STATISTICAL ANALYSIS PLAN

Protocol Number: DRM01B-ACN03

Study Title: A Randomized, Double-Blind, Vehicle-Controlled, Efficacy and Safety Study of Olumacostat Glasaretil Gel in Subjects with Acne Vulgaris

Development Phase of Study: 3

Sponsor: Dermira, Inc.
Sponsor Contact: Beth Zib

Statistical Analysis Plan based on Protocol Version: 06 October 2016

Statistical Analysis Plan Date: 19 January 2018
Statistical Analysis Plan Version: v2.0

Authored by:

SIGNATURE: Brian Armstrong DATE: 22 Jan 2018
Brian Armstrong, M.S.
Associate Director of Clinical Data Standards
QST Consultations, Ltd.

Reviewed by:

SIGNATURE: Laura Hatsis DATE: 22JAN2018
Laura Hatsis, M.S.
Manager of Statistics
QST Consultations, Ltd.

Approved by:

SIGNATURE: DATE: 22Jan 2018
Beth Zib
Clinical Lead
Dermira, Inc.

SIGNATURE: DATE: 22JAN 2018
Ramanan Gopalan, Ph.D.
Biostatistician
Dermira, Inc.

Revisions to the Statistical Analysis Plan described herein must be approved through a formal written amendment with the exception of minor editorial changes to tables, figures, or listing shells and any necessary textual clarifications for programmers that do not affect the stated analysis variables, study endpoints, or statistical methods.

Statistical Analysis Change History

Version	Date	Summary of Changes	Author
1.0	14SEP2017	Original document	Brian Armstrong
2.0	19JAN2017	Addition of subgroup summary for subjects taking oral contraceptives approved for the treatment of acne. Per-Protocol Population changed to be analyzed on the group they were randomized. Addition of treatment-emergent adverse event tables by area treated. Extent of Exposure to also include number of subjects who treated face only, face and chest, face and back, and face, chest, and back. Incidence of Local Skin Reaction of Mild, Moderate or Severe also summarized by visit.	Brian Armstrong

TABLE OF CONTENTS

1. LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

AE(s)	Adverse event(s)
ANCOVA	Analysis of covariance
ATC	Anatomical therapeutic chemical
ECG	Electrocardiogram
ET	Early termination
IGA	Investigator Global Assessment
ITT	Intent-to-treat
IWRS	Interactive web-based randomization system
LSMean	Least squares mean
LSR	Local skin reactions
LSSD	Least squares standard deviation
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
n	Number of observations
N	Number of subjects (sample size)
PP	Per-protocol
Q1	25^{th} percentile
Q3	75^{th} percentile
QST	QST Consultations, Ltd.
QTc	QT interval corrected for heart rate
QTcB	QT interval corrected for heart rate using Bazett's formula
QTcF	QT interval corrected for heart rate using Fridericia's formula
SAE(s)	Serious adverse event(s)
SAP	Statistical Analysis Plan
SAS®	Statistical Analysis System (SAS® Institute Inc., Cary, NC)
SD	Standard deviation
TEAE	Treatment-emergent adverse event
TESAE	Treatment-emergent serious adverse event
TOFA	5-(tetradecyloxy)-2-furancarboxylic acid
WHO-DDE	World Health Organization Drug Dictionary

2. INTRODUCTION

Acne arises from a combination of physiological changes in the skin including altered sebaceous gland cell differentiation, heightened sebum production, localized bacterial colonization and inflammation. Although acne is generally viewed as a benign dermatological disease of adolescence, more severe forms may lead to permanent scar formation, with some of these patients suffering from psychological injury and significant loss of self-worth. Recent scientific advances in the understanding of the complex multi-factorial nature of this common disease offer great opportunity for scientific innovation through selective targeting of key elements of the disease process.

Topical agents routinely prescribed for acne, including antibiotics, retinoids and combinations thereof, often produce only modest therapeutic benefits and do not affect sebum production. A locally-delivered medication that selectively inhibited sebum formation would represent a breakthrough advance in reducing the pathogenic influences of sebum overproduction in acne.

Dermira, Inc. is pursuing the development of olumacostat glasaretil, previously referred to as DRM01B, a pro-drug of TOFA (5-(tetradecyloxy)-2-furancarboxylic acid), sarcosine ester, as a topically applied sebum inhibitor for the treatment of acne vulgaris. Olumacostat glasaretil is a new chemical entity and has been formulated for clinical development as Olumacostat Glasaretil Gel, 5.0%.

Data from the completed Phase 2 dose ranging study, DRM01B-ACN02, and data from study DRM01B-ACN01, have shown that Olumacostat Glasaretil Gel was well tolerated at multiple concentrations. Olumacostat Glasaretil Gel at a strength of 5.0% has been selected as the concentration to take forward in Phase 3 development. Two Phase 3 pivotal trials are planned as part of the Phase 3 development program and will include adults and adolescents (ages 9 years and older) in order to confirm the efficacy and safety of Olumacostat Glasaretil Gel, 5.0% for the treatment of acne vulgaris.

This study, DRM01B-ACN03, is one of the two confirmatory Phase 3 trials being conducted to assess the efficacy and safety of Olumacostat Glasaretil Gel, 5.0% in subjects with acne vulgaris.

3. STUDY OBJECTIVES

The objective of this study will be to assess the efficacy and safety of Olumacostat Glasaretil Gel at a concentration of 5.0%, compared to Olumacostat Glasaretil Gel, Vehicle when applied twice daily for 12 weeks in subjects with acne vulgaris.

