A Multi-Center, Double-Blind, Randomized, Vehicle-Controlled, Parallel-Group Study to Compare Perrigo UK FINCO's Azelaic Acid Foam, 15%, to Bayer Healthcare Pharmaceutical Inc.'s, FINACEA® (AZELAIC ACID) FOAM, 15%, and Both Active Treatments to a Vehicle Control in the Treatment of Inflammatory Lesions of Rosacea

Protocol No.: PRG-NY-16-009

NCT02800148

10-04-2016

STATISTICAL ANALYSIS PLAN

PRG-NY-16-009: Azelaic Acid Foam, 15%

[redacted]

A Multi-Center, Double-Blind, Randomized, Vehicle-Controlled, Parallel-Group Study to Compare Perrigo UK FINCO's Azelaic Acid Foam, 15%, to Bayer Healthcare Pharmaceutical Inc.'s, FINACEA® (AZELAIC ACID) FOAM, 15%, and Both Active Treatments to a Vehicle Control in the Treatment of Inflammatory Lesions of Rosacea

Perrigo New York, Inc.
1701 Bathgate Ave.
Bronx, NY 10457

Written By: [redacted]

Approved By:

[redacted]

[redacted]

[redacted]

[redacted]

Table of Contents

List of Abbreviations

AE	Adverse Event
ANOVA	Analysis of Variance
CI	Confidence Interval
CMH	Cochran–Mantel–Haenszel Test
IGA	Investigator's Global Assessment
ITT	Intent-to-Treat (Population)
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-to-Treat (Population)
PD	Protocol Deviation
PP	Per-Protocol (Population)
PV	Protocol Violation
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
TEAE	Treatment-Emergent Adverse Event
TESAE	Treatment-Emergent Serious Adverse Event
WHO Drug	World Health Organization Drug Dictionary

Statistical Analysis Plan

1 Purpose of Statistical Analysis Plan

The purpose of the statistical analysis plan is to describe in detail all the data, statistical methods, and summary tables required to implement the statistical analysis of Clinical Study Protocol PRG-NY-16-009 ([redacted]).

2 Study Objectives

To compare the safety and efficacy profiles of Perrigo UK FINCO's Azelaic Acid Foam, 15%, to Bayer HealthCare Pharmaceuticals Inc., Finacea® (Azelaic Acid) Foam, 15%, in order to prove bioequivalence between them and to demonstrate the superior efficacy of the two active formulations over that of the vehicle in the treatment of Inflammatory Lesions of Rosacea.

3 Study Design and Sample Size

3.1 Study Design

For the purpose of exploring the above objectives, the study will be conducted as a double-blind, randomized, parallel-group, vehicle-controlled, multicenter trial.

Each subject will be randomly assigned to one of the following treatment groups in a [redacted]

(1) Test: Azelaic Acid Foam, 15%, [redacted]

(2) Reference: Finacea® (Azelaic Acid Foam, 15%) manufactured by Bayer HealthCare Pharmaceuticals Inc.

(3) Vehicle of test product, [redacted].

Subjects will be admitted into the study only after written informed consent has been obtained and all of the inclusion and none of the exclusion criteria have been met. Randomization will be performed according to a computer generated randomization scheme where the treatment group designation has been assigned to the subject number. The treatment designation will remain blinded until after the final database is locked. An independent third party will generate and hold the randomization code throughout the study. Randomized subjects will apply the smallest amount of the assigned study medication necessary as a thin layer to adequately cover each area of the face (chin, left cheek, right cheek, nose, and forehead), avoiding contact with the eyes (upper and lower eyelids), lips, broken skin, inside the nose/nostrils and mucous membranes, twice daily morning and evening ([redacted]) approximately the same time for 12 weeks..

Subjects will come to the study site for clinical evaluations at Visit 1/Day 1 (Baseline), Visit 2/Interim/Week 4 (Day 28 ±4 days), Visit 3/Interim/Week 8 (Day 56 ±4 days), and Visit 4/End of Treatment/ Week 12 Study (Day 84 ±4 days) or at early discontinuation. Safety will be assessed by monitoring adverse events at each visit and at the Week 2/Day 14 (±4 days) Telephone Contact.

