**ORAL HYGIENE AND MATERNITY OUTCOMES MULTICENTER STUDY (OHMOM):**

**A RANDOMIZED CONTROLLED CLINICAL TRIAL TO EVALUATE LATE FIRST TO MID-SECOND TRIMESTER INTRODUCTION OF ADVANCED DAILY ORAL HYGIENE ON GINGIVITIS AND MATERNITY OUTCOMES**

**Study Number 2011001**

**Detailed Statistical Analysis Plan**


**Prepared by:** Julie Grender, PhD
Research Fellow Statistician — 27-March-2015 Date

**Reviewed by:** Dave Ramsey, MS
Senior Scientist Statistician — 27-March-2015 Date

## Table of Contents

## 1. INTRODUCTION

The purpose of this analysis plan is to provide a detailed description of statistical analyses that will be performed to produce the clinical report for Study 2011001. Analyses specified here are pertinent to the objectives stated in the protocol.

## 2. STUDY OBJECTIVES AND DESIGN

This is a multicenter randomized, controlled, single-blind, 2-treatment, parallel group clinical trial to assess gingivitis and maternity outcomes in pregnant women who undertake an advanced daily oral hygiene program. Up to 750 subjects were to be enrolled to complete the study with a target population of approximately 600 evaluable subjects. Subjects were adult female volunteers with moderate-to-severe gingivitis who were in their late first trimester or early-to-mid second trimester of pregnancy (8-24 weeks of pregnancy).

Sufficient screening was conducted to enroll approximately 300 evaluable subjects at each of two study centers. After informed consent was obtained, the baseline measures were recorded and subjects were randomized to either: 1) an advanced oral hygiene regimen consisting of marketed rechargeable toothbrush, toothpaste, mouth rinse and floss plus oral hygiene education counseling (Regimen), or 2) regular oral hygiene with an anticavity toothpaste, regular manual toothbrush and floss (Control).

Separate randomizations were generated for each study site with approximately equal numbers of subjects randomized to both groups. Each randomization was balanced for history of pre-term birth, smoking status and number of bleeding sites.

Study-related evaluations were scheduled, where possible, as part of routine prenatal visits. Dental examinations were conducted at Baseline before study product assignment, and monthly thereafter for 3 consecutive months of study product use. Product distribution took place at the monthly visits or via shipment when the subjects were not present for the monthly evaluation. Adverse events were collected from examination and interview. Oral hygiene and prenatal knowledge and practices were assessed by survey at Baseline and Month 3 visit.

The monthly gingivitis endpoints were the primary outcome measure, with the Month 3 evaluation being the primary time point of interest. Maternity outcome data (gestational age, birth weight, etc.) were abstracted separately from medical records as secondary variables with gestational age being of most importance. An interim analysis of the primary and secondary outcomes was performed by an independent statistician, and the target population size was adjusted. An independent Data and Safety Monitoring Board has assessed study design, conduct and outcomes at prescribed intervals with respect to study integrity, subjects' safety and evaluability.

## 3. CHANGES TO THE DESIGN/PLAN

An interim analysis was planned in the protocol and was limited to sample size calculations based on the observed variability after the interim database lock on February 11, 2013. The database had *approximately* 184 total subjects who had delivery data (approximately 100 from UAB and 84 from Penn).

The final determination of treatment differences will be evaluated by the final and locked database without an adjustment in Type I error rate since the purpose of the interim analysis was only to re-estimate the sample size. No hypothesis testing or treatment difference information was obtained at the interim. Furthermore, the treatment code identity was not revealed to the statistician who carried out the analysis or to any study or Sponsor personnel, but it was available to the Data and Safety Monitoring Board (DSMB) if needed. The DSMB chose to remain blinded at the interim analysis stage.

On February 11-12, 2013 the interim study data was provided to a designated independent (off-site) statistician who was not involved in the study execution or its conduct. Additionally, the independent statistician was also not involved in the evaluability of the final data. The independent statistician received the treatment code via a secure transfer from another identified statistician not supporting this project.

