

A Multi-Center, Randomized, Double-Blind, Parallel-Group Vehicle Controlled Study To Compare The Efficacy And Safety Of CD5789 (Trifarotene) 50µg/g Cream Versus Vehicle Cream In Subjects With Acne Vulgaris.

NCT Number: NCT02556788

Date: 05 Oct 2015

**A MULTI-CENTER, RANDOMIZED, DOUBLE-BLIND, PARALLEL-GROUP, VEHICLE CONTROLLED
STUDY TO COMPARE EFFICACY AND SAFETY OF CD5789 50µg/g CREAM VERSUS VEHICLE
CREAM IN SUBJECTS WITH ACNE VULGARIS**

STATISTICAL ANALYSIS PLAN

RD.06.SPR.18252

Version Final

05Oct2015

This document is electronically approved and signed by: *Biometrics Manager*

Author :



Reviewer:



TABLE OF CONTENTS

| | | |
|--------------|--|-----------|
| 1 | OBJECTIVES | 5 |
| 2 | STUDY DESIGN | 5 |
| 3 | EFFICACY AND SAFETY ASSESSMENT | 6 |
| 3.1 | Efficacy Assessment | 6 |
| 3.1.1 | IGA (Investigator’s global assessment) of facial acne | 6 |
| 3.1.2 | PGA (Physician Global Assessment) of truncal acne..... | 6 |
| 3.1.3 | Lesions counts on the face and trunk..... | 7 |
| 3.1.4 | Subject self-assessment of facial acne improvement..... | 7 |
| 3.2 | Safety Assessment | 7 |
| 4 | EFFICACY AND SAFETY ENDPOINTS | 8 |
| 4.1 | Efficacy Endpoints | 8 |
| 4.1.1 | Primary Efficacy Endpoint | 8 |
| 4.1.2 | Secondary Efficacy Endpoint..... | 9 |
| 4.1.3 | Supportive Efficacy Endpoints | 9 |
| 4.2 | Safety Endpoints | 9 |
| 4.3 | Other Endpoints..... | 9 |
| 5 | POPULATIONS ANALYZED..... | 9 |
| 5.1 | Intention to Treat (ITT) Population..... | 10 |
| 5.2 | Intention to Treat on the Trunk (ITTT) Population..... | 10 |
| 5.3 | Safety (SAF) Population | 10 |
| 5.4 | Safety (SAFT) Population for Local Tolerability on the Trunk..... | 10 |
| 5.5 | Per Protocol (PP) Population | 10 |
| 5.6 | Per Protocol (PPT) Population on the Trunk..... | 11 |
| 6 | SAMPLE SIZE CONSIDERATION | 11 |
| 7 | STATISTICAL METHODS AND DATA CONSIDERATIONS | 12 |
| 7.1 | Study subjects..... | 12 |
| 7.1.1 | Disposition of subjects | 12 |

| | | |
|-------|---|----|
| 7.1.2 | Demographics and baseline characteristics..... | 12 |
| 7.2 | Protocol deviations..... | 13 |
| 7.3 | Medical history, previous and concomitant therapies/procedures..... | 13 |
| 7.4 | Efficacy analysis..... | 13 |
| 7.4.1 | Efficacy analysis methods..... | 13 |
| 7.4.2 | Statistical and analytical issues..... | 17 |
| 7.5 | Safety analysis..... | 20 |
| 7.5.1 | Extent of exposure..... | 20 |
| 7.5.2 | Adverse events..... | 21 |
| 7.5.3 | <i>Cutaneous safety</i> | 21 |
| 7.5.4 | Laboratory parameters..... | 22 |
| 7.6 | Other analyses..... | 22 |
| 7.7 | Analysis visit definition..... | 23 |
| 8 | CHANGES FROM THE PROTOCOL ANALYSIS PLAN..... | 23 |
| 9 | TABLE SHELLS AND REPORTING OUTPUT (GENERAL FEATURES)..... | 23 |
| 10 | REFERENCES..... | 25 |

1 OBJECTIVES

The purpose of this statistical analysis plan (SAP) is to describe the efficacy and safety of CD5789 50µg/g cream applied once daily for 12 weeks in subjects with moderate *acne vulgaris* to be included in the Clinical Study Report for Protocol RD.06.SPR.18252 dated 21 July 2015.

2 STUDY DESIGN

This is a Phase 3, multi-center, randomized, double blind, vehicle controlled study comparing CD5789 cream applied once daily in the evening versus its vehicle.

Group 1: CD5789 50µg/g cream applied once daily

Group 2: Vehicle Cream applied once daily

This Phase 3 study is designed to evaluate the efficacy and safety of CD5789 50µg/g in subjects with moderate acne vulgaris on the face and moderate acne vulgaris on the trunk. Subjects with an IGA of 3 and a minimum of 20 inflammatory lesions and 25 non-inflammatory lesion counts on the face and a PGA of 3 and a minimum of 20 inflammatory lesions and 20 non-inflammatory lesion but no more than 100 non-inflammatory lesion counts on the trunk will be enrolled in the study.

Subjects will be treated once daily for 12 weeks and evaluated at Baseline, Weeks 1, 2, 4, 8 and 12/Early Termination.

Specific requirements for children between 9 and 11 years old who do not have moderate acne on the trunk at Baseline (i.e. who do not have PGA of 3, at least 20 inflammatory lesions on the trunk and at least 20 non-inflammatory lesions on the trunk):

In this case, it is up to the investigators to decide if these 9 to 11 years old subjects should be treated or not for truncal acne. In these subjects, treatment can be started (based on investigator's decision) even if e.g. PGA=1, or the subject has less than 20 inflammatory lesions on the trunk and/or less than 20 non-inflammatory lesions on the trunk). These subjects will not be assessed for efficacy, but they will be assessed for safety and local tolerability. A total of approximately 1200 subjects from USA (approximately 40%), EU and Russia (approximately 60%) will be randomized to the study stratified by center using a 1:1 ratio, to have 600 subjects per treatment arm randomized.

Subjects will be randomized to study treatments in a 1:1 ratio within a block for CD5789 cream and Vehicle cream, respectively. Randomization will be stratified by clinical trial centers using the Interactive Response Technology (IRT) System.

The kit number, a unique number corresponding to the number on the label of the study medication, will be assigned to each eligible subject at Baseline.

3 EFFICACY AND SAFETY ASSESSMENT

3.1 Efficacy Assessment

IGA/PGA and Lesion Counts will be performed for face and trunk separately. The IGA/PGA assessments should be performed before the lesions counting.

3.1.1 IGA (Investigator’s global assessment) of facial acne

The areas defined for IGA assessment are forehead, each cheek, chin and nose. The IGA will be assessed at Screening, Baseline, and Weeks 1, 2, 4, 8 and 12/ET visits according to the following scale:

| Investigator’s Global Assessment Scale (IGA) Face | | |
|---|--------------|--|
| 0 | Clear | Clear skin with no inflammatory or non-inflammatory lesions. |
| 1 | Almost Clear | A few scattered comedones and a few small papules. |
| 2 | Mild | Easily recognizable; less than half the surface is involved. Some comedones and some papules and pustules. |
| 3 | Moderate | More than half of the surface is involved. Many comedones, papules and pustules. One nodule may be present. |
| 4 | Severe | Entire surface is involved. Covered with comedones, numerous papules and pustules. Few nodules may be present. |

3.1.2 PGA (Physician Global Assessment) of truncal acne

The PGA is a snapshot static assessment and is outlined in the following table and will be assessed at Screening, Baseline, and Weeks 1, 2, 4, 8 and 12/ET visits:

| Physician Global Assessment Scale (PGA) Trunk | | |
|---|--------------|---|
| 0 | Clear | Clear skin with no inflammatory or non-inflammatory lesions. |
| 1 | Almost Clear | A few scattered comedones and a few small papules. |
| 2 | Mild | Easily recognizable; less than half the surface is involved. Some comedones and some papules and pustules. |
| 3 | Moderate | More than half of the surface is involved. Many comedones, papules and pustules. One nodule may be present. |
| 4 | Severe | Entire surface is involved. Covered with comedones, numerous papules and pustules. Few nodules may or not be present. |

Specific requirements for subjects between 9 and 11 years old who do not have moderate acne on the trunk at Baseline (i.e. who do not have PGA of 3, at least 20 inflammatory lesions on the trunk and at least 20 non-inflammatory lesions on the trunk):

The PGA scale will be used by the investigators to decide if they want to start the treatment for truncal acne. However, the PGA data from these subjects will not be analyzed.