3.2 Sample Size

Approximately [redacted] subjects will be randomized in ratio [redacted] (Test:Reference:Vehicle) to obtain [redacted] mITT subjects and [redacted] PP subjects. This sample size [redacted]

[redacted]

The actual number of subjects enrolled in the study will be based on blinded review of subject status (related to the per-protocol definition, section 4) to determine that the number of subjects expected to meet the PP criteria is sufficient ([redacted]). If this number is expected to be met prior to enrolling [redacted] subjects, the enrollment will be closed.

4 Populations To Be Analyzed

The analysis populations are defined as follows:

(1) Intent-to-treat (ITT) population: any subject that was randomized, received and used study medication;

(2) Modified Intent-to-treat (mITT) population: any subject, who met the inclusion/exclusion criteria, was randomized, received and used the study medication, and returned for at least one post-baseline efficacy assessment;

(3) Per Protocol (PP) population; any subject who (a) met all inclusion/exclusion criteria; (b) was randomized and received and used study medication; (c) met the protocol criteria for compliance [redacted]); (d) completed Visit 4/Week 12/Day 84 (End of Treatment/Early Termination Visit) within window OR was discontinued from the study due to treatment failure and; (e) Without significant protocol violations that could have interfered with the administration of the treatment or the precise evaluation of treatment efficacy.

[redacted]

- [redacted]
- [redacted]

[redacted]

5 Planned Analyses

5.1 Methodological Considerations

The study will be conducted under the same protocol across all the sites. No formal statistical analyses are planned to evaluate the consistency of efficacy results across the multiple clinical sites. These results, however, will be tabulated and if a site's efficacy data are obviously inconsistent with the results across all sites, this will be explored and addressed in the final study report.

All randomized subjects who received study medication will be evaluated for safety using ITT population. The efficacy analysis will be conducted on both the PP and the mITT populations.

Two-sided hypothesis testing will be conducted for all the tests. Resulting p-values less than 0.05 will be considered statistically significant unless noted otherwise. No adjustments of p-values for multiple comparisons will be made. No interim analyses are planned.

SAS software will be used for all data analyses and tabulations.

5.2 [redacted]

[redacted]

5.3 Demographics and Baseline Characteristics

Baseline variables (e.g., sex, age, ethnic origin) will be analyzed by adjusting for site, to identify differences between treatment groups. Any significant baseline differences will be reviewed for their potential impact on the efficacy findings.

Continuous variables at baseline will be examined by two-way analysis of variance (ANOVA) with treatment and site as fixed effects when ANOVA assumptions (normally distributed error, homogeneous variance) are satisfied, or by the nonparametric rank test when ANOVA assumptions are not met.

Categorical variables such as gender, race, etc., will be examined by Cochran-Mantel-Haenszel test, stratified by site. Summary tables by treatment group will be presented. For each continuous

variable, the summary will include the mean, standard deviation, minimum and maximum. For each categorical variable, the summary will include frequencies and percentages.

5.4 Subject Accountability

A summary of subject disposition will be provided for all subjects. Descriptive summaries of subject disposition, reason for discontinuation, and analyses population will be provided by treatment group. The data will also be presented in subject data listings.

5.5 Efficacy Variables and Analyses

5.5.1 Primary Endpoint

The primary efficacy endpoint of the study is the mean percent change from baseline to Visit 4/Week 12 in the inflammatory (papules and pustules) lesion counts.

Equivalent Efficacy

For the mean percent change from baseline in the inflammatory lesion counts, the Test treatment will be considered to be therapeutically equivalent to the Reference treatment if the 90% confidence interval on the Test-to-Reference ratio of means, calculated by Fieller's Method, falls within the interval 0.80 to 1.25. The compound hypothesis to be tested for therapeutic equivalence between test and reference is:

H_0: $\mu_T/\mu_R \leq 0.80$ or $\mu_T/\mu_R \geq 1.25$ versus

H_A: $0.80 < \mu_T/\mu_R < 1.25$.

Where μ_T and μ_R are the mean percent change from baseline to Visit 4/Week 12 (Day 84) in inflammatory lesions counts for the test treatment and the reference treatments, respectively. The null hypothesis is rejected when the two-sided 90% confidence interval (CI) for the ratio of means between test and reference products is between 0.80 and 1.25. Rejection of the null hypothesis supports the conclusion of therapeutic equivalence between test and reference products for the primary efficacy variable.