Sample size calculations were carried out for the birth outcome endpoints, Gestational Age and Birth Weight. Subsequently, the projected sample size from these analyses (as well as the current sample size) was used to compute an expected difference between products for the gingivitis endpoint, Total Number of Bleeding Sites at Month 3. The subject populations used in these calculations included the following at both sites together and for each study site separately:

- **Evaluable subjects (N=163)**: Subjects who were provided with a sufficient supply of study products to last through at least 2/3 (~67%) of the study participation duration, and had not violated any major protocol criteria. See Interim Analysis Evaluability memo for list of non-evaluable subjects for Birth Outcomes and/or Bleeding endpoint. If a subject did not present for the Month 3 visit, the gingivitis data was assumed to be missing. The evaluable populations were different for Birth and Gingivitis endpoints.
- **Intent-to-Treat subjects, ITT (N=188)**: Subjects who were randomized and used the study products. For the Bleeding endpoint, we did not include those subjects graded by examiner #129 (at Baseline) at The University of Alabama, which were considered as having inaccurate Loe Silness measurements (e.g., '*instrument/examiner error*').

The sample size calculations were performed on Gestational Age and Birth Weight using 80% power in 2-sided testing with Type I error rate of 5% in the statistical model specified below. The delta between the treatment products was assumed to be 3-5% (for Gestational Age) and 5-10% (for Birth Weight) of the Control group (group with lower average Gestational Age and Birth Weight values) adjusted mean, and the error term variance was estimated from the mean square error from the specified models.

Estimating gestational age to be 38.1 weeks for the control product means that a 2.5% difference between treatments results in 0.953 weeks (~7 days), and the 3% difference corresponds to 1.143 weeks (~8 days) while the 5% difference equals 1.905 weeks (~13.3 days). Using this data, the sample sizes per treatment group were calculated and presented below for the two populations using the MSE from the specified models:

| GEST AGE | SAMPLE SIZE per GROUP | |
|---|---|---|
| % Diff Btwn Products | Evaluable Subjects | ITT Subjects |
| 2.5% | 354 | 334 |
| 3% | 246 | 232 |
| 5% | 89 | 84 |

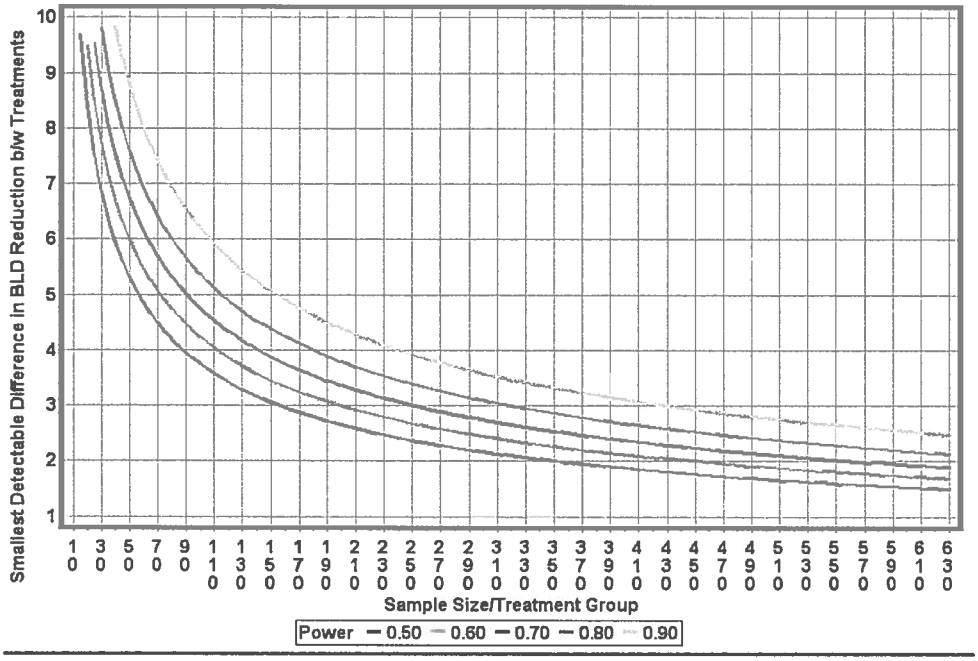
Estimating average birth weight to be 6.87 lbs for the control product among evaluable subjects the 5% difference between treatments results in 0.343 lbs while the 10% difference corresponds to 0.686 lbs. Estimating average birth weight to be 6.73 lbs for the control product among ITT subjects, the 5% difference between treatments results in 0.337 lbs while the 10% difference gives 0.673 lbs. Using this data, the sample sizes per treatment group were calculated and presented below for the two populations using the MSE from the specified models:

| BIRTH WT | SAMPLE SIZE per GROUP | |
|---|---|---|
| % Diff Btwn Products | Evaluable Subjects | ITT Subjects |
| 5% | 439 | 452 |
| 10% | 110 | 114 |

For Bleeding Site Reduction at Month 3, the estimated detectible difference between groups that would lead to a statistically significant result was calculated using 80% power, 2-sided testing, 5% Type I error, projected sample size (per group) from Gestational Age calculations (above) and the mean square error from the gingivitis model.

| GEST AGE | DETECTIBLE DIFFERENCE (n per group) | |
|---|---|---|
| % Diff Btwn Products | Evaluable Subjects | ITT Subjects |
| 2.5% | 2.9 (n=354) | 2.9 (n=334) |
| 3% | 3.4 (n=246) | 3.5 (n=232) |
| 5% | 5.7 (n=89) | 5.9 (n=84) |

In addition to the above detectible differences for given sample sizes from the Gestational Age calculations, below is a power curve with different sample sizes and detectible differences.




## Statistical Models for Interim Analysis

The statistical model to determine treatment differences for mean Gestational Age and mean Birth Weight was an analysis of covariance (ANCOVA) model with previous preterm (< 37 weeks) birth (Y/N), smoking during pregnancy (Y/N), mother's weight at delivery (in pounds), Study Site, and Treatment (A/B).

The model carried out on reduction of Number of Bleeding Sites (Baseline minus Month 3) was an analysis of covariance (ANCOVA) with the baseline Number of Bleeding Sites as the covariate and also included Treatment, Study Site and Treatment by Study Site interaction as factors.

The outcome of the interim analysis could have resulted in early termination of the research due to unexpected safety findings (input provided by the DSMB), continuing the study with the planned sample size or adjusting the sample size. The outcome of this interim analysis resulted in an increase in the sample size of 150 additional subjects (75 per study site).

## 4.0 DATA MANAGEMENT/QUALITY ASSURANCE

Prior to statistical analyses, the following steps were taken to ensure the accuracy, consistency, completeness, and reliability of the data:

- Electronic database review,
- Data management quality control checks (edit checks), and
- Additional review by the clinical/contractors of 100% of data (clinical source to database) from a random sample of 5-10% of subjects at each site.

## 5.0. DATA COLLECTED AND EVALUABILITY

## 5.1 Demographic and Baseline Characteristic Measurements

The following general demographic variables were collected during this study: Age, Race, Marital Status, Education, Income, Employment and Insurance variables. Additional baseline obstetrical variables were collected: Last Menstrual Period, Estimated Delivery Date, In-vitro Fertilization, Date of first Ultra Sound and corresponding Gestational Age reading, Previous Pregnancy (full, preterm, abortions), Smoking Status during pregnancy, Alcohol Use during pregnancy, and Other drug use. Additional variables collected were the subject's height and weight at baseline and subject's weight at each monthly visit and at delivery. The subjects also took an oral hygiene questionnaire at the Baseline and Month 3 visits.

At Baseline each subject had a Loe-Silness Gingival Index exam taken, which was used to assess number of bleeding sites for inclusion into the study.

Gingival crevicular fluid, plaque bacteria and blood samples were taken at the Baseline visit and will be analyzed separately outside of this protocol by the study sites.