3.1.3 Lesions counts on the face and trunk

Inflammatory Lesions will be defined as follows:

Papule - A small, red, solid elevation less than 0.5 cm in diameter.

Pustule - A small, circumscribed elevation of the skin that contains yellow-white exudates.

Non-inflammatory Lesions will be defined as follows:

Open Comedone - A pigmented dilated pilosebaceous orifice (blackhead).

Closed Comedone - A tiny white papule (whitehead).

Note: The truncal lesion counts will not be performed for subjects between 9 and 11 years old who do not have moderate acne on the trunk at Baseline (i.e. who do not have PGA of 3, at least 20 inflammatory lesions on the trunk and at least 20 non-inflammatory lesions on the trunk).

3.1.4 Subject self-assessment of facial acne improvement

Subjects will evaluate by his/her own their facial acne improvement by comparing what they recall on their disease at the start of the study at the Week 12/ET visit according to the following scale:

| Subject's Assessment of Acne Scale | |
|------------------------------------|----------------------|
| 0 | Complete Improvement |
| 1 | Marked Improvement |
| 2 | Moderate Improvement |
| 3 | Minimal Improvement |
| 4 | No Change |
| 5 | Worse |

3.2 Safety Assessment

A safety assessment will be conducted for all subjects at the Screening visit (from the time of ICF signature) and every subsequent visit.

The safety parameters are the recording of adverse events, laboratory safety tests, physical examination, vital signs and local tolerability scores (0 [none] to 3 [severe]) for erythema, scaling, dryness, and stinging/burning as specified in sections 7.2.1 to 7.2.5 of the protocol.

Specific requirements for children between 9 and 11 years old who do not have moderate acne on the trunk at Baseline (i.e. do not have PGA of 3, at least 20 inflammatory lesions on the trunk and at least 20 non-inflammatory lesions on the trunk):

If the investigators decide to start the treatment for truncal acne in these subjects during the study, then the local tolerability scales (for trunk) will be completed and data will be evaluated in the safety analysis.

Local tolerability (erythema, scaling, dryness, and stinging/burning) will be assessed on the face and trunk separately.

3.3 Other Assessment

3.3.1 Photographs (at selected sites)

Facial and truncal photographs will be taken as per standardized procedure and guideline at the Baseline, Week 2, 4, 8, and Week 12/ET Visits at selected sites to illustrate the observed effect.

3.3.2 Dermatology Life Quality Index (DLQI) / Children's Dermatology Life Quality Index (C-DLQI)

The Dermatology Life Quality Index (DLQI, designed for use in adults, i.e. subjects over the age 16 years old) and Children's Dermatology Life Quality Index (C-DLQI, designed for 16 years or younger at the date of Baseline visit) will be completed at Baseline and Week 12/ET visits.

The DLQI/ C-DLQI will measure the dermatology-related limitations of functional ability and the frequency, severity and impact of acne symptoms on subjects' lives and acne-related quality of life. The six areas (domains) addressed in the questionnaire are: symptoms and feelings; daily activities; leisure; work/school; personal relationships; and treatment. The score on the DLQI/ C-DLQI has a possible range of 0 to 30 and can also be presented as a percentage of the maximum possible score of 30.

4 EFFICACY AND SAFETY ENDPOINTS

4.1 Efficacy Endpoints

The imputation of missing data for the efficacy endpoints is described in Section 7.4.2.2.

4.1.1 Primary Efficacy Endpoint

The primary efficacy endpoint consists of the following 3 co-primary endpoints:

1. Success Rate, defined as the percentage of subjects who achieve an IGA score of 1 (almost clear) or 0 (Clear) and an at least a 2-grade improvement from Baseline at Week 12.
2. Absolute Change in facial non-inflammatory lesion count from Baseline to Week 12
3. Absolute Change in facial inflammatory lesion count from Baseline to Week 12

4.1.2 Secondary Efficacy Endpoint

The secondary efficacy endpoint consists of the following 3 co-secondary endpoints:

1. Percentage of subjects who achieve a PGA score of 1 (almost clear) or 0 (Clear) and an at least 2 grade improvement from Baseline to Week 12.
2. Absolute change in truncal non-inflammatory lesion count from Baseline to Week 12
3. Absolute change in truncal inflammatory lesion count from Baseline to Week 12

4.1.3 Supportive Efficacy Endpoints

1. Percent change in facial non-inflammatory lesion counts from Baseline to Week 12.
2. Percent change in facial inflammatory lesion counts from Baseline to Week 12.
3. Percent change in truncal non-inflammatory lesion counts from Baseline to Week 12.
4. Percent change in truncal inflammatory lesion counts from Baseline to Week 12.
5. Subjects' assessment of facial acne improvement

4.2 Safety Endpoints

The safety endpoints are the following:

1. Treatment-emergent Adverse Events (TEAE): defined as an adverse event that occurred on or after the first study drug application.
2. Change from baseline of Local tolerability (Face)
3. Change from baseline of Local tolerability (Trunk)
4. Change from baseline of Laboratory Safety Test (Hematology, Chemistry, UA)

4.3 Other Endpoints

Other endpoints are the following:

1. Dermatology Life Quality Index (DLQI) / Children's Dermatology Life Quality Index(CDLQI)

5 POPULATIONS ANALYZED

The Intention to Treat (ITT) population will be used for all efficacy endpoints on the face. The Intention to Treat on the Trunk (ITTT) population will be used for all efficacy endpoints on the trunk. The Safety population (SAF) will be used for all safety analyses except for the local

tolerability on the trunk; and the Safety Population on the Trunk (SAFT) will be used for the analysis of local tolerability on the trunk. The Per Protocol (PP) population will be used for supportive analyses of the primary efficacy endpoint on the face; and the Per Protocol Population on the trunk (PPT) will be used for supportive analyses of the secondary efficacy endpoint on the trunk.

In the case that the SAFT turns out to be identical to the SAF population, then all safety analysis will be based on the SAF population.

In the case that the ITTT population turns out to be identical to the ITT population, then all efficacy analyses will be based on the ITT population.

In the case that the PPT population turns out to be identical to the PP population, then the supportive analyses of the primary and secondary efficacy endpoints will be based on the PP population.

5.1 Intention to Treat (ITT) Population

The ITT population is defined as any subjects who are randomized. Data from subjects included in the ITT population will be analyzed according to the treatment as randomized.

5.2 Intention to Treat on the Trunk (ITTT) Population

The ITTT population is defined as any subjects in the ITT population with moderate truncal acne at Baseline visit. In practice, this excludes from the ITT population the children between 9 and 11 years old who did not meet inclusion criterion #4 (PGA) or #5 (Truncal lesion counts) at study entry and therefore not planned to be treated on the trunk.

Subjects in the ITTT population will be analyzed according to the treatment as randomized.

5.3 Safety (SAF) Population

The SAF population is defined as comprising the ITT population subjects who applied/were administered the study medication at least once.

The SAF population is used for all safety analyses, except for the local tolerability on the trunk.

5.4 Safety (SAFT) Population for Local Tolerability on the Trunk

The SAFT population is defined as the SAF subjects who also applied/were administered the study medication on the trunk at least once.

5.5 Per Protocol (PP) Population

The PP population is defined as comprising the ITT subjects who have no major protocol deviations. The primary efficacy analyses (on the face) will be repeated based on the PP Population to confirm the results.

