Statistical Analysis Plan

A Multicenter, Double-blind, Randomized, Parallel-group, Vehicle-Controlled Study to Evaluate the Safety and Clinical Equivalence of a Generic Azelaic Acid Foam, 15% and the Reference Listed Finacea® (Azelaic acid) Foam, 15% in Patients with Moderate Facial Rosacea

Study Number ACTA/AZEL/2015

NCT03287791

Statistical Analysis Plan Approval Date: 06 April 2017

Statistical Analysis Plan	
Study code	ACTA/AZEL/2015
Version	Final
Date	*06/Apr/2017*

A MULTICENTER, DOUBLE-BLIND, RANDOMIZED, PARALLEL-GROUP, VEHICLE-CONTROLLED STUDY TO EVALUATE THE SAFETY AND CLINICAL EQUIVALENCE OF A GENERIC AZELAIC ACID FOAM, 15% AND THE REFERENCE LISTED FINACEA® (AZELAIC ACID) FOAM, 15% IN PATIENTS WITH MODERATE FACIAL ROSACEA.

Study Statistician

<<*Name*>> Date

Sponsor Representative

<<*Name*>> Date

TABLE OF CONTENTS PAGE

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse Event
ANOVA	Analysis of Variance
BMI	Body Mass Index
CI	Confidence Interval
CFR	Code of Federal Regulations
DOB	Date of Birth
eCRF	Electronic Case Report Form
ICF	Informed Consent Form
IGE	Investigator's Global Evaluation
IP	Investigational Product
LOCF	Last Observation Carried Forward
LS	Least square
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent to Treat
mm	Millimeter
PI	Principal Investigator
PP	Per Protocol
PT	Preferred Term
TEAE	Treatment Emergent Adverse Event
SAE	Serious Adverse Event
SOC	System Organ Class
UV	Unscheduled Visit

1. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol "A Multicenter, Double-blind, Randomized, Parallel-group, Vehicle-Controlled Study to Evaluate the Safety and Clinical Equivalence of a Generic Azelaic Acid Foam, 15% and the Reference Listed Finacea® (azelaic acid) Foam, 15% in Patients with Moderate Facial Rosacea" dated January 27, 2016.

Marketed by Bayer HealthCare, Finacea® (azelaic acid) Foam, 15% is a safe and effective topical therapy used for the treatment of moderate facial rosacea. Actavis Laboratories UT has developed a generic formulation of azelaic acid 15% foam and the current study is designed to evaluate the safety and efficacy of this formulation. This randomized, vehicle-controlled, parallel-group, multicenter, double-blind study was designed to evaluate the therapeutic equivalence and safety of a generic Azelaic Acid Foam, 15% and Finacea® (Azelaic acid) Foam 15%, in the treatment of moderate facial rosacea, and to demonstrate the superiority of the efficacy of the two active foams over the vehicle control.

This document will give a description of the planned methods of the analysis.

2. OBJECTIVES

The objectives of this study are:

- to evaluate the therapeutic equivalence and safety of a generic Azelaic Acid Foam, 15% and Finacea® (azelaic acid) Foam, 15% in the treatment of moderate facial rosacea;
- to demonstrate superiority of the efficacy of the two active foams over the vehicle control.

3. STUDY OVERVIEW

3.1 Study Design

Up to 1010 patients will be enrolled in the study to randomize 924 patients to obtain at least 840 modified Intent-to-Treat (mITT) patients in a 1:1:1 ratio (280:280:280 patients, respectively) to each treatment group and 672 (224:224:224) evaluable subjects in the Per protocol (PP) population. Patients will be assigned in a 1:1:1 ratio to treatment with the Test product, Azelaic Acid Foam, 15% (Actavis Laboratories UT), the Reference product,

Finacea® (Azelaic acid) Foam, 15% (Bayer HealthCare) or the Vehicle control (Actavis Laboratories UT) in this multicenter, double-blind, randomized, vehicle-controlled, parallel-group study.

The assigned Investigational Product (IP) will be self-applied topically to the entire facial area (cheeks, chin, forehead, and nose) twice daily for 84 consecutive days. The IP should be gently massaged into the affected areas on the face twice daily, in the morning and evening after the patient's face has been washed with a mild cleanser and patted dry with a soft towel. Patients will be required to use diaries to document study treatments, any missed treatments and the occurrence of all adverse events.

The duration of each patient's participation in the study will be 84 days. Scheduled study visits will include: Visit 1 (Baseline Visit, Day 0), Visit 2 (Day 28 ± 4 days), Visit 3 (Day 56 ± 4 days), and Visit 4 (End of Treatment, Day 84 ± 4 days). A window of ± 4 days will be considered acceptable for each scheduled visit following the baseline visit.

If the Principal Investigator determines that the patient's condition has worsened to the degree that it is unsafe for the patient to continue in the study, the patient may be discontinued from the study as a treatment failure and the patient may be treated using the standard care.