## 5.2 Gingivitis Evaluations

Löe-Silness Gingivitis Index (GI)

The Löe-Silness GI was used to evaluate the gingiva based on color, consistency, and bleeding on probing. The entire dentition, with the exception of the third molars, was evaluated. For each tooth, six gingival areas (distobuccal, buccal, mesiobuccal, mesiolingual, lingual, and distolingual) were evaluated using adequate light, a mouth mirror, and a periodontal probe. Prior to evaluation, the teeth and gingiva were air dried as required to provide adequate visability. The probe was inserted about 1mm into the gingival sulcus and passed from interproximal to interproximal. One aspect (either facial or lingual) of each tooth in a quadrant was first "skimmed" and then graded before passing to the next quadrant. Each of the six tooth surfaces was given a score of 0-3. A subject's full mouth GI score was determined by summing the scores and dividing by the number of sites examined. Criteria for the gingivitis scores are as follows:

| Score | Criteria |
|---|---|
| 0 | Normal gingiva. |
| 1 | Mild inflammation – slight change in color, slight edema. No bleeding on probing. |
| 2 | Moderate inflammation – redness, edema, and glazing. Bleeding on probing. |
| 3 | Severe inflammation – marked redness, edema. Ulceration. Tendency to spontaneous bleeding. |
| 8 | Tooth not scored. |
| 9 | Tooth missing. |

Due to the nature of the measurement, the Löe-Silness GI assesses both inflammation and bleeding. Gingival bleeding is derived from the GI scores at each site by assigning the site's Bleeding Score to 1 if the GI score is either a 2 or 3, and assigning it a value of 0 if the GI score

is 0 or 1. The full mouth bleeding score is determined by summing the Bleeding Scores of all scored sites.

When the tooth was missing or it was non-gradable at the Month 3 visit then the corresponding tooth was excluded from the previous visits (including Baseline) whole mouth gingivitis score or number of bleeding sites calculation.

Probing Pocket Depth Measurements

Local supragingival scaling may have been administered as needed to allow periodontal probing. Periodontal pocket was measured with a periodontal probe using standard methods. For each tooth, six gingival areas (distobuccal, buccal, mesiobuccal, mesiolingual, lingual, and distolingual) were scored using adequate light, a mouth mirror, and a periodontal probe. Prior to scoring, the teeth and gingiva may have been air dried as required to provide adequate visibility. The probe was inserted to the depth of the pocket, and pocket depth was measured (in mm).

## 5.3 Birth Outcome Evaluations

For each subject, pertinent medical records were reviewed after delivery to determine maternity outcomes such as Expected Delivery Date and Birth weight. In addition, other delivery details such as delivery route, pregnancy complications, medications, etc. and neonate variables such as Apgar score, length, evidence of congenital anomalies, etc., were recorded as additional information.

## 5.4 Safety Measurements

Safety was assessed by the collection of volunteered adverse events experienced during the study period or observed by the examiner during the gingivitis exams.

The description of an AE included:

- Onset and resolve dates,
- Expectedness of the AE,
- Severity (Mild, Moderate, or Severe),
- Outcome (if known, whether or not AE was resolved or had a fatal outcome),
- Investigator's opinion as to the causality from the study medication (Doubtful, Possible, or Probable),
- Action taken regarding study medication,
- Indication of serious outcome (for each AE, the Investigator must determine if it is an SAE),
- Category of the AE (Non-Oral or Oral Category), and
- Whether the AE was reported by the subject or observed by the examiner.

## 5.5 Rules and Conventions for Analyses

All data will be checked for accuracy, completeness, and compliance with the study protocol prior to breaking the treatment blind. The statistical analyses are the responsibility of the Sponsor's Department of Global Statistics and Data Management Department. In addition, a statistical analysis plan will be completed by the Sponsor prior to breaking the treatment blind. All statistical analyses will be performed using SAS®.