Potential major protocol deviations may include but are not limited to:

1. Entrance Criterion Deviations:
 - a. Out of range baseline lesion counts
 - b. Out of range baseline IGA grade
 - c. Inclusion/exclusion criterion violation
2. Non-Compliance:
 - a. Subjects who have dosing deviations more than 30% of the planned 84 doses
 - b. Subjects who miss doses for 5 or more consecutive days just prior to the last visit
3. Prohibited Medications: Subjects who have taken interfering concomitant therapies during treatment
4. Administrative error:
 - a. Accidental unblinding
 - b. Medication dispensing errors
 - c. Lesion counts and IGA/PGA performed by a non-approved evaluator

The final list of major protocol deviation criteria and subjects who have any major protocol deviations will be documented in the blind review memo before database lock.

5.6 Per Protocol (PPT) Population on the Trunk

The PPT population is defined as any subjects in the PP population with moderate truncal acne at the Baseline visit. In practice, this excludes from the PP population the children between 9 and 11 years old who did not meet inclusion criterion #4 (PGA) or #5 (Truncal lesion counts) at study entry (i.e. who do not have PGA of 3, at least 20 inflammatory lesions on the trunk and at least 20 non-inflammatory lesions on the trunk); or subjects who will be deemed not evaluable for the analyses on the trunk.

The secondary efficacy analyses (on the trunk) will be repeated based on the PPT Population to confirm the results.

6 SAMPLE SIZE CONSIDERATION

The assumptions used to power this study are based on the Phase 2 study (SPR.18223) on the Non-Japanese population. Success rates based on IGA are assumed 23% (CD5789) and 15% (Vehicle). For a 90% power to detect a significant difference versus vehicle using a two-sided test at a type I error of 0.05, 504 subjects per arm are needed for analysis.

Based on the results from statum 1 in the phase 2 study (SPR18223), the mean difference between the active and vehicle is estimated at 3 inflammatory lesions and 6 non-inflammatory lesions. Standard deviation of changes in lesion counts, based on phase 2 and on large phases 3 trials from other projects are estimated at 12 for inflammatory lesions and 20 for non-inflammatory lesions. Using these as assumptions and a sample of 504 evaluable subjects per arm, power is 95% at least for both lesion types.

To compensate for drop-outs and non-evaluable subjects (estimated at 16%), a total of 1200 subjects are to be randomized into each treatment group in a 1:1 ratio stratified by center, with 600 subjects in each treatment group randomized.

The assumptions for truncal acne being the same as for facial acne in terms of efficacy, therefore it is estimated that the power for truncal acne is adequate.

7 STATISTICAL METHODS AND DATA CONSIDERATIONS

For statistical analyses purpose, baseline is defined as the last measurement prior to the first application of the study drug.

A type I error of 0.05 (two-sided test) will be used to declare statistical significance.

In general, no formal inferential analyses are planned for the baseline and safety data, and only summary statistics will be provided.

For the summary statistics, the categorical will be summarized by frequency and percentage for each response category (N, %); and the continuous will be summarized using means, medians, minimum, maximum, and standard deviations.

SAS version 9.3 or above will be used for all statistical analyses.

7.1 Study subjects

7.1.1 Disposition of subjects

The number and percentage of subjects screened, randomized, completed, discontinued, and the primary reason for discontinuation based on the CRF exit form will be displayed.

In addition, the number and percentage of subjects in each study population (i.e., ITT, ITTT, SAF, SAFT, PP, and PPT) will be summarized by treatment group as well as by center.

7.1.2 Demographics and baseline characteristics

The following demographics variables will be summarized for the ITT population for each treatment group: age, Ethnicity, Race, Skin phototype, gender, and country of origin (US, EU, Russia). Age will be analyzed as a continuous variable and will be tabulated by age groups of “< 18years” (“9-11 years”, “12-14 years”, “15-17 years”), and “>= 18 years” (“18-64 years”, “>=65 years”).

Baseline IGA, PGA, facial lesion counts (inflammatory and non-inflammatory), and truncal lesion counts (inflammatory and non-inflammatory) will be summarized by treatment groups.

7.2 Protocol deviations

Number and percentage of subjects with major protocol deviations will be summarized for each treatment group in the ITT population. A listing of subjects with major protocol deviations will also be provided.

7.3 Medical history, previous and concomitant therapies/procedures

For statistical analysis purpose, previous therapies/procedures are defined as those ending at Baseline visit or before; and concomitant therapies/procedures are defined as those ongoing at the Baseline visit or starting after the Baseline visit.

Previous and concomitant therapies will be coded using WHODRUG dictionary). Previous and concomitant procedures will be coded using MedDRA dictionary.

A summary table will be provided for each of the following in the ITT population:

1. number and percentage of subjects who had previous therapies/medications by ATC text and WHO drug name
2. number and percentage of subjects who had concomitant therapies/medications by ATC text and WHO drug name
3. number and percentage of subjects who had previous procedures by System organ class and preferred term
4. number and percentage of subjects who had concomitant procedures by System organ class and preferred term

7.4 Efficacy analysis

7.4.1 Efficacy analysis methods

Table 1 summarizes the efficacy endpoints analyses.

7.4.1.1 *Analysis of the Primary Efficacy Endpoint*

Hypothesis testing

Formally stated the hypotheses to be tested for the primary efficacy endpoint are:

IGA Success Rate, $H_0: P_{\text{CD5789}} = P_{\text{Vehicle}}$,

$H_a: P_{\text{CD5789}} \neq P_{\text{Vehicle}}$

Where P_{\cdot} is the proportion of subjects in each treatment group from the overall population, who achieved the defined success criteria.

Change in facial Inflammatory and non-inflammatory lesions,

$$H_0: \delta_{CD5789} = \delta_{Vehicle}$$

$$H_a: \delta_{CD5789} \neq \delta_{Vehicle}$$

Where δ is the difference from baseline to week 12 for each treatment group from the overall population, in facial inflammatory and non-inflammatory lesion counts.

| Table 1 Analysis Strategy for Efficacy Endpoints | | | | |
|---|----------------------------------|--------------------|---------------------|-----------------------|
| Endpoint | Primary vs. Supportive Approach† | Statistical Method | Analysis Population | Missing Data Approach |
| Co-Primary | | | | |
| <ul style="list-style-type: none"> ▪ Success rate at week 12 (IGA) ▪ Change from baseline of facial inflammatory lesion counts at week 12 ▪ Change from baseline of facial non-inflammatory lesion counts at week 12 | P | CMH, ANCOVA | ITT | MI |
| | S | CMH, ANCOVA | PP | MI |
| | S | CMH, ANCOVA | ITT | MI (MNAR) |
| | S | CMH, ANCOVA | ITT | LOCF |
| | S | CMH, ANCOVA | ITT | Missing as Failure |
| Co-Secondary | | | | |
| <ul style="list-style-type: none"> ▪ Success rate at week 12 (PGA) ▪ Change from baseline of truncal inflammatory lesion counts at week 12 ▪ Change from baseline of truncal non-inflammatory lesion counts at week 12 | P | CMH, ANCOVA | ITTT | MI |
| | S | CMH, ANCOVA | PPT | MI |
| | S | CMH, ANCOVA | ITTT | MI (MNAR) |
| | S | CMH, ANCOVA | ITTT | LOCF |
| | S | CMH, ANCOVA | ITTT | Missing as Failure |
| Supportive | | | | |
| <ul style="list-style-type: none"> ▪ Percent change from baseline in facial inflammatory lesion counts at week 12 ▪ Percent change from baseline in facial non-inflammatory lesion counts at week 12) | P | CMHS | ITT | MI |
| | | | | |
| <ul style="list-style-type: none"> ▪ Subjects' assessment of facial acne improvement at week 12 | P | CMHS | ITT | As Observed |
| † P=Primary approach; S=Supportive approach ITT = Intention to Treat, ITTT = Intention to Treat for the truncal acne; PP = Per Protocol on the face; PPT = Per Protocol on the trunk; MI = Multiple imputation using the missing at random assumption; MI (MNAR) = Multiple imputation using the missing not at random assumption; LOCF=Last | | | | |

Observation Carried Forward.
CMH: Cochran-Mantel-Haenszel test stratified by analysis center.
CMHS: Cochran-Mantel-Haenszel test stratified by analysis center using the RIDIT score.
ANCOVA: Analysis of Covariance model including treatment, analysis center, and baseline lesion count as factors.

