will be used to perform imputations of the missing primary variables, namely, the PI and GI indices. The EM algorithm assumes multivariate normal distribution for the data; PI and GI values (baseline values included). The missing data for a given treatment group will be imputed independently with input data being the PI and GI data at all time points (baseline, 3 months and 6 months) from only that treatment group. The imputation procedure is valid when the missing data are missing at random. %MULTINORM macro can test for multivariate normality and PROC MI can display patterns of missing data.

6.4 Training and Calibration of Clinical Examiner in the GI and PI Measures

The clinical examination of the GI and PI will be performed by experienced examiners who have publications on the GI and PI evaluation, and the examiners will perform the same clinical examination throughout the study. To ensure adequate reliability and consistency, a training and calibration session on the GI and PI evaluation will be conducted for the examiners prior to the baseline, 3-month, and 6-month visits.

The training and calibration will be conducted using digital dental photographs of 15 historical subjects selected from a library at the study site. The digital dental photographs of 15 subjects will be selected to provide adequate representation of the full range of the GI (0 to 3) and PI (0 to 5) scales. In addition, the goal will be to achieve a sample that is representative of the study population in terms of race, age and sex. The digital dental photographs instead of subjects are used for training and calibration because clinical examination of the GI and PI on the same subject by multiple examiners will produce significant variations between evaluations as well as discomfort to the subject.

The Principal Investigator will review the criteria of the GI and PI with the examiner, who will then determine the initial score; after one week, the examiner will again reevaluate the same set of the digital dental photographs in a random order. The two-way random, single measure reliability of the Intraclass Correlation Coefficient (ICC) will be computed and ICC values ≥0.85 will signal acceptable calibration of the study examiner.

7. SUMMARY OF STUDY POPULATION DATA

7.1 Subject Disposition

Subject enrollment, including number of subjects screened, number of subjects randomized and the reasons for screen failures, will be summarized by treatment group (if appropriate) and overall.

Subject disposition, including the number and percentage of subjects included in each analysis data set (ITT or PP), as well as reasons for exclusion from the PP data set, will be summarized by treatment group and overall.

Study completion status (completed the study or discontinued early from the study), as well as the primary reason for study discontinuation, will be summarized by treatment group and overall.

7.2 Protocol Deviations

Protocol deviations are any departure from the protocol. Protocol deviations that affect the evaluability of subjects will be reviewed during a blinded review process and documented before data lock and study analysis. Deviations will be summarized by categories, treatment group and overall. Listing of protocol deviations will also be provided.

7.3 Baseline and Demographic Characteristics

Demographic variables will be examined to assess comparability of the treatment groups prior to the initiation of treatment. Characteristics evaluated include age, sex, race, ethnicity and tobacco use status. An overall Fisher's Exact test (discrete variables) or ANOVA model (continuous variables) will be used to compare each demographic or baseline variable among all four treatment groups.

Descriptive summary statistics will be provided for demographic characteristics for each treatment group and overall. For age, the number of subjects, mean, standard deviation, median, minimum and maximum will be provided. For sex, race, ethnicity and tobacco use status, the number and percentage of subjects in each demographic category will be summarized.

In addition, screened subjects participating in the wash-out period before randomization will be summarized by demographic variables.

7.4 Treatment Duration and Oral Rinse Dosing Compliance

Treatment duration will be computed as the number of days from randomization to the last dosing day: (Last Dosing Date-Randomization Date +1). If the last dosing date is missing, the day of last study visit will be used.

The oral rinse compliance will be calculated as a percentage of the documented number of doses taken in the subject diary of the total number of expected doses. The subjects will be expected to take two doses per day unless it is the last day of the treatment period, on which no dose will be expected (based on dosing instruction for PI and GI assessment days). In addition, subjects will be expected to dose for 6 calendar months. For example, subject starting treatment on January 15th will be expected to dose until July 14th with their expected 6-month visit occurring on July 15th.

8. EFFICACY ANALYSIS

8.1 Primary Efficacy Analysis (3M Oral Rinse vs. Vehicle Control)

Performance of the 3M Oral Rinse will be assessed based on two primary endpoints. Statistical analysis will be performed on the PI and GI scores at 6-months using an ANCOVA with treatment as the factor and baseline score as a covariate. The success criterion for the study will be based on the single comparison between the 3M Oral Rinse group and the vehicle control group. The primary efficacy analysis will exclude data from the other two treatment groups, namely, the PerioShield (predicate) and the water control groups.