There are 2 evaluable populations, one for Birth outcomes and one for Gingivitis outcomes, since the evaluability criteria are different for the two sets of variables. A subject can be non-evaluable for partial data (e.g., birth outcomes only, all monthly gingivitis outcomes, or only specific monthly gingivitis outcomes). The criteria for evaluability are outlined in Section 5.6. There will be one population for Intent-to-Treat, which is defined as those subjects who were randomized and received study product

## 5.6 Subject Evaluability Criteria for Randomized Subjects

### 5.6.1 Evaluability for Birth Outcomes

Subjects' birth outcome data (e.g. gestational age (GA), birth weight) will be assessed for evaluability based on the following criteria:

1) Drops: Declare as non-evaluable those subjects who were randomized and dropped from the study
2) Visit and Product Criterion
   For purposes of evaluability, a *missed visit* occurred when an eligible subject (still pregnant) did not attend a visit.
   a) Visit Criterion: Declare as non-evaluable those subjects who *missed* 2 or 3 eligible post-baseline monthly visits.
   b) Product Criterion *(% of entire pregnancy with product)*: If the supply of products was less than 33.3% of the duration of the subject's entire pregnancy, then that subject will be non-evaluable.
3) Failed Entrance Criteria: Subjects who failed any entrance criteria at Baseline will be non-evaluable
   - Autoimmune disease
   - Type 2 Diabetes
   - > 24 weeks GA
   - < 30 bleeding sites
4) Failed Continuance Criteria: Subjects who had any of the following will be non-evaluable
   - Evidence/documentation of consistent non-compliance with assigned test product use (either from general comments or continuance criteria)
   - Elective termination of pregnancy
5) Pre-existing Condition: Subjects who had any of the following diseases/conditions or treatments that may impact GA will be non-evaluable
   - Uterine Abnormality (medically-significant)
   - Rh Incompatibility (untreated)
   - Previous Classical C-Section
   - Sickle Cell Disease
6) Adverse Event: Subject who had the following treatments that may impact GA will be non-evaluable

- Emergency C-Section with concomitant appendectomy due to stump appendicitis

7) Congenital Anomaly: Subjects who had the following medically significant congenital anomalies that may impact GA will be non-evaluable
   - Trisomy 21
   - Fetal Hydrocephalus
   - Tetralogy of Fallot

### 5.6.2 Evaluability for Gingivitis (GI) Outcomes

Subjects' gingivitis outcome data (e.g. Löe-Silness, Number of Bleeding Sites, Pocket Depth) will be assessed for evaluability based on the following criteria:

1) Visit and Product Criterion
   For purposes of evaluability, a *missed visit* occurred when an eligible subject (still pregnant) did not attend a visit.
   a) Visit Criterion: All monthly GI data for subjects who *missed* 2 or 3 eligible post-baseline monthly visits will be non-evaluable.
   b) Product Supply Criterion: (for those who missed 0 or 1 post-baseline visit): If product supply was not received (via shipment or in person) at the previous monthly visit, then the current monthly GI visit will be non-evaluable.
2) Examiner Error: Subjects who had their baseline examination performed by examiner #129 will have all of their monthly GI data (BL, M1, M2, M3) declared non-evaluable.
3) Failed Entrance Criteria: Subjects who had any of the following at Baseline will have all of their monthly GI data declared as non-evaluable
   - Autoimmune disease
   - Type 2 Diabetes
   - > 24 weeks GA
   - < 30 bleeding sites
4) Protocol Deviation: Subject's monthly GI data where any of the following occurred will be declared non-evaluable
   - Multiple GI examinations at a single scheduled visit (2 examinations on 1 day) or same examinations repeated on 2 separate days
   - GI examination performed prior to examiner calibration
5) Failed Continuance Criteria: Subject's monthly GI data where there is evidence of the following will be non-evaluable
   - Insufficient product availability to assure test product compliance (from general comments or Continuance Criteria)
   - Evidence/documentation of consistent non-compliance with assigned test product use (either from general comments or continuance criteria)
6) Intervening dental procedures that might affect GI assessment will be non-evaluable
7) Oral topical antibiotics within 1-2 weeks of monthly GI visit (Peridex, essential oils, triclosan, chlorhexidine) will have that GI visit data declared non-evaluable.

## 6.0. Statistical Methods and Determination of Sample Size

All statistical tests will be 2-sided tests except where noted; and any treatment difference p-values less than 0.05 will be cited as statistically significant. The analysis for the evaluable population will be considered as the primary population and the Intent-to-Treat (ITT) analyses will be considered secondary.

The primary test will be the treatment effect test for 3-month change from baseline for Löe-Silness. No adjustments for multiple testing will be employed.