An Unscheduled Visit is allowed at any time, for any reason, if in the Principal Investigator's opinion, it is warranted. If a patient is discontinued from the study during an Unscheduled Visit, the Unscheduled Visit will be referred to as an Early Discontinuation Visit and all procedures scheduled for Visit 4 will be performed. If the Unscheduled Visit is not an Early Discontinuation Visit (i.e., the patient will continue to take part in the study), then the activities performed will depend on the reason for the unscheduled visit and will be left to the discretion of the Investigator. The Investigator should perform any activities necessary to appropriately evaluate the patient at this Unscheduled Visit. If the Unscheduled Visit is due to an AE, the Investigator will determine whether additional visits are needed.

At Visit 1, an informed consent will be obtained from the potential study patient before any study procedures take place. After the patient has been consented, the patient's medical history will then be documented, including the patient's concomitant medications. A urine pregnancy test will be performed for all female patients of child-bearing potential. A baseline

facial rosacea grade will be assigned to the patient using the Investigator's Global Evaluation (IGE) and a baseline lesion count will be performed. The patient will undergo a physical examination, including the recording of vital signs. The patient will be evaluated for signs and/or symptoms of erythema and telangiectasia. The patient will be reviewed against the inclusion/exclusion criteria. Blinded IP will be dispensed to patients who meet all of the inclusion and exclusion criteria. Patients will be instructed on the application of IP and completion of patient diaries.

Patients will return to the study site for Visit 2, Visit 3, and Visit 4. The patient's concomitant medications will be reviewed and documented. A urine pregnancy test will be performed for female patients of child-bearing potential. The patient's facial rosacea will be assessed using the IGE, the patient's lesions will be counted and all results will be documented. The signs and/or symptoms of local irritation will be evaluated for the patient and any other adverse events will be documented. The patient will bring their used IP and their study diaries to each study visit after the baseline visit. Compliance with drug applications will be assessed at each visit after the baseline visit. New can(s) of IP will be dispensed during Visit 2 and Visit 3. In addition, all IP and diaries will be collected from the patient during each scheduled visit or the Early Discontinuation Visit.

3.2 Sample Size

The sample size for this protocol is based on published data (Finacea® Summary Basis of Approval); Finacea® produced a percent reduction for inflammatory lesions of approximately 55% and the Vehicle had a percent reduction of 40% after 12 weeks of treatment.

The anticipated standard deviation for inflammatory lesions is < 47%. A sample size of n1 = 280, n2 = 280, and n3 = 280 evaluable mITT patients (1:1:1 ratio) to provide at least 224 PP patients in each of the Test and Reference treatment groups should provide at least an [redacted] of showing that, the 90% confidence interval on the Test/Reference ratio of percent reduction from baseline is contained within the interval 0.800 to 1.250 in the PP population and at the same time demonstrating that the Test and Reference products are superior to the vehicle in the mITT population.

In this study, enrolled patients will be randomized in the ratio of 1:1:1 to test, reference and vehicle treatments. It is anticipated that 75% of the mITT patients will qualify for the PP population. Accordingly, approximately 924 patients will be randomized in the study to obtain at least 840 mITT patients in a 1:1:1 ratio (280:280:280 patients, respectively) to each treatment group.

3.3 Randomization and Unblinding Procedures

Patients will be randomly assigned in a 1:1:1 ratio to receive the Test product, the Reference product, or the Vehicle control, respectively. The randomization assignment will be a block randomization, with 3 sequential patient numbers in each block.

An independent third party will generate the randomization code using SAS and hold it throughout the conduct of the study in order to minimize bias. A sealed copy of the randomization scheme will be retained at each study site.

The treatment assignments will remain blinded until the final database is locked. The contents of the cans of IP may not be viewed by the PI, any Sub-investigator, evaluator or any other (blinded) member of the site staff.

In the event of an emergency, the patient-specific treatment may be identified; however, every effort should be made to maintain the blind.

In the event of unblinding, the patient should be excluded from the PP population. The Sponsor or designee must be notified in the event the blind is broken.

4. STUDY ENDPOINTS/OUTCOMES

Primary Efficacy Endpoint

The primary efficacy endpoints are the percent change from Baseline to Week 12 in the inflammatory (papules and pustules) lesion count.

Secondary Efficacy Endpoint

The secondary endpoint is the clinical response of "success" or "failure" at Week 12 on the IGE. Success is defined as an IGE score of 0 (clear) or 1 (almost clear). Any patient who is not considered to be a success will be considered to be a failure.

5. HYPOTHESES TESTING

Hypothesis of Equivalence (Primary Endpoint)

A two-sided, 90% confidence interval on the test/reference ratio for mean percent change from Baseline in the inflammatory lesion count will be constructed using Fieller's method. The estimates of treatment means and standard errors will be obtained from a two-way Analysis of Variance of the Test and Reference results, using a statistical model containing terms for treatment and center. Non-parametric methods will be used if the skewness factor for the residuals from the Analysis of Variance (ANOVA) model is outside the range -2 to +2.

Bioequivalence will be established if the 90% confidence interval for the ratio of test/reference means is contained within the interval [0.80, 1.25].