4. STUDY DESIGN

4.1 Overall Study Design

This study is a randomized, double-blind, vehicle-controlled, parallel group study, designed to assess the efficacy and safety of Olumacostat Glasaretil Gel, 5.0% compared to Olumacostat Glasaretil Gel, Vehicle in subjects with acne vulgaris on the face. All subjects will provide informed consent or assent (pediatric subjects) and undergo screening for study eligibility.

Approximately 700 eligible subjects ≥ 9 years of age will be randomized, in a 2:1 fashion, to Olumacostat Glasaretil Gel, 5.0% or Olumacostat Glasaretil Gel, Vehicle treatment, respectively. Study drug will be applied twice daily to the face, and if applicable, to other affected areas on the chest, shoulders, and back, for 12 weeks. Subjects will return to the study clinic at Weeks 1, 4, 8 and 12 (Study Exit). A follow-up phone call at Week 2 will be made to assess safety (adverse events [AEs]) only.

Primary efficacy will be assessed through acne lesion counts (inflammatory and non-inflammatory lesions) and investigator global assessment of acne (IGA). All efficacy endpoints will be assessed as the change from Baseline to Week 12 on the face only.

Safety will be assessed through AEs, local skin reactions (LSRs), serum chemistry, hematology and urinalysis laboratory testing, electrocardiogram (ECG) testing, physical examination and vital signs.

4.1.1 Schedule of Visits and Assessments

The schedule of assessments can be found in Appendix 1 of the protocol.

4.1.2 Method of Assigning Subjects to Treatment Groups

A randomization schedule will be generated by a member of the Statistical Services department at QST Consultations, Ltd. (QST) who is not associated with the conduct or analysis of the study, using a validated system. The randomization list will not be stratified but stratification by site will be completed within the Interactive Web-based Randomization System (IWRS).

At the Baseline/Day 1 visit, qualified subjects will be randomized to treatment using the IWRS. The IWRS will assign a study drug kit number based on the subject's treatment assignment and available inventory at the investigational site. The kit number will be recorded in the electronic case report form. Approximately 467 subjects will be randomized to active treatment and approximately 233 subjects will be randomized to vehicle treatment for a total of approximately 700 subjects. Subjects will be randomized to treatment groups in such a manner to balance treatment allocation within study sites in a 2:1 fashion.

4.1.3 Blinding

The Sponsor, the contract research organization, the investigator, study site personnel and subjects will be blinded to the treatment assignment. The randomization schedule will be kept strictly confidential and accessible only to authorized persons. Only when the study has been completed, the protocol violations determined and the study database locked will the randomization schedule be made available for analysis.

The integrity of this clinical study must be maintained by observing the treatment blind. If an AE occurs which cannot be managed without knowing whether the subject is receiving active study drug or vehicle solution, the IWRS will be used to obtain treatment assignment information. The Medical Monitor must be notified whenever study medication is unblinded, preferably prior to unblinding a subject.

4.2 Determination of Sample Size

The sample size for this study was based on estimates from the DRM01B-ACN02 study observed Week 12 results. Estimates were combined for all three active treatment groups as well as for the two vehicle groups. All power calculations were performed with nQuery Advisor Version 7.0 for a two-sided test at alpha = 0.05.

For inflammatory lesion change a sample size of 300 Olumacostat Glasaretil Gel and 150 Olumacostat Glasaretil Gel, Vehicle subjects has at least 95% power to detect a statistically significant difference with a significance level of 0.05 using the estimated absolute change from Baseline in treatment means of -15.21 and -9.7 for Olumacostat Glasaretil Gel versus its vehicle, respectively, with a standard deviation (SD) of 11.00.

For non-inflammatory lesion change a sample size of 450 Olumacostat Glasaretil Gel and 225 Olumacostat Glasaretil Gel, Vehicle subjects has 95% power to detect a statistically significant difference with a significance level of 0.05 using the estimated absolute change from Baseline in treatment means of -14.73 and -9.70 for Olumacostat Glasaretil Gel versus its vehicle, respectively, with a SD of 16.72.

The success rate for dichotomized IGA is estimated to be 20.5% in the Olumacostat Glasaretil Gel subjects and 9.6% in vehicle subjects. A sample size of 444 Olumacostat Glasaretil Gel and 222 Olumacostat Glasaretil Gel, Vehicle will have 95% power to detect a statistically significant difference in proportions of successes. These calculations were performed with nQuery Advisor Version 7.0 for a two-sided continuity corrected chi-square test at alpha = 0.05.

To allow for a slight loss in power due to dropouts and withdrawals as well as to accommodate sufficient power in all three co-primary variables, a sample of 700 subjects randomized to 2 to 1 for Olumacostat Glasaretil Gel (n=467) and vehicle (n=233), respectively, will be enrolled in the trial.

5. CHANGES IN THE CONDUCT OF THE STUDY OR PLANNED ANALYSES

There are no changes in the conduct of the study, but there are minor changes to the planned analyses.

The pooling analysis described in the protocol is updated in the statistical analysis plan (SAP) to reflect the intended analysis. The protocol discusses pooling sites with five or fewer subjects enrolled in each treatment arm. Taking into consideration the randomization ratio, the SAP requires ten subjects in the active treatment arm and five subjects in the vehicle group. The SAP also clarifies the pooling of sites will be performed for each country separately.

As part of the pooling analysis, the treatment effect within investigational site must be investigated. The protocol states the site main effect will be examined by using a one-way analysis of variance (for lesion count variables) or a logistic regression analysis (for dichotomized IGA) with a factor of site. The SAP instead explains site will be incorporated into the primary analysis models. Change from Baseline in lesion counts will be analyzed with an analysis of covariance (ANCOVA) (unranked or ranked) with factors of treatment group, investigational site and treatment group by investigational site interaction, and the respective baseline lesion count variable as a covariate. The dichotomized primary endpoint will be analyzed with a logistic regression with factors of treatment group, investigational site and the interaction term of treatment group by investigational site.