As the primary efficacy endpoint consists of 3 co-primary endpoints, all of which need statistical significance for the primary efficacy endpoint to be successful, there is no need for Type I error adjustment for the multiple comparisons of the 3 co-primary efficacy endpoints.

The study will be declared positive if the two-sided p-values is statistically significant (< 0.05) for the difference between the 2 treatment groups in all 3 co-primary efficacy endpoints, and the superiority of CD5789 50 $\mu\text{g/g}$ to Vehicle will be established.

Statistical method

- IGA Success Rate:

Success rate at Week 12 based on the IGA will be analyzed using the Cochran Mantel Haenszel (CMH) test stratified by analysis center (see definition in Section 7.4.2.4). The primary analysis will be performed using the overall ITT population based on the Multiple Imputation methodology (MI) as the primary imputation method for missing values, see Section 7.4.2.2.

The p-value for the treatment comparison will be generated from the general association statistic of the stratified CMH test. Difference in success rate between treatment groups (CD5789 50 $\mu\text{g/g}$ – Vehicle) and the 95% confidence interval of the difference will be based on the large sample approximation method for binary data⁽¹⁾. The success rate for each treatment group will also be displayed for the overall ITT population.

The treatment-by-analysis center interaction for success rate will be assessed for the overall ITT population by using the pair-wise Breslow-Day tests for homogeneity of the odds ratio across analysis centers at Week 12 at alpha level of 0.10. The consistency of the treatment effect across analysis centers will be evaluated using descriptive statistics and using graphical methods.

- Facial Lesion Counts:

Changes from baseline in facial lesion counts at Week 12 will be analyzed separately by type (inflammatory and non-inflammatory) using an ANCOVA (Analysis of Covariance) model including baseline count, analysis center, and treatment as factors. The primary analysis will be performed using the overall ITT population using the primary imputation method for missing values.

The p-value for the treatment comparison, estimate of the LSMEANS treatment difference (CD5789 50 $\mu\text{g/g}$ – Vehicle), and the 95% confidence interval of the LSMEANS difference will be generated from the ANCOVA model.

The treatment-by-analysis centers and the treatment-by-baseline lesion counts interactions for Change in Lesions Counts will be assessed separately by the interaction term at Week 12 at alpha level of 0.10 for the overall ITT population. The consistency of the treatment effect across

analysis centers will be evaluated using descriptive statistics and using graphical methods. The consistency of the treatment effect by baseline lesion counts will be evaluated using graphical methods.

In addition, change in Lesions Counts will be summarized descriptively at each visit for the overall ITT population.

Sensitivity analyses

The above primary efficacy analysis will be repeated using the PP population using the primary multiple imputation method for missing values.

To assess the robustness of the primary efficacy result, the following sensitivity analyses will be conducted for each of the 3 co-primary efficacy endpoints (see Section 4.1) using the ITT population:

- a. The same imputation method above will be repeated based on the Pattern-Mixture Model⁽²⁾ under the missing not at random (MNAR) assumption, by using the profiles from vehicle subjects with observed data to impute missing data.
- b. LOCF (Last observation carry forward) approach
- c. Subjects with missing IGA data at Week 12 will be considered as ‘failures’ for IGA success rate; and subjects with missing lesion counts data at Week 12 will be assigned the median change of lesion counts data from subjects who are ‘failures’ of IGA success within the same treatment group.

7.4.1.2 Analysis of the Secondary Efficacy Endpoint

The same statistical methods planned for the 3 co-primary efficacy endpoints will be used for the analyses of the 3 co-secondary efficacy endpoints using the overall ITTT population, with Multiple Imputation (MI) as primary imputation method for missing values (see Section 7.4.2.2).

If the superiority of CD5789 50 µg/g to Vehicle is established for the primary efficacy endpoint, then CD5789 50 µg/g will be declared superior to Vehicle in the secondary efficacy endpoint for the overall ITTT population if the two-sided p-value is statistically significant (< 0.05) for the difference between the 2 treatment groups in all 3 co-secondary efficacy endpoints.

As the testing of secondary efficacy endpoint is conditional on the success of primary efficacy endpoint, and all of the 3 co-secondary efficacy endpoints need statistical significance for the secondary efficacy endpoint to be successful, no Type I error adjustment is needed for the multiple comparisons of the 3 co-secondary efficacy endpoints.

The above secondary efficacy analysis will be repeated using the PPT population with the primary multiple imputation method for missing values.

The same sensitivity analysis methods to handle missing data planned for the primary efficacy endpoint will also be applied for the analyses of the secondary efficacy endpoint in the overall ITTT population.

7.4.1.3 Analysis of the Supportive Efficacy Endpoints

No Type I error adjustment will be needed for the analyses of the supportive efficacy endpoints.

- **Facial Lesion Counts:**

The analyses of Percent Change in facial lesion counts by type (Inflammatory Lesion Count and non-inflammatory lesion count) will be performed by the Cochran-Mantel-Haenszel test (CMH) stratified by analysis center. The analyses will be carried out in the overall ITT population with the primary multiple imputation methodology for missing data (see Section 7.4.2.2).

The p-value will be from the row mean difference statistic of the CMH test using RIDIT score.

In addition, percent change in facial lesion counts by type will be summarized descriptively by visit.

- **Truncal Lesion Counts:**

The same analysis method planned for the facial lesion counts will be used for the analyses of Percent Change in truncal lesion counts using the overall ITTT population.

- **Subjects' assessment of facial acne improvement:**

The analysis of the subjects' assessment of facial acne improvement will be based on the overall ITT population using observed data at Week 12 (i.e., no missing data imputation).

The p-value will be from the row mean difference statistic of the CMH test using RIDIT score.

In addition, the number and percent of subjects will be summarized for each treatment group by each scale of improvement.

7.4.2 Statistical and analytical issues

This section is intended to be completed by actual results during CSR writing.

7.4.2.1 Adjustment for covariates

The analysis of change from baseline in lesion counts will use an adjustment for the number of baseline lesions as described in Section 7.4.1.1.

7.4.2.2 Handling of dropouts or missing data

The primary method of imputation for missing data will be MI (Multiple Imputation) using the Missing At Random (MAR) assumption.

For the primary MAR based multiple imputation, the MI procedure of the SAS system will be used to generate sets of data with missing values imputed from observed data. It is expected that the pattern of missing data will be monotonic, with slight deviations being corrected by the Markov Chain Monte Carlo (MCMC) method of the MI procedure.

Linear regression will be employed to model the missing lesion count data and a logistic regression model will be used for the ordinal IGA/PGA scores, with the following covariates included in the imputation model: treatment and non-missing data from earlier timepoints. IGA/PGA success will be calculated from the imputed IGA/PGA scores. The imputed datasets will be analyzed using the methodology described in Section 7.4.1.1 for changes in lesion counts, and IGA/PGA success. The results from the analysis of the multiple imputed datasets will be combined by the MIANALYZE procedure of the SAS system. The seed number to be used will be the protocol number (18252) and the number of imputation is planned to be 5.

For success rate, the final p-value from the CMH test stratified by analysis center on each multiply imputed dataset will be generated using the Schafer, J. L.⁽³⁾ Multiple Imputation Methodology. The final difference of success rate between treatment groups (CD5789 50 µg/g – Vehicle) and the final 95% confidence interval of the difference from each multiply imputed dataset will be generated using the method by Bohdana Ratitch ed. al.⁽⁴⁾ .

For changes in lesion counts, the final p-value, treatment difference (CD5789 50 µg/g – Vehicle), and the 95% confidence interval of the treatment difference from the ANCOVA model on each multiply imputed dataset will be generated from the MIANALYZE procedure of SAS.

To assess the robustness of the primary efficacy results, the following sensitivity analyses will be also conducted:

- a. The same imputation method above will be repeated based on the Pattern-Mixture Model under the missing not at random (MNAR) assumption, by using the profiles from vehicle subjects with observed data to impute missing data.
- b. LOCF (Last observation carry forward) approach
- c. Subjects with missing IGA/PGA data at Week 12 will be considered as ‘failures’ for IGA/PGA success rate; and subjects with missing lesion counts data at Week 12 will be assigned the median change of lesion counts data from subjects who are ‘failures’ of IGA/PGA success within the same treatment group.