Significance of the overall treatment effect will be assessed by the p-value associated with the F-test statistic for the treatment group. The level of significance will be two-sided alpha of 0.05. The main null hypothesis of interest to support that 3M Oral Rinse reduces plaque and gingivitis is as follows:

$$H_0: \mu_{3M\ Oral\ Rinse} \geq \mu_{Vehicle} \text{ versus } H_A: \mu_{3M\ Oral\ Rinse} < \mu_{Vehicle}$$

Both primary endpoint comparisons will be based on the intent-to-treat (ITT) data set. The study will be considered successful if both primary endpoints are met.

In the event of a significant demographic or baseline variable, the primary efficacy ANCOVA model will be expanded to include this significant variable as a covariate or factor in the model. Statistical models with interaction terms or subgroup analyses will be explored if appropriate to help further understand the treatment effect.

8.2 Secondary Efficacy Analysis (3M Oral Rinse vs. Vehicle Control)

Secondary efficacy hypotheses will be tested in a fixed sequential order, conditional on the expected significance of the primary analyses. The fixed-sequence testing of a subsequent hypothesis will be conditional on the significant testing of the previous hypotheses. The sequential order of testing will preserve the one-sided alpha of 0.025 and is outlined as follows:

1. Modified Quigley-Hein Plaque Index (PI) at 3 months between the 3M Oral Rinse and the Placebo Vehicle Control
2. Löe-Silness Gingivitis Index (GI) at 3 months between the 3M Oral Rinse and the Placebo Vehicle Control
3. Eastman Interdental Bleeding Index at 6 months between the 3M Oral Rinse and the Placebo Vehicle Control
4. Eastman Interdental Bleeding Index at 3 months between the 3M Oral Rinse and the Placebo Vehicle Control.

8.3 Secondary Efficacy Analysis with all Four Treatment Groups

All pairwise comparisons involving the four treatment groups, namely the 3M Oral Rinse, vehicle control, PerioShield predicate and water control groups, will be carried out with the PI and GI endpoints at 6 months. Analyses will be performed using the ANCOVA model and the p-values will be adjusted using the Tukey method.

9. SAFETY ANALYSIS

Safety variables will include adverse events, abnormal soft and hard oral tissue, and reported side effects.

9.1 Adverse Events

All adverse events (AEs) occurring after initiation of study treatment (treatment emergent AEs) will be summarized by severity, relatedness, overall and treatment group. AEs occurring during the screening period will be summarized separately. All AEs will be classified using the MedDRA dictionary.

All verbatim descriptions will be listed for all AEs, along with information regarding onset, duration, severity, relationship to treatment, and action taken.

Serious adverse events will be summarized and presented according to nature, time to onset, duration, relationship to treatment and outcome.

First, an overall Fisher's Exact test will be carried out to compare each safety parameter among all four treatment groups. If this overall test is significant (p-value ≤ 0.05), pairwise comparisons will be carried out to compare the 3M Oral Rinse and the vehicle control. Each active vs vehicle comparison will be carried out at an alpha level of 0.05.

9.2 Reported Side Effects

The incidence of reported side effects from the subject diary will be summarized by week (from dosing) and treatment group.

First, an overall Fisher's Exact test will be carried out to compare each safety parameter among all four treatment groups. If this overall test is significant (p-value ≤ 0.05), pairwise comparisons will be carried out to compare the 3M Oral Rinse and the vehicle control. Each active vs vehicle comparison will be carried out at an alpha level of 0.05.

9.3 Soft and Hard Oral Tissue Assessment

The incidence of normal and abnormal soft and hard tissue findings will be summarized by visit and treatment group. Shifts in findings from normal/abnormal to normal/abnormal observed within the first 3 months will be summarized by treatment group. The worst assessment up to the 3-month visit will be used in this summary. In addition, shifts in findings from normal/abnormal to normal/abnormal between the baseline visit and the last available visit will be summarized by treatment group.

Revision History

Revision	Details:
1	To establish statistical analysis plan before study enrollment.
2	(1) To clarify when per protocol dataset will be applied (Section 5.3). (2) To clarify dosing compliance computation (Section 7.4) (3) To indicate the classification of AEs using the MedDRA dictionary (Section 9.1) (4) To clarify the incorporation of the unscheduled oral assessment data in planned summaries (Section 9.3).