The protocol describes repeated measures sensitivity analyses for the primary efficacy endpoints. Treatment group, analysis center and visit are listed as factors for the repeated measures analyses. In addition to these factors, the SAP includes a factor of treatment group by visit interaction.

A subgroup summary of oral contraceptive used was added. Specifically, subjects will be categorized as Oral Contraceptives Approved for Acne Treatment Usage and No Oral Contraceptives Approved for Acne Treatment Usage.

6. EFFICACY AND SAFETY ENDPOINTS

6.1 Efficacy Endpoints

Primary and secondary efficacy endpoints will be assessed as the change from Baseline to Week 12 on the face only.

6.1.1 Primary Efficacy Endpoints

The primary efficacy endpoints are as follows:

- Mean absolute change in acne lesion counts (inflammatory and non-inflammatory lesions) from Baseline to Week 12;

- Proportion of subjects who achieved ≥2-grade improvement and a grade of 0 or 1, in the IGA from Baseline to Week 12.

6.1.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints are as follows:

- Percent change in acne lesion counts (inflammatory and non-inflammatory lesions) from Baseline to Week 12;
- Proportion of subjects who achieved ≥ 2-grade improvement in the IGA from Baseline to Week 12.

6.2 Safety Endpoints

Safety will be assessed through AEs, LSRs, serum chemistry, hematology and urinalysis laboratory testing, ECG testing, physical examination and vital signs.

7. STATISTICAL METHODS AND ANALYSIS

7.1 General Methodology

All analyses will be performed using SAS® Version 9.3 or later. No interim analyses are planned. Except where noted, all statistical tests will be two-sided and will be performed at the 0.05 level of significance.

Descriptive statistics will be used to provide an overview of the efficacy and safety results. For categorical parameters, the number and percentage of subjects in each category will be presented. For continuous parameters, descriptive statistics will include n (number of observations), mean, SD, median, minimum and maximum. Appropriate inferential statistics will be used for the primary, secondary and exploratory efficacy variables.

The primary method of handling missing efficacy data will be the method of Markov Chain Monte Carlo (MCMC) multiple imputation.

The primary method of dichotomizing the IGA score will be to consider a subject a “success” if the IGA at a post-baseline visit is 0 or 1 and 2 or more grades lower than at Baseline; otherwise the subject will be considered a treatment “failure”.

Reported AEs, medical history terms, prior and concomitant procedures and therapies and washout procedures will be classified on the basis of Medical Dictionary for Regulatory Activities (MedDRA) terminology, Version 19.1. Prior and concomitant medications and washout medications will be classified on the basis of World Health Organization Drug Dictionary (WHO-DDE) terminology, Format B2, Version March 1, 2016.

7.2 Adjustments for Covariates

Baseline inflammatory lesion count will be a covariate for the primary efficacy endpoint for inflammatory acne lesion counts. Similarly, baseline non-inflammatory lesion count will be a covariate for the primary efficacy endpoint for non-inflammatory lesion counts. No other covariates are planned to be used in the analyses for this study.

7.3 Handling of Dropouts or Missing Data

Missing 12 week efficacy data will be imputed and subsequently analyzed. The primary method of handling missing efficacy data will be the method of MCMC multiple imputation which does not rely on the assumption of data missing at random. Additionally, the pattern of missing observations in each treatment group will not influence the missing value estimation in the other because the imputation will be conducted independently for each treatment group.

The multiple imputation process for imputing missing lesion count and IGA data and subsequent analysis will involve 4 distinct phases with these principal tasks:

1. Calculate the number of missing 12-week values for each treatment group to be estimated by MCMC. Let nmiss be the maximum number of missing between the two treatment groups.
2. Create a data set of subjects, one for each treatment group, with observed values and those needing estimation by MCMC. The missing values in each data set will be filled in using the MCMC method '5 x nmiss' times to generate '5 x nmiss' data sets. The resulting data sets for each treatment arm will be combined into one complete data set for each imputation.
3. For each complete data set, compute the necessary derived variables. Each complete data set will be analyzed with the appropriate analysis model. See Sections 7.3.1.1 and 7.3.1.2 for details.
4. The results from these analyses will be combined into a single inference using SAS PROC MIANALYZE.

Sensitivity analyses will be performed on the primary efficacy endpoints to investigate the impact of the method of imputation. The first set of sensitivity analyses will use repeated measures. See Sections 7.3.2.1 and 7.3.2.2 for details.

The second set of sensitivity analyses will use model-based multiple imputation. The model-based multiple imputation process requires the data to have a monotone missing pattern. However, the imputation model will include treatment group and analysis center, and for lesion counts will include a covariate of baseline lesion count. Since treatment group, analysis center and baseline lesion count will be non-missing, the data will have a monotone missing pattern. The model-based multiple imputation process will involve 4 principal tasks:

1. Calculate the number of missing 12-week values to be imputed (nmiss).
2. Missing values will be filled in '5 x nmiss' times to generate '5 x nmiss' complete data sets. The imputation model used will be the analysis model.
3. For each complete data set, compute the necessary derived variables. Each complete data set will be analyzed with the appropriate analysis model. See Sections 7.3.2.1 and 7.3.2.2 for details.
4. Results from these analyses will be combined into a single inference using SAS PROC MIANALYZE.