7.4.2.3 *Interim analyses and data monitoring*

No interim analysis is planned for this study.

7.4.2.4 *Multicenter studies*

A small center is defined as the center which randomizes less than 8 subjects. Small centers will be pooled prior to analyses. First, centers will be sorted by country, number of randomized subjects (descending order) and center number (ascending order). Pooling will start with

combining the largest of the set of small centers with the smallest center within that country. If there is a further need to combine data (the size of the pooled centers includes less than 8 subjects), the next smallest center will be combined with the next largest of the small centers, until the criterion of a minimum of 8 subjects is met. The process will continue until all pooled centers have a minimum of 8 subjects within the country. Any remaining centers will be pooled with the last pooled center within the country. The pooled centers will be referred to as 'analysis centers' in the statistical analyses.

7.4.2.5 *Multiple comparison/multiplicity*

Control of type 1 error, the experiment-wise error rate is controlled at alpha level of 0.05.

No adjustment for multiplicity is required for the primary efficacy endpoint as the statistical significance needs to be met for all 3 co-primary endpoints.

The statistical inferential testing for the secondary primary efficacy endpoint is conditional on the statistical significance of the primary efficacy endpoint, and all 3 co-secondary efficacy endpoints are required to reach statistical significance for the secondary efficacy endpoint to be statistically significant.

Therefore, no adjustment for multiplicity is required in this study.

7.4.2.6 *Use of an efficacy subset of patients*

The ITTT population is potentially a subset of the ITT population with moderate truncal acne at the Baseline visit. In practice, this excludes from the ITT population the children between 9 and 11 years old who did not meet inclusion criterion #4 (PGA) or #5 (Truncal lesion counts) at study entry (i.e. who do not have PGA of 3, at least 20 inflammatory lesions on the trunk and at least 20 non-inflammatory lesions on the trunk).

The PP/PPT Population will be determined prior to breaking the study blind using the criteria described in Section 5.5.

7.4.2.7 *Active-Control studies intended to show equivalence*

Not applicable.

7.4.2.8 *Examination of Subgroups*

All subgroup analyses are exploratory in nature.

To determine whether the treatment effect is consistent across various subgroups, the estimate of the between-treatment effect (with a nominal 95% CI) for the primary efficacy endpoint will be estimated and plotted within each category of the following classification variables:

- Age Group (<18 years, >=18 years)
- Gender (female, male)

- Race (white, non-white)
- Country (US, EU, Russia)
- Skin Phototype (I-III, IV-VI)

Treatment effects and nominal 95% confidence intervals by category for the classification variables listed above will be reported as well as presented graphically. Formal statistical testing of these interactions will not be performed.

For success rate on IGA, treatment effects and the nominal 95% confidence intervals will be summarized by category for the subgroup variable using the same method as the primary analysis.

For the change in lesion counts by type, treatment effects and the nominal 95% confidence intervals will be summarized by category for the subgroup variable using two-sample t-test.

7.5 Safety analysis

All safety analyses are descriptive, and no formal inferential testing will be performed.

7.5.1 Extent of exposure

7.5.1.1 Study Duration

Study duration (day) is defined as the date of the last visit minus the date of the Baseline visit plus one. Study duration will be summarized by descriptive statistics using the ITT population.

7.5.1.2 Treatment Duration

Treatment duration (day) is defined as “the date of the last application on face or the trunk, whichever is later, minus the date of the first application on face or the trunk, whichever is earlier, plus one”.

Treatment duration will be summarized by descriptive statistics using the SAF population.

7.5.1.3 Study Medication Usage (number of applications)

The study medication usage will be based on data collected on the CRF in terms of the number of applications applied.

Number of applications applied is calculated as the expected number of applications minus the total number of missed doses/applications during the treatment. If the number of missed doses/applications is unknown, then 50% doses will be assumed to be missed between two adjacent visits.

The expected number of applications is defined as the planned total number of applications during the exposed period.

Number of applications applied and the daily average applications applied will be summarized by treatment group for the SAF population.

In addition, study medication usage may also be summarized for the face using SAF population and for the trunk using the SAFT population separately.

7.5.2 Adverse events

A treatment-emergent adverse event (TEAE) is defined as an adverse event that occurred on or after the first study drug application.

The frequency (N %) of each adverse experience in the trial will be presented by system organ class and preferred term by treatment group using the SAF population.

Adverse events will be coded using the MedDRA dictionary

Overall summary of TEAEs [frequency (N, %)] will be provide for each treatment group. Additional summaries by System Organ Class (SOC) and Preferred Term (PT) will also be provided, along with TEAEs of special interest, serious TEAEs (SAEs), TEAEs related to the study drug, and TEAEs leading to discontinuation and severe TEAEs.

Subgroup analyses by Age Group (<18 years, >=18 years), Race (white, non-white), Gender (female, male), Country (Russia, EU, US) and Skin Phototype (I-III, IV-VI) will be provided.

All AE summary tables are based on the number and percentage of subjects who have experienced TEAE(s). For a given TEAE, a subject will be counted once even if he or she has experienced multiple episodes for that particular TEAE.

7.5.3 Cutaneous safety

Only local tolerability (erythema, scaling, dryness, stinging/burning) worsened from baseline will be analyzed.

Analyses of the local tolerability on the face will be based on the SAF population; and the analyses of the local tolerability on the trunk will be based on the SAFT population.

Number and percent of subjects with worsened local tolerability from baseline will be presented for each treatment group by the severity scale for each of the assessments (erythema, scaling, dryness, stinging/burning) by visit.

Number and percent of subjects with worst local tolerability worsened from baseline during treatment period will also be presented for each treatment group by the severity scale for each of the assessments (erythema, scaling, dryness, stinging/burning). The same will be presented for the final local tolerability worsened from baseline during the treatment period.

Summary statistics by treating the severity scale as continuous outcome will also be presented.

Subgroup analyses by Age Group (<18 years, >=18 years), Race (white, non-white), Gender (female, male), Country (Russia, EU, US) and Skin Phototype (I-III, IV-VI) will be provided.

7.5.4 Laboratory parameters

A summary table of the categorical grade shift changes using the normal ranges from baseline to Week 12/ET will be presented for each treatment group by each laboratory category parameter using the SAF population.

Only subjects with both baseline and Week 12/ET assessments will be included in this analysis.

7.6 Other analyses

For the other variables, only observed data will be summarized without imputation for missing data. These data will be analyzed using the same methodology as described in Section 7.4.1.3.

The Dermatology Life Quality Index (DLQI, designed for use in adults, i.e. subjects over the age 16 years old) and Children’s Dermatology Life Quality Index (C-DLQI, designed for use in children, i.e. subjects from age 5 to age 16) will be summarized descriptively.

Only SAF subjects who have both baseline and Week 12/ET assessments using the same age-appropriate questionnaire at both visits will be included in the analyses.

Total Scores and Change from Baseline at Week 12/ET for DLQI/C-DLQI, DLQI only, and C-DLQI only will be summarized.