Hypothesis of Equivalence (Secondary Endpoint)

A two-sided, continuity-corrected, 90% confidence interval on the Test-to-Reference difference for the proportion of patients with treatment success on the IGE will be constructed.

Bioequivalence will be established if the 90% confidence interval for the difference is contained within the interval [-0.20, +0.20].

Hypothesis of Superiority (Primary Endpoint)

The null hypothesis to be tested is that there is no difference in the mean percent change (reduction) from baseline in the inflammatory lesion count between the active treatment and the Vehicle treatment. The hypothesis testing will be performed separately for the Test treatment versus the Vehicle treatment and for the Reference treatment versus the Vehicle treatment using ANOVA under assumption of normal error and homogeneity of variance. Only data for the relevant two treatment arms will be included in each ANOVA. Similarly, to bioequivalence, non-parametric methods will be used if the skewness factor for the residuals from the ANOVA model is outside the range -2 to +2.

Superiority will be established if the mean percent change (reduction) from Baseline in the inflammatory lesion count for each active treatment is greater than, and statistically different from ($p < 0.05$, two-sided), that for the Vehicle.

Hypothesis of Superiority (Secondary Endpoint)

The null hypothesis to be tested is that there is no difference in the proportions of patients with treatment success on the IGE. The evaluation of superiority will be conducted separately for the Test treatment versus the Vehicle treatment and for the Reference treatment versus the Vehicle treatment. The analysis will be conducted using two-sided, $\alpha = 0.05$, Fisher's exact test.

Superiority will be established if the success proportion for each active treatment is greater than, and statistically different from, that of the Vehicle.

6. ANALYSIS SUBSETS

6.1 Safety Population

The safety population includes any individual who was randomized into the study and has evidence of usage of at least one dose of IP according to subject's diary. For patients who are randomized but fail to return their diary card and for whom no evidence of study drug use is available, the following conservative approach will be applied:

- if some safety or efficacy data for these patients is available after Visit 1, they will be included in the Safety population;
- if no safety or efficacy data for these patients is available after Visit 1, they will be considered lost to follow-up and excluded from the Safety population.

The Safety population will be the primary population for the safety analysis.

6.2 Modified Intent to Treat Population (mITT Population)

The mITT population includes all Safety population patients who met all inclusion/exclusion criteria and return for at least one post-baseline efficacy evaluation. This population will be considered as supportive for testing the clinical equivalence and as definitive while testing the superiority.

6.3 Per Protocol Population (PP Population)

The PP population includes all mITT patients who apply 75% to 125% of the scheduled applications of the assigned product for 12 weeks, do not miss more than 6 consecutive applications, return for the 12-week evaluation within +/- 4 days OR have discontinued from the study due to lack of treatment effect (after completing at least 8 weeks of compliant study medication use), and have no protocol major violations.

Patients who are discontinued prematurely from the study due to lack of treatment effect after completing at least 8 weeks of compliant study medication use and without major protocol violations will be included in the PP population as treatment failures (i.e., non-responders) even if they do not have Week 12 evaluation within the visit window or at all. An LOCF approach will be used for imputing missing lesion counts in these PP patients and they will be considered as failures in the IGE evaluations.

Patients discontinued prematurely for other reasons (including those who discontinue due to lack of treatment effect with less than 8 weeks of treatment) will be excluded from the PP population, but included in the mITT population, using LOCF.

Patients can be additionally excluded from the PP population due to non-compliance, if the Investigator's and Sponsor's review of their dosing times suggests a clinically meaningful departure from a twice-daily dosing pattern.

For the purpose of determining the PP status of the patient, a "major protocol violation" is any patient or investigator activity that could have possibly interfered with the therapeutic administration of the treatment or the precise evaluation of the treatment efficacy.

The PP population will be considered as definitive for testing the clinical equivalence and as supportive while testing the superiority.

7. STATISTICAL METHODS OF ANALYSIS

7.1 General Principles

The statistical analyses will be performed by [redacted], under the direction of the Sponsor, Actavis Laboratories UT, using SAS Version 9.3(or higher). All tables, figures, and listings will be produced in the landscape format.

In general, all data will be listed by treatment group, patient and visit/time point where appropriate. The summary tables will also be stratified by, or have columns corresponding to, treatment group.

All patients will be identified by their unique patient numbers. Data from the screen failures will not be included in tables, listings, or figures.

The total number of patients in the study group (N) under the stated population will be displayed in the header of summary tables.

Data will be summarized using descriptive statistics for continuous variables. Unless otherwise stated, descriptive statistics will include number of patients, mean, standard deviation, minimum, median, and maximum. The minimum and maximum statistics will be presented to the same number of decimal places as the original data. The mean and median will be presented to one more decimal place than the original data. The standard deviation will be presented to two more decimal places than the original data. The number of missing observations will be presented only if non-zero.

In summary tables of categorical variables, counts, and percentages will be used. The count [n] indicates the actual number of patients in a particular category, which should always be less than or equal to the total number of patients in the respective study group with known (non-missing) category [N]. Percentage will be obtained by: % = n/N*100. Unless otherwise stated, all percentages will be expressed to one decimal place.