7.3.1 MCMC Multiple Imputation

7.3.1.1 Lesion Count Variable Missing Data Imputation

A total of 4 random seeds will be needed to use PROC MCMC to impute inflammatory lesion counts and non-inflammatory lesion counts for the two treatment groups. Those 4 random seeds have been pre-specified by using a random number generator:

- Inflammatory Lesion Counts; Olumacostat Glasaretil Gel, 5.0%: Seed= 195361148;
- Inflammatory Lesion Counts; Olumacostat Glasaretil Gel, Vehicle: Seed= 2099102059;
- Non-Inflammatory Lesion Counts; Olumacostat Glasaretil Gel, 5.0%: Seed= 325230329;
- Non-Inflammatory Lesion Counts; Olumacostat Glasaretil, Vehicle: Seed= 936128187.

After imputing with PROC MCMC, the absolute change in lesion counts for Baseline minus the 12-week value will be computed. Each complete data set will be analyzed with an ANCOVA with factors of treatment group and analysis center and the respective baseline lesion count as a covariate. A factor of treatment group by analysis center will be added if necessary, and appropriate modifications will be made should the analysis be based on a non-parametric method. The results from these analyses will be combined using SAS PROC MIANALYZE.

7.3.1.2 IGA Missing Data Imputation

A total of 2 random seeds will be needed to impute IGA for the two treatment groups. Those 2 random seeds have been pre-specified by using a random number generator:

- IGA; Olumacostat Glasaretil Gel, 5.0%: Seed= 322979853;
- IGA; Olumacostat Glasaretil Gel, Vehicle: Seed= 1313477781.

After imputing with PROC MCMC, the dichotomous success rate (clear or almost clear with at least a 2-point change from Baseline) will be computed. The 12-week estimated IGA values will be rounded to the nearest integer value prior to evaluating the success rate. Each complete data

set will be analyzed with a logistic regression with factors of treatment group and analysis center. The results from these analyses will be combined using SAS PROC MIANALYZE.

7.3.2 Sensitivity Analyses

7.3.2.1 Lesion Count Variable Missing Data Imputation

The first sensitivity analysis for absolute change in lesion count will use a repeated measures ANCOVA, with factors of treatment group, analysis center, visit (i.e., Week 4, Week 8) and treatment group by visit interaction and a covariate of baseline lesion count. In this analysis, data from all post-baseline visits will be included with no imputation for missing data.

The second sensitivity analysis will use the model-based multiple imputation method to impute missing data for the absolute change in lesion counts at Week 12. The imputation model will include treatment group, analysis center, and baseline lesion count. A total of 2 random seeds will be needed to impute inflammatory lesion counts and non-inflammatory lesion counts for the two treatment groups. Those 2 random seeds have been pre-specified by using a random number generator:

- Inflammatory Lesion CountsSeed= 1134935092;
- Non-Inflammatory Lesion Counts: Seed= 1601290001.

7.3.2.2 IGA Missing Data Imputation

The first sensitivity analysis for the dichotomized IGA success will use a repeated measures logistic regression model (generalized estimating equations), with dichotomized IGA success as the dependent variable and treatment group, analysis center, visit (i.e., Week 4, Week 8) and treatment by visit interaction. In this analysis, data from all post-baseline visits will be included with no imputation for missing data.

The second sensitivity analysis will use the model-based multiple imputation method to impute missing data for the dichotomized IGA data at Week 12. The imputation model used will be a logistic regression with factors of treatment group and analysis center. One random seed will be needed to impute IGA for the two treatment groups. The random seed has been pre-specified by using a random number generator:

- IGA: Seed= 977235733.

7.4 Interim Analyses and Data Monitoring

No interim analysis or data monitoring is planned for this study.

7.5 Multicenter Studies

The clinical study will be conducted under a common protocol for each investigational site with the intention of pooling the data for analysis. Every effort will be made to promote consistency in study execution at each study site. The study is intended to be conducted in a manner such that a minimum of 15 subjects will be randomized and included in the ITT population (i.e., at least ten subjects in the active treatment arm and five subjects in the vehicle treatment arm) for any investigator. In the event that there are too few subjects in a treatment arm for an investigational site, then the site's data will be combined with other site's data to achieve the desired sample size minimum per treatment arm. Country will be considered when pooling sites, such that sites from the United States will be combined with other sites from only the United States, and similarly for sites from Canada and sites from Australia. The combining of investigator data will be accomplished by taking the investigator with the smallest enrollment and combining it with the investigator with the largest enrollment, restricted to investigational sites who did not meet minimum enrollment. If there is a further need to combine data, then the data of the investigator with the second smallest enrollment will be combined with the investigator's data which had the second largest enrollment (restricted to investigational sites who did not meet minimum enrollment), and so on. This process will continue for all investigators who did not have a minimum of 15 subjects enrolled. The process of combining investigator data that have insufficient subjects per treatment arm will result in redefining the groups of investigators for the purposes of statistical analyses. These combined groups will be referred to as "analysis centers" in the statistical analyses.

The consistency of treatment response will be investigated across the analysis centers subsequent to combining the data as described above. Statistical tests will be conducted to identify if there are extreme analysis centers that could affect the interpretation of common statistical and clinical conclusions. An analysis center by treatment interaction will be included in the primary and secondary variable analyses to test for parallel treatment effect at an alpha level of 0.10. Change from baseline in inflammatory lesions and non-inflammatory lesions will be analyzed with an ANCOVA (unranked or ranked) with factors of treatment group, analysis center and treatment group by analysis center interaction and the respective baseline lesion count variable as a covariate. For the purpose of testing consistency of treatment response, the dichotomized IGA will be analyzed with a logistic regression with factors of treatment group, analysis center, and the interaction term of treatment group by analysis center. Further examination will follow for any variables that have a significant ANCOVA or logistic regression interaction term.