All statistical tests will be carried out using SAS® version 9 (SAS Institute Inc., Cary, NC).

### 6.1 Subject Disposition

The number and percent of subjects who were enrolled, randomized, randomized and then dropped (reasons for drops), and completed will be summarized.

### 6.2 Demographic and Baseline Characteristics Summaries

Demographic variables will be summarized across all evaluable subjects {Pregnancy Outcome and Gingivitis Outcome (Month 3) populations} and ITT subjects by: 1) treatment group and 2) treatment group within site. For variables that are categorical in nature (race, marital status, insurance, etc.), the number and percentage of subjects in each category will be presented by site and treatment group. For variables that are continuous in nature (age, baseline bleeding, etc.), the mean, standard deviation, minimum and maximum values will be presented for each treatment, site and overall.

The Medical History data collected at Baseline will be summarized by treatment group and the Obstetrical History will be summarized and tested by treatment to assess balancing across groups.

Comparisons between treatment groups will be performed for categorical variables with Pearson's Chi-square, Fisher's Exact test or Cochran-Mantel Haenszel test (for ordinal responses) as appropriate, and comparisons of continuous variables will be assessed using a two-sample t-test or Wilcoxon Rank Sum test.

These comparisons will be done to assess to degree to which the randomization procedure balanced the treatment groups with respect to the variables of interest.

### 6.3 Gingivitis Endpoints

**Null Hypothesis**: There is no difference between treatment groups with respect to the mean gingivitis change from baseline when an adjustment is made for the respective baseline gingivitis score and other demographic parameters.

**Alternative Hypothesis**: There is a difference between treatment groups with respect to the mean gingivitis score change from baseline when an adjustment is made for the respective baseline gingivitis score and other demographic parameters.

Statistical analyses for gingivitis efficacy endpoint will be based on whole-mouth average Löe-Silness change from baseline scores (Baseline score minus post treatment score) and will be considered the primary variable, with Month 3 the primary time point of interest. An analysis of covariance (ANCOVA) will be performed to determine treatment differences on the whole mouth average gingivitis reduction scores with the baseline gingivitis score as the covariate and also including study site, treatment, and the interaction of site and treatment as factors in the model. The interaction of treatment with the covariate will also be assessed for significance. If the interactions are not significant at the 10% level then they will be eliminated from the final model.

Additionally, 95% confidence intervals will be generated on the treatment difference for the average change from baseline scores. Separate analyses will be performed for each visit. The within-treatment difference from baseline gingivitis score will also be tested versus zero for each visit using the adjusted means from the ANCOVA model

The same analyses will also be generated for the reduction of Total Number of Bleeding sites and Pocket Depth reduction from baseline at each visit.

Testing for normality of the gingivitis endpoints will be carried out using Shapiro-Wilk test and residual plots. For non-normal data, a transformation such as taking the natural log (e.g., Total Number of Bleeding Sites) will be implemented or a non-parametric analysis will be carried out.

In addition to the ITT and Evaluable populations, the GI analysis will be carried out on the Intent-to-Treat subjects without examiner #129 data.

Additional subgroups may be defined for analysis on the gingivitis data, such as the following:

- By Study Site
- Subjects who had the same examiner for Baseline exam as for their post-baseline exam
- Subjects with the highest (top 50%) and Lowest (lowest 50%) Number of Bleeding Sites at Baseline
- By Age (< median age vs. ≥ median age)
- Subjects who completed all monthly visits
- By Gestational Age at enrollment (8 to ≤12 weeks, >12 to ≤16 weeks, >16 to ≤20 weeks, >20 to ≤24 weeks)

Oral Hygiene questionnaire data will be summarized by treatment group at Baseline and Month 3 visits.

Additional analyses may also be performed in order to more fully understand the data.

### 6.4 Birth Outcome Endpoints

**Null Hypothesis**: There is no difference between treatment groups with respect to the mean birth outcome when an adjustment is made for baseline demographic variables (e.g. smoking, evidence of preterm birth, etc.)

**Alternative Hypothesis**: There is a difference between treatment groups with respect to the mean birth outcome when an adjustment is made for baseline demographic variables (e.g. smoking, evidence of preterm birth, etc.)