In addition, number and percent of subjects will be reported by treatment group for each of the following categories of the effect of disease on the quality of life as specified in Table 2:

| Table 2 DLQI/C-DLQI Categories | | | |
|--------------------------------|----------------------|------------------------|---------------|
| DLQI | | C-DLQI | |
| DLQI Category | Range of Total Score | C-DLQI Category | Score Maximum |
| No effect | 0 - 1 | No effect | 0 - 1 |
| Small effect | 2 - 5 | Small effect | 2 - 6 |
| Moderate effect | 6 - 10 | Moderate effect | 7 - 12 |
| Very large effect | 11 - 20 | Very large effect | 13 - 18 |
| Extremely large effect | 21 - 30 | Extremely large effect | 19 - 30 |

Summaries of DLQI only and C-DLQI only by each dimension as defined in Table 3 below will also be provided.

| Table 3 DLQI/C-DLQI Dimensions | | | |
|---------------------------------------|---------------|------------------------|---------------|
| DLQI | | C-DLQI | |
| DLQI dimensions | Score Maximum | C-DLQI dimensions | Score Maximum |
| Total Score | 30 | Total Score | 30 |
| Symptoms and feelings | 6 | Symptoms and feelings | 6 |
| Daily Activities | 6 | Sleep | 3 |
| Leisure | 6 | Leisure | 9 |
| Work and School | 3 | School or holidays | 3 |
| Personal relationships | 6 | Personal relationships | 6 |
| Treatment | 3 | Treatment | 3 |

7.7 Analysis visit definition

All efficacy variables and local tolerability variables will be summarized and analyzed by analysis visit. Analysis visit will be imputed according to the following algorithm to summarize the data by proper visit window interval. Study day is calculated as visit date minus the date of first application plus 1.

| Table 4 Analysis Visit and Visit Window | | | |
|--|---|-------------------------|---------------------------------|
| Analysis Visit | Analysis Visit Number in Derived Dataset | Target Study Day | Visit Window (Study Day) |
| Baseline | 1 | 1 | [<= 1] |
| Week 1 | 2 | 8 | [2 – 11] |
| Week 2 | 3 | 15 | [12 – 21] |
| Week 4 | 4 | 29 | [22 – 42] |
| Week 8 | 5 | 57 | [43 – 70] |
| Week 12 | 6 | 85 | >70 |

If multiple measurements are taken in the same interval, the one closest to the target study day will be used for the analysis. If two measurements are taken with equal differences in timing compared with the target date, the nominal visit number (recorded on the CRF page) will be used.

8 CHANGES FROM THE PROTOCOL ANALYSIS PLAN

Any change from the protocol will be justified and fully documented.

9 TABLE SHELLS AND REPORTING OUTPUT (GENERAL FEATURES)

The final list of tables and figures and their shells for the reporting of this study will be available in a separate document to be developed and will be finalized before database lock.

Details of analysis specifications including but not limited to the SAS code will be documented on the table shells.

10 REFERENCES

- (1). Fleiss, J. L. “Statistical methods for rates and proportions”, 2nd ed. 1981, New York: John Wiley & Sons, Inc.
- (2). B. Ratitch, M. O’Kelly and R. Tosiello, “Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models”, January 2013, Pharmaceutical Statistics.
- (3). Schafer, J.L. “Analysis of Incomplete Multivariate Data”, page 115, 1st ed. 1997, Chapman & Hall/CRC Monographs on Statistics & Applied Probability.
- (4). Bohdana Ratitch, et al. “Combining Analysis Results from Multiply Imputed Categorical Data”, 2013, PharmaSUG Proceedings, Paper SP-03.

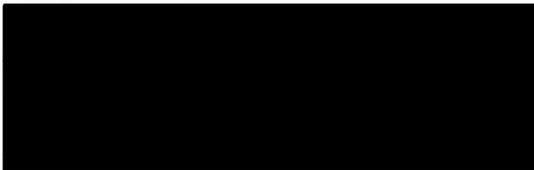
RD.06.SPR.18252
Statistical Analysis Plan Addendum

A MULTI-CENTER, RANDOMIZED, DOUBLE-BLIND,
PARALLEL-GROUP, VEHICLE-CONTROLLED STUDY TO
COMPARE EFFICACY AND SAFETY OF CD5789 50 μ g/g
CREAM VERSUS VEHICLE CREAM IN SUBJECTS
WITH ACNE VULGARIS

Phastar Statistician



Galderma Statistician
Name:



1. Purpose

The purpose of this document is to give details and clarifications about the analysis as presented in the Statistical Analysis Plan dated October 5, 2015. Any changes from the planned analysis as described in that document are also detailed here, and any differences described here supersede the analysis as presented in the SAP.

2. General Clarification and Considerations

2.1. Decimal Precision

Unless otherwise noted, means and medians will be presented to one more decimal place than the source data; standard deviations will be presented to two more decimal places than the source data; minimums and maximums will be presented to the same number of decimal places as the source data, percentages will be presented to one decimal place; confidence intervals will be presented to two decimal places; and p-values will be presented to three decimal places.

2.2. Missing Date Procedures

Adverse events with completely missing dates will be considered treatment-emergent. Medications with completely missing end dates will be considered concomitant. Adverse events and medications with partially missing start or end dates will be considered treatment-emergent and concomitant respectively unless the non-missing portion of the dates definitively proves otherwise.

For example, if a subject starts treatment on 10FEB2015, then adverse events with onset dates of FEB2015, 2015, or 05DEC would all be considered treatment-emergent, while onsets dates of 2014 or JAN2015 would not be considered treatment-emergent. Medications starting or ending in FEB2015 or 2015 would be considered concomitant, medications ending in JAN2015 or 2014 would not.

2.3. Coding Versions

Adverse events, procedures, and non-drug therapies are coded using MedDRA 18.0. Medications and therapies are coded using WHO-DDE v2013 March.

2.4. Subgroup Analysis

Subgroup analyses for IGA and PGA success rate at Week 12, absolute change from baseline for inflammatory and non-inflammatory lesion counts at Week 12 on the face and trunk, treatment-emergent adverse events by SOC and PT, and local tolerability parameters worsened from baseline on the face and trunk will include an analysis by:

- Age: 9-11 years, 12-17 years, and ≥ 18 years of age
- Gender: female and male
- Age and gender: female < 25 years, male < 25 years, female ≥ 25 years, and male ≥ 25 years
- Race: White, Black or African American, Asian, and Other
- Ethnicity: Hispanic or Latino and Not Hispanic or Latino
- Skin Phototype: I-III and IV-VI

In addition, a subgroup analysis for IGA and PGA success rate at Week 12, treatment-emergent adverse events by SOC and PT, and local tolerability parameters worsened from baseline on the face and trunk will be provided for the additional age categories of: <14 Years, 14-17 Years, and >=18 Years.

2.5. Subpopulations

Three subpopulations of the ITT population will be studied using the primary and secondary efficacy analyses. Those populations are defined as:

- Subjects with at least 75% non-inflammatory baseline lesions. A subject's baseline total non-inflammatory lesion count will be divided by their baseline total lesion count (inflammatory and non-inflammatory). If the result is greater than or equal to 0.75, they will be included in the subpopulation. Presence of nodules and/or cysts will not be included for the calculation of inflammatory lesion counts.
- Subjects with no prior use of topical or oral retinoids. Use of topical and oral retinoids are identified using the ATC4 codes of D10AD and D10BA from the concomitant medication dataset.
- Subjects with no prior use of medications for acne. Subjects with use of prior medication of acne are identified using the following indication terms from the concomitant medications dataset:
 - ACNE
 - ACNÉ
 - ACNE ULGARIS
 - ACNE CONGLOBATA
 - ACNE FACE WASH
 - ACNE FACE/CHEST/BACK
 - ACNE ON FACE
 - ACNE VULARIS
 - ACNE VULGARIS
 - ACNE VULGARIS AND PITYROSPORUM FOLLICULITIS
 - ACNE VULGARIS OF FACE
 - ACNE VUOLGARIS
 - ACNE WASH
 - AKNE VULGARIS
 - FACIAL ACNE
 - FACIAL ACNE VULGARIS

2.6. Site Pooling

Analysis centers will be created by totaling the number of subjects at each site and pooling smaller sites together.

Sites with fewer than 8 subjects will be pooled. Sites will be sorted by country, the number of subjects at that site in descending order, and then site number in ascending order. For each country separately, the site at the top of the list (the largest site with less than 8 subjects and, if tied, the smallest site number) will be pooled with the site at the bottom of the list (the smallest site with less than 8 subjects and, if tied, the largest site number). If the pooled site now has at least 8 subjects, it is removed from the list.

Otherwise, it will continue to be combined with the smallest available site until it has at least 8 subjects or no sites remain.

After all possible pooling is done, if all remaining small sites cannot be pooled together into an analysis center with at least 8 subjects, those sites will all be pooled into the analysis center created in the previous iteration of the pooling, even if that analysis center is not the smallest. If there was no other pooling done (because all small sites in the country total to less than 8 subjects), those sites will be pooled with the smallest site in the country. If no other sites are available in that country, the site will remain with fewer than 8 subjects.