The two-sided 90% confidence interval will be constructed using an ANOVA model with treatment and site as factors. [redacted]

The therapeutic equivalence evaluation in the per-protocol (PP) population will be considered definitive and that in the mITT will be considered supportive.

Analysis for therapeutic equivalence will be performed based on the following SAS code (SAS Institute v.9.1.3):

The key procedure used to determine CIs for ratio of treatment means:

```
PROC GLM DATA=inf_pp;
 CLASS trt site;
 MODEL pchg_inf = trt site;
 ESTIMATE 'ratio=c' intercept (1-c) trt 1 -c;
RUN;
```

Where the ESTIMATE statement is run targeting at two-sided p=0.10 to locate the c value (our Test/RLD) for the 90% CI.

Superiority

For the percent change from baseline in the inflammatory lesion counts, each active treatment will be evaluated to determine if it has superior efficacy to that of the Vehicle at Visit 4/Week 12 (Day 84) via an ANOVA model containing terms for treatment and site. The compound hypothesis to be tested for superiority of test and reference over Vehicle is:

$H_0: \mu_T \leq \mu_V$ or $\mu_R \leq \mu_V$ versus

$H_A: \mu_T > \mu_V$ and $\mu_R > \mu_V$

Where μ_T, μ_R and μ_V are the mean percent change from baseline to Visit 4/Week 12 (Day 84) in inflammatory lesions counts for the test, the reference and the vehicle treatments, respectively. The null hypothesis is rejected when both p-values from the ANOVA are less than 0.05 (two-sided test). Rejection of the null hypothesis supports the conclusion of superiority of test and reference products over the vehicle product for the primary efficacy variable.

A skewness test (SAS® PROC UNIVARIATE) will be performed using the residuals from the ANOVA of the primary efficacy variable. If the skewness statistic is greater than 2 or less than -2, the analysis will be performed on the ranked mean percent change from baseline to Visit 4/Week 12 (Day 84) in inflammatory lesion counts.

Superiority analyses in the mITT population will be considered definitive and those in the PP will be considered supportive.

Analysis for superiority will be performed based on the following SAS code (SAS Institute v.9.1.3):

The key procedure used to compare mean values of an active treatment arm to vehicle:

```
PROC GLM DATA=inf_pp;
  CLASS trt site;
  MODEL pchg_inf = trt site /ss3;
RUN;
```

5.5.2 Secondary Endpoints

The secondary efficacy endpoint is the proportion of subjects with clinical success defined as a score of clear or almost clear (score of 0 or 1) on the Investigator Global Assessment (IGA) at Visit 4/Week 12 (Day 84).

Equivalent Efficacy

The compound hypothesis to be tested for clinical equivalence between test and reference is:

H_0: $p_T - p_R < -0.20$ or $p_T - p_R > 0.20$ versus

H_A: $-0.20 \leq p_T - p_R \leq 0.20$.

Where p_T and p_R are the proportions of subjects with clinical success at Visit 4/Week 12 (Day 84) for the test and reference products, respectively. The test product will be considered to be therapeutically equivalent to the reference product if the 90% CI on the difference in their proportions of subjects with clinical success, calculated by Wald's method with Yates' continuity correction, is contained within the limits -0.20 to +0.20. Rejection of the null hypothesis supports the conclusion of therapeutic equivalence between the test and reference products for the secondary efficacy variable.

The result of the analysis in the PP population will be considered definitive and that in the mITT population as supportive.

Analysis for therapeutic equivalence will be performed based on the following SAS code (SAS Institute v.9.1.3):

P-value is chosen from Continuity Adj. Chi-Square test.

The SAS code for 90% confidence interval (trt=1 for Test, trt=2 for Reference):

```
proc freq data=XX ;
  where trt in (1,2);
  tables trt* success / alpha=0.10 riskdiff;
  output out=CIDIFF (keep= l_rdif2 u_rdif2 ) riskdiff;
run;
```

For the final 90% continuity-corrected CI, the lower limit = (l_rdif2 - yates) and the upper limit = (u_rdif2 + yates), where yates, the Yates' continuity-correction factor, is derived as (1/n1+1/n2)/2, n1=number of subjects in Test arm and n2=number of subjects in Reference arm.