All statistical tests will be two-sided at a significance level of $\alpha = 0.05$, unless otherwise indicated. No adjustment will be made for multiplicity.

Relative days will be calculated relative to date of first dose of study medication. In general, relative days will be calculated as follows only when the full assessment date is known (i.e., partial dates will have missing relative days).

For assessment on or after the day of first dose of study drug:

Relative Day = Date of Assessment – Date of First Dose of study Drug+1.
For assessment before the day of first dose of study drug:
Relative Day = Date of Assessment – Date of First Dose of study Drug.
All dates will be displayed in DD/MMM/YYYY format.

7.2 Patient Disposition

The number of patients enrolled in the study, randomized to treatment, included in the Safety, PP, mITT populations, prematurely discontinued from the study (along with the reasons for discontinuation) will be calculated. The percentages will be based on the number of patients

randomized to each treatment group. Percentages for discontinuation reasons will be based on the sub-population of patients who discontinued from the study.

Number and percentage of patients enrolled by site will be tabulated for all enrolled patients, Safety, mITT, and PP populations.

7.3 Demographic and Baseline Characteristics

Demographic characteristics will include:

- age;
- gender;
- race;
- ethnicity;
- Rosacea history including time since diagnosis (years).

Baseline characteristics include:

- Baseline inflammatory lesion count;
- Baseline telangiectasia assessment;
- Vital signs: systolic and diastolic blood pressure, heart rate, respiratory rate, oral body temperature;
- Physical examination: height, weight;
- Medical history other than rosacea history.

Descriptive statistics will be presented for age (years), height, weight, body mass index (BMI), time since diagnosis (years), lesion count. Frequency counts and percentages will be presented for race, ethnicity, and baseline telangiectasia grade. Height will be reported in centimeters and weight in kilograms.

Age will be derived from Informed Consent Signed Date (INFCSD) and Date of Birth (DOB) as the number of whole years between those two dates.

Demographic and baseline/randomization characteristics will be evaluated for comparability across treatment groups in the following manner. Continuous variables (age, height, and weight) will be analyzed with an ANOVA with factors of treatment and investigational site. Overall p-value for the global null hypothesis of all groups being equal will be displayed. Categorical variables (gender, ethnicity, and race) will be analyzed with a Cochran-Mantel-Haenszel general association test, stratified by investigational site.

Separately, a categorical summary will be created for physical examination results and vital signs results (normal, abnormal not clinically significant, and abnormal clinically significant). Medical history other than rosacea history will be summarized by Medical Dictionary for the Drug Regulatory Activities (MedDRA) System Organ Class and Preferred Term. Baseline vital signs and physical examination will be listed.

These analyses will be performed for Safety, mITT, and PP populations.

All parameters reported during screening or baseline phase (including informed consent information, inclusion/exclusion criteria, randomization information, method of contraception, etc.) will be presented in the by-patient listings.

7.4 Protocol Deviations

Protocol deviations will be derived algorithmically. The following deviation categories will be derived:

1. Use of prohibited medication: use of any concomitant medication identified by the clinical review as critical and influencing the efficacy outcomes (major deviation).
2. Concomitant Medication – not critical: use of any other concomitant medication (minor deviation).
3. Diary accountability: subjects failing to return their diary at at least one scheduled visit (minor deviation).
4. Missed > 6 consecutive doses (major deviation).
5. IP compliance < 75% or > 125% (major deviation).
6. Dosing – Missed Doses(s): subjects who missed at least one dose, but do not fall under the previous two items (minor deviation).
7. IP Accountability: subjects who failed to return the IP at at least one scheduled visit (minor deviation).
8. Visit 4 out of window: subjects who attended Week 12 visit outside of ± 4-day window (major deviation).

9. Failed to return to Visit 4: subjects who do not attend Visit 4 at all. This deviation will not apply to subjects who discontinued from study due to lack of treatment effect after at least 8 weeks of treatment (major deviation).
10. Inclusion/Exclusion criteria violated: subjects who were enrolled despite violating inclusion/exclusion criteria (major deviation).
11. Pregnancy: subjects with positive pregnancy test results at any time during the study or adverse event with preferred term "Pregnancy" or discontinuing from the study with reason "Pregnancy" (major deviation).
12. Urine Pregnancy Test not done: subject is female and of child-bearing potential, but UPT was not performed at Visits 2, 3, or 4 (minor deviation).
13. Facial lesion count not done at Visits 2 or 3 (minor deviation);
14. Facial lesion count not done at Visit 4 (major deviation);
15. IGE not done at Visits 2, 3, or 4 (minor deviation);
16. Sign and symptoms of local irritation not done at Visits 2, 3, or 4 (minor deviation).

Additional deviations may be reported by sites classified into deviation categories and graded as major or minor.

Protocol deviations will be summarized by deviation type and treatment group. This analysis will be performed for Safety, mITT, and PP populations.