3. Protocol Deviations

Protocol deviations will be provided in the DV SDTM dataset. Subjects with a deviation categorized as “Major” or “Major for face” will be excluded from the PP dataset. Subjects with a deviation categorized as “Major,” “Major for face,” or “Major for trunk” will be excluded from the PPT dataset.

The number of subjects in each analysis population will be summarized, and specific protocol deviations will be summarized in a listing.

4. Disposition

For the purposes of disposition summaries, subjects are considered to be Screened if they have signed the Informed Consent form. Subjects will be considered randomized once they’ve been assigned a treatment.

5. Prohibited Medications

Prohibited medications will be flagged on the concomitant medications listing. A list of prohibited medications with ATC/Drug names will be identified prior to the database lock.

6. Efficacy Analyses

6.1. IGA/PGA Success Rate

Success rate of IGA and PGA will be calculated for each visit. Success rate will be calculated as the number of subjects considered a success at that visit divided by the number of subjects with IGA/PGA data at that visit.

The large-sample approximation for binary data will be used for the 95% confidence intervals around the difference in success rates. Continuity corrections will not be used for either the CMH test or the generation of 95% confidence intervals.

The success rate of IGA and PGA using missing data procedures will be presented at Week 12. The success rate of IGA and PGA will be presented at each visit using the observed data; no missing data imputation will be used.

In addition, overall success rate at Week 12 will be calculated using missing data procedures assuming that the data is missing at random. Overall success rate is defined as having an IGA score of “clear” or

“almost clear” at Week 12 and a grade change of at least two from baseline and having a PGA score of “clear” or “almost clear” at Week 12 and a grade change of at least two from baseline.

6.2. Multiple Imputation – Missing at Random

A multiple imputation method will be used to account for missing data on efficacy analyses. The following steps will be followed:

- 1) For lesion counts, the pattern of missingness in the data will be evaluated. If the data is not monotone missing, the MCMC method will be used to make it monotone missing. The single chain method will be used, with 200 burn-in iterations and 100 iterations between imputations. Fifty imputations will be created. The seed number will be 18252. Lesion counts generated as a result of the MCMC method will be re-imputed if they are below zero; they will not be rounded.
- 2) The MI procedure in SAS will be used to create multiple imputations of the data that have no missing values. If the MCMC method was previously employed, one imputation will be made using each of the fifty MCMC-imputed datasets. If the MCMC method was not used, fifty imputations will be created using the MI procedure assuming the data is Missing at Random. The seed number will be 18252. These imputations will use the following models:
 - a. For IGA and PGA scores, a logistic regression model will be used with covariates for treatment and non-missing data from earlier scheduled time points including baseline. Weeks will be imputed sequentially, with IGA/PGA data at prior non-baseline weeks treated as continuous data (baseline will be treated as categorical).
 - b. For lesion counts, a linear regression model will be used with covariates for treatment and non-missing data from earlier scheduled time points including baseline.
- 3) The imputed datasets will be analyzed as follows:
 - a. For IGA and PGA scores, the number of successes at Week 12 will be analyzed using the CMH test stratified by analysis center. The sample difference in proportions and its standard error will also be calculated. No continuity corrections will be used.
 - b. For lesion counts, change from baseline to Week 12 in lesion counts will be analyzed using an ANCOVA model with covariates for baseline count, analysis center, and treatment.
- 4) The resulting analysis on the imputed datasets will then be combined to produce a single set of statistics.
 - a. For IGA and PGA scores, the results from the CMH analysis will be combined using the procedure by Rubin (1987) and Li et al. (1991) to produce a pooled CMH statistic and p-value. The differences in proportions and standard errors will be combined using the MIANALYZE procedure in SAS; the resulting pooled difference and standard errors will be used to produce 95% confidence intervals based on the large-sample approximation method for binary data without the use of a continuity correction. Both methods will be used as described in the Bohdana Ratitch et al. paper *Combining Analysis Results from Multiply Imputed Categorical Data*.
 - b. For lesion counts, results from the ANCOVA analysis will be combined using MIANALYZE procedure in SAS.

Below is an example of the SAS syntax for the MCMC and logistic MI procedures for IGA/PGA:

```
proc mi data=transI seed=18252 out=miI1 nimpute=50 noprint;
class trt01pn baseline week_1;
var trt01pn baseline week_1;
monotone logistic(week_1=trt01pn baseline);
run;
```

```
proc mi data=miI1 seed=18252 out=miI2 nimpute=1 noprint;
by _IMPUTATION_;
class trt01pn baseline week_2;
var trt01pn baseline week_1 week_2;
monotone logistic(week_2=trt01pn baseline week_1);
run;
```

```
proc mi data=miI2 seed=18252 out=miI3 nimpute=1 noprint;
by _IMPUTATION_;
class trt01pn baseline week_4;
var trt01pn baseline week_1 week_2 week_4;
monotone logistic(week_4=trt01pn baseline week_1 week_2);
run;
```

```
proc mi data=miI3 seed=18252 out=miI4 nimpute=1 noprint;
by _IMPUTATION_;
class trt01pn baseline week_8;
var trt01pn baseline week_1 week_2 week_4 week_8;
monotone logistic(week_8=trt01pn baseline week_1 week_2 week_4);
run;
```

```
proc mi data=miI4 seed=18252 out=miI(rename=_imputation_=imp) nimpute=1 noprint;
by _IMPUTATION_;
class trt01pn baseline week_12;
var trt01pn baseline week_1 week_2 week_4 week_8 week_12;
monotone logistic(week_12=trt01pn baseline week_1 week_2 week_4 week_8);
run;
```

Below is an example of the SAS syntax for the MCMC and linear MI procedures for lesion counts:

```
proc mi data=transI seed=18252 out=mcmcI(rename=_imputation_=imp) nimpute=50 minimum=0 noprint;
mcmc impute=monotone chain=single nbiter=200 niter=100;
var baseline week_1 week_2 week_4 week_8 week_12;
run;
```

```
proc mi data=mcmcI seed=18252 out=miI nimpute=1 minimum=. 0 0 0 0 0 0 noprint;
by imp;
class trt01pn;
var trt01pn baseline week_1 week_2 week_4 week_8 week_12;
monotone reg(week_12=trt01pn baseline week_1 week_2 week_4 week_8);
run;
```

6.3. Multiple Imputation – Missing Not at Random

As a sensitivity analysis, the multiple imputation process will be repeated assuming that the data is missing not at random. This process will be the same as the one described for the Missing at Random analysis, but with the assumptions changed. The Missing Not at Random dataset will use data collected from vehicle subjects to impute missing observations.

Below is an example of the SAS syntax for the logistic MI for IGA/PGA for the missing not at random assumption:

```
proc mi data=transI seed=18252 out=miI1 nimpute=50 noprint;
class trt01p baseline week_1;
var baseline week_1;
monotone logistic(week_1=baseline);
mnar model(week_1 / modelobs=(trt01p="Vehicle"));
run;
```

```
proc mi data=miI1 seed=18252 out=miI2 nimpute=1 noprint;
by _IMPUTATION_;
class trt01p baseline week_2;
var trt01p baseline week_1 week_2;
monotone logistic(week_2=baseline week_1);
mnar model(week_2 / modelobs=(trt01p="Vehicle"));
run;
```

```
proc mi data=miI2 seed=18252 out=miI3 nimpute=1 noprint;
by _IMPUTATION_;
class trt01p baseline week_4;
var trt01p baseline week_1 week_2 week_4;
monotone logistic(week_4=baseline week_1 week_2);
mnar model(week_4 / modelobs=(trt01p="Vehicle"));
run;
```

```
proc mi data=miI3 seed=18252 out=miI4 nimpute=1 noprint;
by _IMPUTATION_;
class trt01p baseline week_8;
var trt01p baseline week_1 week_2 week_4 week_8;
monotone logistic(week_8=baseline week_1 week_2 week_4);
mnar model(week_8 / modelobs=(trt01p="Vehicle"));
run;
```

```
proc mi data=miI4 seed=18252 out=miI(rename=( _imputation_ =imp)) nimpute=1 noprint;
by _IMPUTATION_;
class trt01p baseline week_12;
var trt01p baseline week_1 week_2 week_4 week_8 week_12;
monotone logistic(week_12=baseline week_1 week_2 week_4 week_8);
mnar model(week_12 / modelobs=(trt01p="Vehicle"));
run;
```