In the event that the ANCOVA or logistic regression interaction (referred to henceforth as the "appropriate test") p-value is ≤ 0.10, a sensitivity analysis that excludes analysis centers with the extreme efficacy result will be performed to determine the robustness of the treatment effect. On the other hand, if all three analyses result in interaction terms with p-values >0.10, then the

conclusions from the pooled data will be considered to be free of the impact of extreme analysis centers.

The first step in conducting a sensitivity analysis is to identify the extreme analysis center or centers that contribute to the statistical significance of the interaction term of the appropriate test. The process involves submitting subsets of analysis centers to the appropriate test and observing the interaction p-value for the subset. Subsets producing interaction p-values >0.10 are considered homogeneous.

The search for an extreme analysis center begins by analyzing all subsets that can be created by excluding one analysis center. If one or more of the subsets result in an interaction p-value >0.10, then the analysis center excluded from the subset with the largest interaction p-value is deemed to be the extreme analysis center.

If all subset interaction p-values are ≤ 0.10, then the process will analyze the interaction for all subsets that can be created by excluding two analysis centers. If one or more of these subsets generate interaction p-values > 0.10, then the analysis centers excluded from the subset with the largest interaction p-value are deemed the extreme analysis centers.

Thus, the process of identifying the extreme analysis centers will continue in a stepwise manner by first excluding one, then two, then three, etc., analysis centers until the appropriate test interaction p-value is > 0.10.

Once the extreme analysis center or centers have been identified, then the treatment p-values of the remaining analysis centers will be computed. Inferences will be drawn from the treatment p-value, as well as any pertinent observations regarding the extreme analysis center or centers. Additionally, it is noted that this process excludes subjects from the analysis in a nonrandom manner and has an unpredictable impact on the power of the treatment effect test. In the event that the treatment effect of the remaining subset is not statistically significant, due consideration of the post hoc aspects of the process will be given when the results are interpreted. Conclusions will be presented by the Sponsor as appropriate to the findings of the sensitivity analysis.

Prior to investigating the treatment effect within the analysis centers, the treatment effect within investigational site will be investigated to determine if the site-to-site variability is such that it could mask the analysis center effects. Thus, prior to pooling, change from baseline in inflammatory lesions and non-inflammatory lesions will be analyzed with an ANCOVA (unranked or ranked) with factors of treatment group, investigational site, and treatment group by investigational site interaction, and the respective baseline lesion count variable as a covariate. The dichotomized primary endpoint will be analyzed with a logistic regression with factors of treatment group, investigational site, and the interaction term of treatment group by investigational site. If any of the analyses are not computationally feasible due to some investigational sites having very few subjects enrolled, the low-enrolling investigational sites will be excluded from the analysis.

7.6 Multiple Comparisons/Multiplicity

The following stepwise process will be conducted for testing the secondary efficacy endpoints in order to control for multiplicity. These tests will be performed for only the intent-to-treat (ITT) population. The testing process will terminate whenever a statistical test for a step is not significant. All subsequent tests for the remaining steps will be considered not significant. The order of testing is percent change in inflammatory lesion count from Baseline to Week 12, percent change in non-inflammatory lesion count from Baseline to Week 12 and the proportion of subjects who are dichotomized to success (minimum 2-grade improvement from Baseline).

7.7 Examination of Subgroups

Subset analyses will be conducted for the ITT population for subgroups based on the following:

- Baseline IGA;
    - Categorized as Baseline IGA = 3 (Moderate) and Baseline IGA = 4 (Severe); due to the inclusion criteria requiring subjects to have either a 3 or 4 IGA score at study entry other IGA response categories will not be presented;
- Gender;
- Age:
    - Dichotomized as less than the median age of ITT subjects and greater than or equal to the median age of ITT subjects;
    - Categorized as less than 18, 18 to less than the median age of ITT subjects and greater than or equal to the median age of ITT subjects; If the median age of ITT subjects is less than 18 then the subgroup will be <18 and >= 18
- Ethnicity;
- Race:
    - Categorized as White and Non-White; subjects indicating more than one race category which includes both white and a non-white race will be summarized under the Non-White subgroup.
- Oral Contraceptives Approved for the Treatment of Acne:
    - Categorized as Oral Contraceptives Approved for Acne Treatment Usage and No Oral Contraceptives Approved for Acne Treatment Usage. All male subjects will be grouped into the No Oral Contraceptives Approved for Acne Treatment Usage category.

Subset analyses will be conducted on the variables absolute change from Baseline in inflammatory lesions and non-inflammatory lesions at Week 12 as well as the dichotomized IGA score at Week 12. These analyses will contain descriptive statistics only.

7.8 Analysis Populations

7.8.1 Randomized Population

All subjects who are randomized to study treatment will be included in the randomized population and will be analyzed according to the treatment group they were randomized.

7.8.2 Intent-to-Treat (ITT) Population

All subjects in the randomized population who are dispensed study drug will be included in the ITT population and will be analyzed according to the treatment group they were randomized. All efficacy analyses will be presented using the ITT population.

7.8.3 Safety Population

All subjects in the randomized population who receive at least one confirmed dose of study drug and have at least one post-baseline safety assessment will be included in the safety population and will be analyzed according to the treatment they received. All safety analyses will be performed using the safety population.

7.8.4 Per-Protocol (PP) Population

All subjects in the safety population who complete the Week 12 evaluation without any significant protocol violations will be included in the PP population and analyzed according to the treatment group they were randomized. The PP population will include subjects in the safety population who do not meet any of the following criteria:

- Violated the inclusion/exclusion criteria;
- Have taken any interfering concomitant medication;
- Did not attend the Week 12 visit;
- Have missed both the Week 4 and Week 8 study visits;
- Have not been compliant with the dosing regimen (i.e. subjects must apply 80-120% of the expected applications of study medication during participation in the study);
- Out of visit window at the Week 12 visit by ±5 days.