7.5 Efficacy Analyses

7.5.1 Center pooling

To eliminate potential effect of random fluctuations at small site on the primary endpoints small centers will be pooled. A study center will be pooled if it doesn't meet both of the following conditions:

- It has at least 10 patients in the PP population;
- It has at least one patient in each treatment group in the PP population.

The smallest center that does not meet the above requirements will be pooled with the next smallest center. The procedure will be repeated until all pooled centers meet the above two requirements.

Pooled center will be used in all the efficacy analyses.

7.5.2 Analyses of Primary Endpoint

At each visit, an Investigator will assess the patient's facial rosacea by counting the number of pustules, papules, and nodules. The number of pustules, papules, total number of inflammatory lesions (pustules and papules) and nodules will be reported on the CRF.

Missing Week 12 assessments will be imputed as follows. In the mITT analysis, if a patient discontinues the study prior to Week 12 or misses the Week 12 assessment for any other reason, the LOCF rule will be used to impute the number of lesions. In the PP analysis, if a patient discontinues the study prior to Week 12, due to lack of treatment effect, after at least 8 weeks of treatment, the LOCF rule will be used to impute the number of lesions. If a patient is missing the Week 12 assessment for any other reason, the patient will be excluded from the PP population.

The baseline lesion count will be defined as the results from the latest examination prior to the start of study drug therapy. For each post-baseline visit the change from Baseline and percent change from Baseline in the total number of inflammatory lesions will be calculated. The baseline value, timepoint value, change from baseline, and percent change from baseline at the timepoint will be analyzed using descriptive statistics and will be tabulated by site, visit, and treatment group.

7.5.2.1 Analysis of clinical equivalence of test and reference treatments

To show the clinical equivalence, estimates of mean percent change from Baseline in the inflammatory lesion count will be calculated for the Test and Reference treatment, and then the 90% CI for the mean ratio will be constructed using Fieller's method. Bioequivalence will be established if the 90% confidence interval for the ratio of Test/Reference means is contained within the interval [0.80, 1.25] for the PP population.

To this end, first an ANOVA model will be fit with percent change from Baseline in lesion count as outcome and treatment, center and treatment-by-center interaction, as factors on the data from Test and Reference treatments only (excluding Vehicle patients). If the treatment-by-center interaction factor is not significant at the 0.05 level, the model will be rerun without the interaction term. Treatment means and standard errors will be estimated from this model.

Then Fieller's formula will be applied; covariance between the treatment means will be assumed to be 0. See Appendix 11.4 for complete description of the Fieller's formula.

A non-parametric rank based ANOVA will be considered when the data is highly skewed. The evaluation of skewness, (using SAS® PROC UNIVARIATE) will be performed using the residuals from ANOVA and if the skewness statistic is less than -2 or greater than +2, the analysis will be performed on the ranks of the percent change in the inflammatory lesion count values.

Analysis of bioequivalence will be performed both on the PP population and on the mITT population. The results on the PP population will be considered definitive and the results on the mITT population will be considered supportive.

7.5.2.2 Analysis of superiority to vehicle control

The analysis of superiority will be performed separately for the Test treatment versus the Vehicle treatment and for the Reference treatment versus the Vehicle treatment. Each of these analyses will be performed using an ANOVA model with percent change from baseline in the lesion count as outcome and treatment, center and treatment-by-center interaction as factors. If the treatment-by-center interaction factor is not significant at the 0.05 level, the model will be rerun without the interaction term. The model will be fit on data from the Test and Vehicle treatment for analysis of superiority of the Test treatment and separately on data from the Reference and Vehicle treatment for analysis of superiority of the Reference treatment. From this model, the least square (LS) mean estimate for each treatment group with the 95% CI will be calculated; further, an estimate of the LS mean difference between the active treatment (Test or Reference) and Vehicle with 95% CI and the p-value for test of no difference will be calculated.

Superiority will be established if the mean percent change from Baseline (reduction) for each active treatment is estimated to be greater than, and statistically significantly different from ($p < 0.05$ for test of no difference) that for the Vehicle, for the mITT population.

Analysis of superiority will be performed both on the PP population and on the mITT population. The results on the mITT population will be considered definitive and the results on the PP population supportive.

A non-parametric rank based ANOVA will be considered under the same circumstances as for the analysis of bioequivalence.

7.5.3 Analyses of Secondary Endpoint Outcome

The secondary endpoint is the IGE score (see Appendix 11.2), expressed in terms of treatment success or failure. Success is defined as an IGE score at Week 12 of 0 (clear) or 1 (almost clear). Any other outcome will be considered a failure. Patients who are discontinued prematurely from the study due to lack of treatment effect after at least 8 weeks of compliant treatment will be considered as treatment failures both in the PP and mITT analyses. Patients who do not have a valid Week 12 assessment for any other reason will be excluded from the PP analysis, but included in the mITT analysis using LOCF approach to impute the response.

7.5.3.1 Analysis of clinical equivalence of test and reference treatments

A two-sided, continuity-corrected, 90% confidence interval on the Test-to-Reference difference for the proportion of patients with treatment success on the IGE will be constructed. Bioequivalence will be established if the 90% confidence interval for the difference is contained within the interval [-0.20, +0.20] for the PP population.