Statistical analyses to determine treatment differences for mean Gestational Age and mean Birth weight will be analyzed using an analysis of variance (ANOVA) model with demographic factors such as previous pre-term birth (< 37 weeks; Y/N), self-reported smoking during pregnancy (Y/N), mother's weight at delivery (in grams), Study site and treatment. Interactions of these demographic variables with treatment will also be investigated for inclusion into the model. If the interactions are not significant at the 10% level then they will be eliminated from the final model. 95% confidence intervals will also be generated on the treatment difference of the mean gestational age and mean birth weight.

If any of the data does not satisfy the normality criterion, then transformations will be considered to improve normality or analogous nonparametric methods will be employed.

In addition, gestational age will be categorized as preterm (gestational age < 37 weeks) or term (gestational age ≥ 37 weeks) for subsequent analyses and tested for treatment differences using logistic regression with the same demographic factors as described above. Gestational age will also be analyzed for treatment differences as a time-to-event variable using the log-rank test (and Kaplan Meier curves) and Cox Regression (adjusting for demographics/covariates). Other categorizations may also be used to identify early preterm births: 20 to < 34 weeks (early preterm), 34 to <37 weeks (late preterm), ≥ 37 weeks; or low birth weight (< 2500 grams).

The distribution of birth weight will also be generated within the following gestational age categories (< 34, 34 to <35, 35 to <36, 36 to <37, 37 to <38, 38 to <39, ≥ 39 weeks) for each treatment group using box plots.

Additional subgroups may be defined for analyzing birth outcome data, such as the following:

- By Study Site
- By Race (Caucasian, Black, Other)
- By Gestational Age at enrollment (8 to ≤12 weeks, >12 to ≤16 weeks, >16 to ≤20 weeks, >20 to ≤24 weeks)
- Women with no labor delay medication or procedures.
- By Age group (< median age vs. ≥ median age)
- Pregnancy History (previous pregnancies (Y/N)

Additional analyses may also be performed in order to more fully understand the data.

### 6.5 Concomitant Medications

Listings of concomitant medications will be presented in an appendix.

### 6.6 Compliance and Exposure

Exposure and compliance will be summarized by treatment groups.

### 6.7 Safety Analyses

All subjects who were randomized and received product (intent-to-treat) will be included in the safety analysis. Adverse events (AEs) will be summarized separately by treatment group.

The AEs will be summarized by tabulating the frequency of subjects with one or more adverse events. The number and/or percentages of adverse events will also be reported. An appendix will list all subjects who incurred adverse events, whether serious or not, during the study by treatment group and subject.

Adverse event data will be presented in the following tables:

- Overall summary of the adverse events by treatment group including the following: number of subjects who reported at least one AE, number of AEs reported, number of subjects who withdrew due to AEs, serious AEs, and number of subjects with possible/probably related AEs.

- Summary of the most common adverse events by MedRA terms and treatment group.

All adverse events of the same kind, i.e., AEs with the same MedRA classification term will be counted only once for each subject.

Pregnancy Complications, Labor and Delivery information and Neonatal Data will be summarized by treatment group.

### 6.8 Determination of Sample Size

The initial selection of sample size (n=300/group) was based on variability of gestational age from preliminary data in a similar study design without a randomized control group. The final sample size was re-estimated with more accurate data from this exact design with the current study population from the interim analysis.

With a sample size of 300/group and an estimate of variability of 4.1, a difference in gestational age between the two groups of approximately 1.1 weeks with power of 90% and carrying out two-sided testing with an $\alpha$=0.05 can be detected. Similarly, for the Löe-Silness gingivitis endpoint, a difference between treatments of 0.05 using an estimate of variability of 0.185 with a power of at least 90% can be detected.

An interim sample size re-calculation was conducted after approximately 200 subjects were completed (approximately 100 from each study site). Based on the interim analysis, the sample size was adjusted by adding approximately 150 subjects (75 per site) providing a total of 750 subjects overall: to i) replace non-evaluable subjects, ii) account for increased variability from the interim analyses compared to the initial estimate of variability and iii) increase power for some subset analyses (See Section 3.0).