Subjects that discontinue treatment or from the study due to an AE related to study treatment or documented lack of treatment effect will be included in the PP population. Prior to breaking the

blind, other additional criteria may be added to the list to accommodate for unforeseen events that occurred during the conduct of the trial that result in noteworthy study protocol violations.

All efficacy analyses will be performed on the PP population.

7.9 Statistical Analysis

All analyses will be performed by QST. All summary tables and data listings will be prepared utilizing SAS® software. The standard operating procedures of QST will be followed in the creation and quality control of all data displays and analyses.

All data collected at scheduled time points will be summarized, regardless of treatment with study drug. Post-treatment assessments for subjects who discontinue treatment but continued to be followed in the study will be included in summaries.

7.9.1 Baseline Definition

Baseline is defined as the last non-missing assessment prior to the first dose of study drug.

7.9.2 Visit Windowing

Data will be summarized based on nominal visit indications with the exception of data captured at early termination and unscheduled visits. Data from early termination and unscheduled visits will be summarized based on mapped visit values. The analysis windows for early termination and unscheduled visits are presented in the following table.

Analysis Windows for Efficacy and Safety Assessments

Scheduled Visit	Target Study Day	Window (Days)
Week 1	8	5 to 18
Week 4	29	19 to 42
Week 8	57	43 to 70
Week 12	85	71 to 98

Data collected at early termination and unscheduled visits prior to study day 5 will not be analyzed, with the exception of those identified as baseline values. Data collected at early termination and unscheduled visits after study day 98 will not be included in analyses.

The definition for the study day included in each study window is defined as below:

Study Day Prior to Day 1 = Visit Date – Day 1 Date

Study Day On or After Day 1 = Visit Date – Day 1 Date + 1.

If an assessment's mapped visit is a visit at which the subject has data from a scheduled visit present, or if no analyses are planned for the assessment at the mapped visit, the data collected at the early termination or unscheduled visit will not be included in analyses.

In the event of multiple values from unscheduled or early termination assessments within an analysis window, the value closest to the scheduled visit target study day will be used for analyses. If two values tie as closest to the time point (for example, one value is before and the other value is after the time point), then the later value will be selected.

Data collected at all visits will be included in the data listings with visit presented as reported by the site.

7.9.3 Subject Disposition

The number of subjects included in each analysis population (randomized, ITT, safety and PP) will be summarized by treatment group, by treatment group within analysis center and by treatment group within investigational site. The number of subjects completed and discontinued (including the primary reason for discontinuation) will be summarized for each treatment group.

Subjects who are excluded from an analysis population will be summarized by the reasons for exclusion. Reasons subjects were not randomized will be summarized for subjects who did not enroll in the study.

7.9.4 Protocol Deviations

Protocol deviations will not be entered into the database. Only deviations leading to exclusion from analysis populations will be identified and summarized.

7.9.5 Demographic and Baseline Characteristics

All baseline summaries will be performed on the ITT, PP and safety populations.

Sex, race, ethnicity and Fitzpatrick skin type will be summarized by counts and percentages. Age will be summarized with descriptive statistics. Summaries of dichotomized age (less than median and greater than or equal to median) and categorized age (less than 18, 18 to less than median and greater than or equal to median) will also be included.

Subjects' baseline characteristics related to efficacy analyses will be summarized and compared between treatment groups to evaluate baseline balance between the treatment arms. Baseline inflammatory lesion count and baseline non-inflammatory lesion count will be compared using a two-way analysis of variance with factors of treatment group and analysis center. Baseline IGA will be compared using a Cochran-Mantel-Haenszel test of row means scores, stratified by analysis center.

Acne vulgaris history, including number of years since onset, anatomical areas affected and prior acne treatments, will be summarized by counts and percentages. If only the year is known for start of acne vulgaris the number of years since onset will be calculated by using January for

month and "01" as day. If only the year and month are known for start of acne vulgaris then onset will be calculated by using "01" as day.

Medical histories will be coded using the MedDRA dictionary and presented in a by-subject listing.

7.9.6 Prior and Concomitant Medications

Prior and concomitant medications will be coded to preferred name and anatomical therapeutic chemical (ATC) classification of ingredients using the WHO-DDE.

Counts and percentages will be provided to summarize the use of prior and concomitant medications other than the study drug reported throughout the study. The number and percent of subjects reporting medications will be summarized by ATC level 2 term and preferred name. Medications which start prior to first dose will be considered prior medications. Ongoing medications and medications ending after the date of first dose will be considered concomitant medications. Medications which are both prior and concomitant will be included in both summaries. Incomplete start and end dates which could be prior to first dose or after first dose will be considered prior to first dose.

7.9.7 Efficacy Analyses

All efficacy analyses will be presented by treatment group for both ITT and PP populations.

7.9.7.1 Primary Efficacy Analyses

Absolute change from Baseline to Week 12 in inflammatory and non-inflammatory lesion counts will be analyzed using either parametric or non-parametric methods consistent with the statistical assumptions required to support the analyses. Specifically, the tests of superiority will be based on an ANCOVA with factors of treatment group and analysis center and the respective baseline lesion count as a covariate or on ranked data submitted to an ANCOVA with factors of treatment group and analysis center and the respective baseline lesion count as a covariate. If the treatment group by analysis center interaction effect is significant at an alpha less than 0.10, then the effect will be included in the model; otherwise it will be omitted.