Analysis of bioequivalence will be performed both on the PP population and on the mITT population. The results on the PP population will be considered definitive and the results on the mITT population supportive.

7.5.3.2 Analysis of superiority to vehicle control

The evaluation of superiority will be conducted separately for the Test treatment versus the Vehicle treatment and for the Reference treatment versus the Vehicle treatment, comparing the proportions of patients with treatment success on the IGE. The analysis will be conducted using two-sided Fisher's exact test.

Superiority will be established if the success proportion for each active treatment is greater than, and statistically significantly different from ($p < 0.05$ for the Fisher's test) that of the

Vehicle for the mITT population. Analysis of superiority will be performed both on the PP population and on the mITT population. The results on the mITT population will be considered definitive and the results on the PP population supportive.

7.5.4 Analyses of IGE

Investigator Global Evaluation (IGE) expressed as severity grade 0 to 4 will be performed at each visit. For study enrollment, the patient must have an IGE rosacea severity score of 3 at Baseline visit.

Number and percentage of patient at each IGE severity grade will be tabulated by visit. In addition, a shift table for changes in IGE grades from Baseline to each post-baseline visit will be created. This analysis will be performed for both mITT and PP populations.

7.6 Safety Analyses

7.6.1 Adverse Events

Adverse Events will be coded using the MedDRA, Version 18.0, AE coding system for purposes of summarization.

Only Treatment Emergent Events (TEAEs) will be used for the summary analysis. An AE will be considered as treatment-emergent if the time of onset is after the time of the first study drug administration or if it increased in severity during the study period. AEs with unknown start dates will be counted as treatment-emergent unless the AE resolution date is prior to the study drug start date. If the start date is partially missing, then month and year (when available) will be used to determine if the event occurred prior to or post dosing.

A summary of the frequencies (number and percentage of patients) of TEAEs, serious TEAEs (AEs with missing seriousness will be treated as serious), treatment-related TEAEs will be presented by system organ class and preferred term. Adverse events will also be analyzed by their severity (Mild, Moderate, Severe). In case the severity was not assessed, the most conservative result – severe will be chosen for the analysis.

A TEAE is termed as treatment-related if it is recorded as definitely, probably, or possibly related to the study medication on the CRF. In case the relatedness was not assessed, the most conservative result – related will be chosen for the analysis.

A patient experiencing the same AE multiple times will only be counted once for that preferred term. Similarly, if a patient experiences multiple AEs within the same system organ class, that patient will be counted only once in that system organ class. All AEs will be listed in alphabetical order of SOC and preferred terms.

An overall summary will include, by treatment group and overall, the number of TEAEs and the number and percentage of patients reporting at least 1 TEAE in the following categories:

- Any TEAE,
- Treatment-related TEAE,
- Serious TEAE,
- TEAE leading to discontinuation of the study medication.

The following TEAE frequency tables will be prepared summarizing the overall number of TEAEs, the number and percentage of patients reporting at least one TEAE by MedDRA SOC and PT:

- All TEAEs,
- TEAEs by Severity,
- TEAEs by Relationship to Study Medication.

Additionally, TEAEs will be summarized by the preferred terms in the descending order of frequency in the total treatment group. In this table a p-value from Fisher's exact test comparing event rates between the Test and the Reference treatment groups will be provided for those preferred terms that have frequency > 1% in either Test or Reference group.

All information pertaining to adverse events noted during the study will be listed by patient, detailing verbatim, preferred term, system organ class, start date, stop date, intensity, outcome, action taken, and causal relationship to the study drug. The adverse event onset will also be shown relative (in number of days) to the date of first administration of the study medication. In addition, the adverse event duration (if AE Stop Date is available) will be evaluated as below and presented (in number of days).

$$\text{AE Duration} = \text{AE Stop Date} - \text{AE Start Date} + 1$$

7.6.2 Signs/Symptoms of Local Irritation

At each study visit, beginning with Visit 1, patients will be evaluated for any signs and symptoms of local irritation (application site reactions), including erythema, dryness, scaling, pruritus, burning/stinging and edema. Baseline values will be used for comparative purposes against the scores documented at subsequent visits for each treatment group. Each patient will be assigned a severity score by an Investigator based on the scale as presented in Appendix 11.3.

Number and percentage of subjects with each severity will be presented by visit and by symptom. Additionally, the shifts from baseline in irritation score will be tabulated by site, treatment, symptoms and scheduled post-baseline visits. The denominator for the proportions will be the number of patients with evaluated signs and symptoms both at baseline and at given visit.

7.6.3 Exposure to Product

The patients will be instructed to use the diary to document all doses taken by checking the yes or no box for the appropriate date and am/pm time. The date(s) and reason for each dosing noncompliance must be recorded.

Compliance with scheduled application of IP will be determined from the patient's diary as [Number of recorded applications] / [Planned number of applications] * 100%, where Planned number of applications is 168.