Superiority

The hypotheses to be tested for superiority of the test and reference products over Vehicle are:

H_0: $p_T \leq p_V$ versus H_A: $p_T > p_V$

H_0: $p_R \leq p_V$ versus H_A: $p_R > p_V$

Where p_T, p_R and p_V are the proportions of subjects with clinical success at Visit 4/Week 12 (Day 84) for the test, reference and Vehicle products, respectively. The tests will be conducted independently for the test product vs. vehicle and the reference product vs. vehicle using two-sided, $\alpha = 0.05$, continuity-corrected Z-tests. Superiority will be established if the

proportion of subjects with clinical success in the active treatment group is greater than and statistically different ($p< 0.05$) from that in the Vehicle. Rejection of the null hypothesis supports the conclusion of superiority of the test and reference products over the Vehicle product for the secondary efficacy variable.

The analyses in the mITT population will be considered primary and those in the PP population as supportive.

Analysis for superiority will be performed based on the following SAS code (SAS Institute v.9.1.3):

The SAS code for p-value (trt=2 for Reference vs. trt=3 for Vehicle):

```
proc freq data=XX ;
 where trt in (2,3);
 table trt * success /chisq ;
run;
```

5.5.3 [redacted]

[redacted]

5.6 Safety Variables and Analyses

Duration of Treatment and Medication Compliance

Number of applications, days of exposure, and compliance rate will be summarized by treatment group using descriptive statistics. For each subject, the overall duration of treatment (days) will be calculated using the following formula:

(Date of last application of study medication) - (Date of first application of study medication) + 1.

Medication compliance rate (%) will be calculated for each subject as follows:

(Total number of applications used) / (Expected number of applications) *100%.

[redacted]

Adverse Events

Adverse events (AEs) will be coded in MedDRA, version 18.1. Treatment-Emergent Adverse Event (TEAE) is defined as any AE occurs on or after application of the first dose of study drug. Number and percent of subjects reporting TEAEs will be tabulated by treatment group. Summaries will be presented by body system and preferred term for the ITT population, and

further by severity and relationship to study medication. TEAEs reported by 5% or more subjects for any treatment group will also be summarized. In the summaries of incidence rates (frequencies and percentages), severity and relationship to study drug, subjects who report more than one event that are mapped to the same preferred term will be counted only once under the strongest severity and relationship, accordingly. The difference between Test and Reference treatments with regard to the severity and frequency of their dermatological adverse events will be statistically evaluated. Chi-Square or Fisher's exact test will be used to compare the proportions of subjects of the two active treatment groups who report any TEAE.

Treatment-Emergent Serious Adverse Events (TESAEs) will be discussed within the clinical study report. TEAEs, TESAEs and TEAEs that led to treatment interruption or discontinuation will be presented in data listings.

Concomitant Medications, Laboratory Values, and Vitals Signs
Concomitant medications will be coded using the WHO Drug Dictionary, version September 2015, and will be presented in data listings. All vital signs data will be displayed in listings.

Erythema Severity and Application Site Reaction Assessments
Frequency and distribution of erythema severity assessment and application site reactions of dryness, scaling/peeling, pruritus, burning/stinging, pain and will be summarized and compared descriptively by visit.

Safety comparisons will be performed only for the ITT population.

6 Appendices

6.1 Handling of Missing or Incomplete Dates for Adverse Events and Concomitant Medications

Adverse Events
Handling of partial dates is only considered for the start date. An adverse event with a partial start date is considered treatment emergent if:

- only the day is missing and the start month/year is the same or after the month/year of the first dose
- the day and month are missing and the start year is the same or greater than the year of the first dose date
- the start date is completely missing

Concomitant Medications
Handling of partial dates is only considered for the stop date. A medication with a partial stop date is considered concomitant if:

- only the day is missing and the stop month/year is the same or after the month/year of the first dose
- the day and month are missing and the stop year is the same or greater than the year of the first dose date

- the stop date is completely missing or the medication is ongoing

6.2 Summary of Assessments

The schedule of visits and procedures to be conducted at each visit are summarized in the Schedule of Study Procedures.

Study Medication						