A skewness test, based on the methods presented by J.H. Zar (1984) [1], will be applied to the residuals resulting from an ANCOVA. A two-sided p-value for the skewness test significant at 0.01 will imply the use of the non-parametric method. If a parametric analysis is indicated, the results of the parametric analysis will be considered the primary analysis. Should a non-parametric analysis be indicated, the absolute or percent changes in inflammatory and non-inflammatory lesions will be rank-transformed prior to submitting them to the ANCOVA. Results of the rank-transformed analyses then will be considered the primary analysis; however, results of the non-ranked transformed analyses will also be presented. That is, the primary

analyses for inflammatory and non-inflammatory lesion counts will be determined by the treatment group by analysis center interaction p-value and the skewness p-value as described in the table below:

Determination of Model for Primary Analyses

Interaction P-Value	Skewness P-Value	Analysis Model
< 0.10	< 0.01	Ranked ANCOVA with factors of treatment, analysis center and treatment by analysis center interaction and the respective baseline lesion count as a covariate
≥ 0.10	< 0.01	Ranked ANCOVA with factors of treatment and analysis center and the respective baseline lesion count as a covariate
< 0.10	≥ 0.01	ANCOVA with factors of treatment, analysis center and treatment by analysis center interaction and the respective baseline lesion count as a covariate
≥ 0.10	≥ 0.01	ANCOVA with factors of treatment and analysis center and the respective baseline lesion count as a covariate

Note that interaction and skewness p-values will be computed for each imputation of multiply-imputed data. Model determination will depend on averaged p-values, computed by finding the mean p-values across all imputations.

The proportion of subjects who are dichotomized to success at Week 12 will be analyzed using a logistic regression with factors of treatment group and analysis center.

The sensitivity analysis for absolute change in lesion count will include a skewness test, based on the methods presented by J.H. Zar (1984) [1]. A ranked repeated measures analysis of covariance with factors of treatment group, visit, analysis center and treatment by visit interaction, and the respective baseline lesion count as covariate will also be included. Values will be ranked across all visits (i.e. visit will be ingored when ranking) and the baseline covariate will not be ranked.

7.9.7.2 Secondary Efficacy Analyses

The percent change from Baseline to Week 12 in inflammatory and non-inflammatory lesion counts will be analyzed using the same ANCOVA method described for the primary endpoint.

The proportion of subjects who are dichotomized to success (minimum 2-grade improvement from Baseline) at Week 12 will be analyzed using the same logistic analysis method described

for the primary endpoint. A stepwise process will be conducted for testing the secondary efficacy endpoints in order to control for multiplicity, refer to Section 7.6 for details.

7.9.7.3 Other Supportive Efficacy Analyses

The percent change and absolute change from Baseline to Weeks 4 and 8 inflammatory and non-inflammatory lesion counts will be analyzed using the same ANCOVA method described for the primary endpoint of absolute change in lesion counts from Baseline to Week 12. The inclusion of p-values in these efficacy analyses is to assist in characterizing the therapeutic efficacy of the active formulation and dosage regimen. No adjustments will be made for multiple comparisons for the supportive efficacy analyses.

Descriptive statistics will be presented for the following parameters by treatment group for both the ITT and PP populations:

- Frequency and percent distributions of the IGA Score at Baseline and Weeks 4, 8 and 12;
- Frequency and percent distributions of the dichotomized IGA Score at Baseline and Weeks 4, 8 and 12;
- Descriptive statistics will be used to summarize inflammatory and non-inflammatory lesion counts at Baseline and Weeks 4, 8 and 12;
- Descriptive statistics will be used to summarize the absolute and percent change in inflammatory and non-inflammatory lesion counts at Weeks 4, 8 and 12;
- Descriptive statistics will be used to summarize the Acne Patient Self-Questionnaire assessments at Weeks 4, 8 and 12. Satifasction questions answered at Week 12 will be summarized with frequency and percent distributions.

7.9.8 Safety Analyses

7.9.8.1 Extent of Exposure

The extent of exposure to study drug in each treatment group will be summarized by amount of study drug used (computed from study drug tube weights) and anatomical areas treated. Facial exposure will be summarized by total number of days of exposure, total number of applications, number of missed applications and number and percentage of subjects who are compliant. A subject will be considered compliant with the dosing regimen if the subject applied 80% to 120% of the expected number of applications while enrolled in the study. Total number of days of exposure will be computed as follows:

Total Exposure = Date of Last Application – Date of First Application + 1.

Dosing deviations reported for areas other than the face will not be considered in compliance calculations. On the date of first application and the date of last application, some subjects were

to apply one dose and others were to apply two doses, depending on the time of the study visit. After the date of first application and before the date of last application subjects were expected to apply two doses. Missed applications will be calculated from dosing deviations which report zero or one application on a given date. Extra doses will be calculated from dosing deviations which report more than the expected number of applications on a given date. The total number of applications will be calculated as follows:

If the Date of First Application is not equal to Date of Last Application, then

Total Applications = 2*(Date of Last Application – Date of First Application – 1)
+ Expected Applications on Date of First Application
+ Expected Applications on Date of Last Application
– Missed Applications + Extra Applications.

If the Date of First Application = Date of Last Application, then

Total Applications = Expected Applications on Date of First Application
– Missed Applications + Extra Applications.

Treatment compliance will be based on the expected number of doses given the treatment period duration. Treatment period duration will be computed from the Baseline/Day 1 visit date and the Week 12/Early Termination (ET) visit date. If a subject does not have a Week 12/ET visit, treatment period duration will be calculated based on the date of last dose or last completed visit date, whichever is later. That is, treatment period duration will be computed from the Baseline/Day 1 date and the end of treatment period date, where the end of treatment period date is given by

- If the subject completed the Week 12/ET visit, the end of treatment period date will be:
    - The date of the Week 12/ET visit if the expected number of applications on the date of last application is one;
    - The day prior to the Week 12/ET visit if the expected number of applications on the date of last application is two;
- If the subject did not complete the Week 12/ET visit, the end of treatment period date will be the date of last dose or last completed visit, whichever is later.