Number of missed doses will be calculated as follows. The number of doses missed in the dosing period (between the first and the last study drug application) will be taken directly from the CRF. Additionally, number of doses missed outside of the dosing period is the number of missed dosing times between the date and time (AM/PM) of the last dose and the date and time when the subject would have reached the required 168 doses (i.e. the date and time (AM/PM) of the first dose plus 84 days, inclusive)..

Patients who have more than 6 consecutive missed doses will be considered non-compliant and will be excluded from the PP population. For the dosing period >6 consecutive applications will be collected directly on the eCRF. Additionally, patients will be considered having >6 consecutive missed doses if after the dosing period patients missed more than 6

dosing times after their last dose and prior to the morning of Visit 4 date or the date and time (AM/PM) when they would have reached the required 168 doses, whichever occurs earlier. Patients can be additionally excluded from the PP population due to non-compliance, if the Investigator's and Sponsor's review of their dosing times suggests a clinically meaningful departure from a twice-daily dosing pattern.

Patients will be considered compliant if they apply at least 75% and not more than 125% of doses, with no more than 6 consecutive missed doses. The compliance will be analyzed using the descriptive statistics by treatment group. The proportion of compliant vs. non-compliant patients will be tabulated for each treatment.

Duration of exposure will be calculated as Date of last use of study medication – Date of first use of study medication + 1. Duration of exposure will be summarized descriptively by treatment group.

Compliance and duration of exposure will be summarized for the Safety and mITT populations.

7.6.4 Exposure to Concomitant Medication

Prior and concomitant medications and concomitant non-drug therapies, including the use of sunscreen, in addition to the reason for the medication use, will be assessed at baseline and at each subsequent study visit. The start and stop date of concomitant medication use during the study will be provided in the data set, in addition to the reason for the medication use.

Medication or non-drug therapy will be classified as prior, if the end date is known and is prior to the first use of the study medication. Medications and non-drug therapies that are ongoing or ended after the first use of the study medication will be classified as concomitant. If the end date of the medication or non-drug therapy is unknown, it will also be considered concomitant.

The concomitant medications will be summarized by preferred term for each treatment group as coded by the World Health Organization Drug (WHODRUG) dictionary Sept-2015 B2 and be listed in a by patient listings. Prior medications will be listed only.

8. CHANGES FROM PROTOCOL-SPECIFIED ANALYSES

There are no changes from the protocol-specified analyses.

9. LIST OF PLANNED TABLES, FIGURES, AND LISTINGS

See separate document with the table, figure and listing shells.

10. LITERATURE CITATIONS / REFERENCES

1. Study Protocol: A Multicenter, Double-blind, Randomized, Parallel-group, Vehicle-Controlled Study to Evaluate the Safety and Clinical Equivalence of a Generic Azelaic Acid Foam, 15% and the Reference Listed Finacea® (azelaic acid) Foam, 15% in Patients with Moderate Facial Rosacea.

11. APPENDICES

11.1 Study visit Schedule

Visit Number	Visit 1	Visit 2	Visit 3	Visit 4
Visit Day	**Day 0**	**Day 28±4**	**Day 56±4**	**Day 84±4**
Visit Name	**Baseline**	**Interim**	**Interim**	**End of Treatment/ Early Termination**
Informed Consent	X			
Demographics	X			
Medical History	X			
Inclusion/Exclusion Criteria	X			
Physical Exam including Vital Signs	X			
Pregnancy Test *	X	X	X	X
Concomitant Medication	X	X	X	X
Adverse Event		X	X	X
Inflammatory Lesion Counts	X	X	X	X
Investigator's Global Evaluation (IGE)	X	X	X	X
Telangiectasia Assessment	X			
Irritation Assessment	X	X	X	X
Subject Instruction/Compliance Review	X	X	X	X
Dispense Study Medication and Diary **	X	X	X	
Collect Study Medication and Diary (Accountability)		X	X	X

* Urine pregnancy test for all females of childbearing potential enrolled in study.
** Day 1 will occur the date patient begins dosing. This may be the evening of Visit 1 or the following day.

11.2 IGE Scale

Score	Grade	Description
0	Clear	No inflammatory lesions present; at most, mild erythema
1	Almost Clear	Very mild erythema present. Very few small papules/pustules.
2	Mild	Mild erythema. Several small papules/pustules.

3	Moderate	Moderate erythema. Several small or large papules/pustules, and up to 2 nodules.
4	Severe	Severe erythema. Numerous small and/or large papules/pustules, up to several nodules.

11.3 Application Site Reactions Scores

Erythema:

Score	Grade	Definition
0	None	No redness present.
1	Very Mild	Slight pinkness.
2	Mild	Pink to light red.
3	Moderate	Definite redness, easily recognized.
4	Severe	Marked erythema; fiery red.

Dryness:

Score	Grade	Definition
0	None	No dryness.
1	Mild	Slight but definite roughness.
2	Moderate	Moderate roughness.
3	Severe	Marked roughness.

Scaling:

Score	Grade	Definition
0	None	No scaling.
1	Mild	Barely perceptible shedding, noticeable only on light scratching or rubbing.
2	Moderate	Obvious but not profuse scaling.
3	Severe	Heavy scale production.