Then treatment period duration may be calculated as:

Treatment Period Duration = End of Treatment Period Date – Day 1 Date + 1.

If the treatment period duration exceeds 88 days, treatment period duration will be considered 88 days. In addition to treatment period duration, missed applications due to treatment-related AEs

will be considered when determining the expected number of doses for each subject. If a treatment-related AE (regardless of location) requires interruption of study drug during the treatment period, missed applications during the AE (onset date to resolution date) will be deducted from the expected number of doses. Therefore, the expected number of applications will be calculated as:

Expected Applications = 2*(Treatment Period Duration – 2)

+ Expected Applications on Date of First Application

+ Expected Applications on Date of Last Application

– Applications Missed Due to a Treatment-Related AE.

If Treatment Period Duration = 1, then

Expected Applications= Expected Applications on Date of First Application

– Applications Missed Due to a Treatment-Related AE.

Percent compliance will be calculated from total number of applications and total number of expected applications as follows:

Percent Compliance = 100*(Total Applications/Expected Applications).

Percent compliance will not be calculated for subjects who are lost to follow-up.

7.9.8.2 Adverse Events

All AEs that occur during the study will be recorded and classified on the basis of MedDRA terminology. Treatment-emergent AEs (TEAEs) are defined as AEs with an onset on or after the date of the first study drug application. AEs noted prior to the first study drug administration that worsen after Baseline will also be reported as AEs and included in the summaries.

All information pertaining to an AE noted during the study will be listed by subject, detailing verbatim term given by the investigator or designee, preferred term, system organ class, onset date, resolution date, severity, seriousness, action taken, outcome and drug relatedness. The event onset will also be shown relative (in number of days) to date of first application.

Treatment-emergent AEs will be summarized by treatment group, the number of subjects reporting a TEAE, system organ class, preferred term, severity, relationship to study drug (causality) and seriousness. When summarizing AEs by severity and relationship, each subject will be counted once within a system organ class or a preferred term by using the event with the highest severity and greatest relationship within each classification.

Serious AEs (SAEs) will be summarized by treatment group, severity and relationship to study drug, and individual SAEs will be listed by subject. In addition, a summary of AEs leading to discontinuation from the study due will be provided.

Treatment-emergent AEs will also be summarized by area treated (subjects who treated face only and subjects who treated face plus truncal area or areas).

No statistical inference between treatment groups will be performed on AEs.

Listings will be presented for all AEs as well as for SAEs and AEs leading to discontinuation from the study.

7.9.8.3 Clinical Laboratory Evaluations

Laboratory test results will be summarized descriptively at Baseline and Week 12. Additionally, shifts from Baseline to Week 12 in laboratory test results based on normal ranges will be summarized with descriptive statistics. Individual laboratory test results will be presented in a by-subject listing.

7.9.8.4 Local Skin Reactions

LSRs include erythema, peeling, dryness, burning/stinging and pruritus. These will be scored as 0 (None), 1 (Mild), 2 (Moderate) or 3 (Severe). LSRs will be summarized by treatment group and visit using descriptive statistics. Each subject's worst post-baseline grade for each LSR will also by summarized. Additionally, the number of subjects who experience mild, moderate or severe LSRs post-baseline will be presented overall, and by visit. By-subject listing of all LSRs, as well as subjects with any severe LSRs will be included.

7.9.8.5 Other Observations Related to Safety

7.9.8.5.1 ECG Measurements

Descriptive statistics by treatment group and visit will be provided for the following ECG parameters: heart rate (HR), PR duration, RR duration, QRS duration, QT duration, QT interval corrected for heart rate using Bazett's formula (QTcB) and QT interval corrected for heart rate using Fridericia's formula (QTcF). Change from Baseline in ECG abnormalities will be summarized using shift tables at the Week 12 visit. Shift tables will be based on the following categories:

- PR Interval: < 100 msec, 100 – 220 msec and > 220 msec;
- QRS Interval: < 50 msec, 50 – 110 msec and > 110 msec;
- QTcF Interval: < 450 msec, 450 – 500 msec and > 500 msec.

The following treatment-emergent ECG abnormalities will be summarized for subjects in the safety population with at least one post-application ECG:

- PR Interval
    - > 220 msec;
    - > 220 msec and Change from Baseline > 25%;
- QRS Interval
    - > 110 msec;
    - > 110 msec and Change from Baseline > 25%;
- QTcF Interval
    - > 450 – 480 msec;
    - > 480 – 500 msec;
    - > 500 msec;
    - Change from Baseline > 30 - 60 msec;
    - Change from Baseline > 60 msec;
    - > 450 - 480 msec and Change from Baseline > 30 - 60 msec;
    - > 450 - 480 msec and Change from Baseline > 60 msec;
    - > 480 - 500 msec and Change from Baseline > 30 - 60 msec;
    - > 480 - 500 msec and Change from Baseline > 60 msec;
    - > 500 msec and Change from Baseline > 30 - 60 msec;
    - > 500 msec and Change from Baseline > 60 msec.

7.9.8.5.2 Vital Signs

Vital signs will be presented by treatment group and visit as observed values and changes from Baseline using descriptive statistics.

7.9.8.5.3 Physical Examination

Physical examination data will be presented in a by-subject listing.

8. REFERENCES

1. Zar, JH (1984), *Biostatistical Analysis, 2nd ed*. Englewood Cliffs, New Jersey: P 118-119.