Pruritus:

Score	Grade	Definition
0	None	No itching.
1	Mild	Slight itching but not bothersome.
2	Moderate	Definite itching, somewhat bothersome without loss of sleep.
3	Severe	Intense itching that has caused pronounced discomfort; interrupted sleep and excoriation of the skin may be present.

Stinging/Burning:

Score	Grade	Definition
0	None	No stinging/burning.
1	Mild	Slight warm tingling sensation but not bothersome.
2	Moderate	Definite warm tingling/stinging sensation somewhat bothersome.
3	Severe	Hot tingling/stinging sensation that has caused definite discomfort

Edema:

Score	Grade	Definition
0	None	No swelling.
1	Mild	Slightly or barely perceptible swelling.
2	Moderate	Distinct presence of swelling.
3	Severe	Marked or intense swelling.

11.4 Code Fragments

ANOVA model for superiority analysis in primary endpoint

```
proc glm data=<datasets name>;
   class <treatment> <center>;
   model <Percent change from baseline in lesion count at week 12> = <treatment> <center>
   <treatment>*< center>/ ss3;
   lsmeans <treatment> / pdiff cl;
```

```
    output out=residuals residual=residual;
run;
quit;
```

Note: this analysis needs to be performed separately for test and reference treatments on a dataset containing only test and vehicle or reference and vehicle treatment patients. If treatment-by-center interaction term is not significant at 0.05 level, the model will be rerun without this term.

Superiority analysis in secondary endpoint

```
proc freq data=<dataset>
    tables <treatment>*<success> / fisher;
run;
```

Note: this analysis needs to be performed separately for test and reference treatments on a dataset containing only test and vehicle or reference and vehicle treatment patients.

Analysis of clinical equivalence in primary endpoint

ANOVA model:

```
proc glm data=<datasets name>;
    class <treatment> <center>;
    model <Percent change from baseline in each type of lesions at week 12> = <treatment>
    <center> <treatment>*<center>/ ss3;
    lsmeans <treatment> / stderr;
    output out=residuals residual=residual;
run;
quit;
```

Here dataset contains test and reference treatment patients only.
Note: if the treatment-by-center interaction term is not significant at the 0.05 level, the model will be rerun without this term.

Fieller's method.

Generally, Fieller's formula allows to calculate the confidence interval for the ratio of two (possibly correlated) means of two samples a and b with expectations μ_a and μ_b, and variances $\nu_{11}\sigma^2$ and $\nu_{22}\sigma^2$ and covariance $\nu_{12}\sigma^2$. If $\nu_{11}, \nu_{12}, \nu_{22}$ are all known, then a (1 − 2α) confidence interval (mL, mU) for μ_a/μ_b is given by

$$(m_L, m_U) = \frac{1}{(1-g)}\left[\frac{a}{b} - \frac{g\nu_{12}}{\nu_{22}} \mp \frac{t_{r,\alpha}s}{b}\sqrt{\nu_{11} - 2\frac{a}{b}\nu_{12} + \frac{a^2}{b^2}\nu_{22} - g\left(\nu_{11} - \frac{\nu_{12}^2}{\nu_{22}}\right)}\right]$$

where

$$g = \frac{t_{r,\alpha}^2 s^2 \nu_{22}}{b^2}.$$

Here s^2 is an unbiased estimator of σ^2 based on r degrees of freedom, and $t_{r,\alpha}$ is the α-level deviate from the Student's t-distribution based on r degrees of freedom.

In the case of this study the two samples are independent (different patients), thus covariance can be assumed to be zero, and the formula simplifies to:

$$(m_L, m_U) = \frac{1}{1-g}\left[\frac{m_t}{m_r} \mp \frac{t_{r,\alpha}}{m_r}\sqrt{se_t^2 + \frac{m_t^2}{m_r^2}se_r^2 - g \cdot se_t^2}\right]$$

where

$$g = \frac{t_{r,\alpha}^2 se_r^2}{m_r^2}$$

Here m_t and se_t are mean and standard error estimate for test treatment and m_r and se_r for reference treatment correspondingly obtained from the above model. The degrees of freedom r can be obtained from the model as degrees of freedom for the error term in the overall ANOVA table, $\alpha = 0.05$ (for the 90% confidence interval). The term $t_{r,\alpha}$ can be calculated in SAS as tinv(α, r).

Clinical equivalence analysis in secondary endpoint

```
proc freq data=<dataset>
    tables <treatment>*<success> / riskdiffc;
run;
```

Note: this analysis needs to be performed on a dataset containing test and reference treatment patients only.

Evaluation of skewness

This evaluation will be performed on the residuals output by the "output" statement in PROCGLM above:

```
proc univariate data=residuals;
    var residual;
run;
```

Non-parameteric rank ANOVA

If decision to use the rank ANOVA is taken, the data will be ranked first:

```
proc rank data=<dataset>
    var <change_from_baseline>;
    ranks rank;
run;
```

After that the analysis will proceed using the same SAS code as described above but using rank variable instead of the original change from baseline.