Below is an example of the SAS syntax for the MCMC and linear MI procedures for lesion counts for the missing not at random assumption:

```
proc mi data=transI seed=18252 out=mcmcI(rename=_imputation_=imp) nimpute=50 minimum=0 noprint;
mcmc impute=monotone chain=single nbiter=200 niter=100;
var baseline week_1 week_2 week_4 week_8 week_12;
run;

proc mi data=mcmcI seed=18252 out=miI nimpute=1 minimum=. 0 0 0 0 0 0 noprint;
by imp;
class trt01p;
var baseline week_1 week_2 week_4 week_8 week_12;
monotone reg(week_12=baseline week_1 week_2 week_4 week_8);
mnar model(week_1 / modelobs=(trt01p="Vehicle"));
mnar model(week_2 / modelobs=(trt01p="Vehicle"));
mnar model(week_4 / modelobs=(trt01p="Vehicle"));
mnar model(week_8 / modelobs=(trt01p="Vehicle"));
mnar model(week_12 / modelobs=(trt01p="Vehicle"));
run;
```

6.4. Multiple Imputation – Categorical Sensitivity Analysis

The primary multiple imputation method will treat IGA/PGA as continuous when used as covariates for the logistic regression model. As a sensitivity analysis, the multiple imputation process for IGA/PGA will be repeated treating those predictors as categorical.

First, the pattern of missingness in the data will be evaluated. If the data is not monotone missing, the MCMC method will be used to make it monotone missing. The single chain method will be used, with 200 burn-in iterations and 100 iterations between imputations. Fifty imputations will be created. If all subjects have the same baseline IGA/PGA value, the baseline visit will not be included in the model. The seed number will be 18252. IGA and PGA scores generated as a result of the MCMC method will be rounded prior to the next step; scores will be re-imputed if, after rounding, they are lower than 0 or higher than 4.

Due to the categorical nature of the data, and the rarity of some values of IGA/PGA, there is the potential for convergence issues when using logistic regression with categorical covariates. If this is the case, each week will be imputed sequentially and predictor weeks that cause convergence issues (and thus can't be used as valid predictors) will be dropped.

The model will be chosen by first checking if the inclusion of any covariates causes convergence issues. If so, predictor weeks will be removed from the model and the model rerun until any convergence issues are solved. Earlier weeks will be dropped from the model before later weeks, and combinations of weeks will only be dropped if the removal of only one week does not cause the model to converge. Thus, weeks will be dropped in the following order:

1. Week 1
2. Week 2
3. Week 4
4. Week 8
5. Weeks 1 and 2
6. Weeks 1 and 4
7. Weeks 2 and 4
8. Weeks 1 and 8
9. Weeks 2 and 8
10. Weeks 4 and 8

11. Weeks 1, 2, and 4
12. Weeks 1, 2, and 8
13. Weeks 2, 4, and 8

After each removal, the model will be rerun. The first model that converges will be kept, and the process will restart with the next week until Week 12 is imputed. The final chosen models will be run with seed number 18252. Below is an example of the SAS syntax for the MCMC and logistic MI procedures treating predictor weeks as categorical, with no weeks removed:

```
proc mi data=transI seed=18252 out=mcmcI(rename=_imputation_=imp) nimpute=50 minimum=0 maximum=4 round=1;
mcmc impute=monotone chain=single nbiter=200 niter=100;
var trt01pn baseline week_1 week_2 week_4 week_8 week_12;
run;

proc mi data=mcmcI seed=18252 out=miI nimpute=1 noprint;
by imp;
class trt01pn baseline week_1 week_2 week_4 week_8 week_12;
var trt01pn baseline week_1 week_2 week_4 week_8 week_12;
monotone logistic(week_1=trt01pn baseline);
monotone logistic(week_2=trt01pn baseline week_1);
monotone logistic(week_4=trt01pn baseline week_1 week_2);
monotone logistic(week_8=trt01pn baseline week_1 week_2 week_4);
monotone logistic(week_12=trt01pn baseline week_1 week_2 week_4 week_8);
run;
```

6.5. Analysis of Percent Change in Lesion Count

The percent change in lesion count will be analyzed using a CMH test using ridit scoring. The resulting test statistics for the row-mean scores difference from the CMH analysis will be combined using the procedure by Rubin (1987) and Li et al. (1991) to produce a pooled CMH statistic and p-value.

6.6. Missing as Failure Analysis of Lesion Counts

The missing-as-failure analysis will be used for lesion counts to impute missing lesion count change from baseline at week 12.

For missing facial lesion count data, the median change from baseline in lesion counts from among all observed IGA failures at Week 12 will be substituted. For missing truncal lesion count data, the median change from baseline in lesion counts from among all observed PGA failures at Week 12 will be used.

In both cases the median will be calculated separately for each treatment group, using actual treatment, and the imputation will use the median that corresponds to the treatment group of the subject being imputed.

6.7. Efficacy Figures

Figures will be presented for IGA and PGA success rate by visit using observed data. The week 12 success rate using the same MAR multiple imputation method as used in the primary analysis will also be graphed.

Figures will also be presented for the success rate at week 12 for each subgroup using MAR multiple imputation.

6.8. Efficacy Subgroup Analysis

The subgroup analyses will repeat the analysis of the three co-primary and co-secondary endpoints on all subgroups. The MAR multiple imputation method will be used to handle missing data.

For lesion count subgroup analysis, the t-tests will be done assuming unequal variances using Satterthwaite's formula for variance.

6.9. Dermatology Life Quality Index

Descriptive statistics treating the (C)DLQI totals scores as a continuous measure will be presented with observed and change from baseline summaries. Counts and percentages of subjects within categorized ranges ("No effect" to "Extremely large effect") will also be presented. These summaries will be presented for the DLQI and CDLQI.

Total score is calculated assigning a numerical equivalent to each answer, and summing them across the 10 questions. Answers are scored as follows:

- Very much: 3
- A lot/Quite a lot: 2
- A little/Only a little: 1
- Not at all: 0

For question 7, answering a "Yes" scores as 3 and a "No" scores as 0, however answering "No" requires answering question 7b where "A lot," "A little," and "Not at all" as scored as normal.

Descriptive statistics treating the (C)DLQI dimensional scores as a continuous measure will also be presented with observed and change from baseline summaries. These will be presented for DLQI and CDLQI separately.

For the DLQI, dimensional scores will be calculated using the following questions:

- Symptoms and Feelings: questions 1 and 2
- Daily Activities: questions 3 and 4
- Leisure: questions 5 and 6
- Work and School: question 7
- Personal Relationships: questions 8 and 9
- Treatment: question 10

For the CDLQI, dimensions cores will be calculated using the following questions:

- Symptoms and Feelings: questions 1 and 2
- Sleep: question 9
- Leisure: questions 4, 5, and 6
- School or Holidays: question 7
- Personal Relationships: questions 3 and 8
- Treatment: question 10

7. Safety Analyses

7.1. Drug Exposure

Study and Treatment Duration

Study duration will count the number of days a subject is on the study, regardless of their treatment requirements. It is calculated as study end date minus study start date plus one.

Subjects are expected to take one application per day on the study. Treatment duration for the face will be calculated as the subject's treatment end date minus treatment start date plus one; treatment duration on the trunk will be calculated as the subject's treatment end date minus treatment start date plus one.

Total Applications

Applications taken is not collected on the CRF, so to estimate the number of applications a subject took, the number of recorded missed doses will be subtracted from the number of doses the subject should have taken to arrive at the number of taken applications.

Subjects who have visits after their treatment end date record the time between in one of two different ways:

1. Some record a number of missed doses between treatment end and their current visit. These do not count as actual missed doses, as the subject was off treatment.
2. Others record "NA" or leave it missing.

Thus, the calculation of total applications on the face will be done in the following steps:

1. Start with the subject's treatment start date.
2. Chose an end date to use to calculate the starting number of doses measured on the CRF. If the subject has a non-missing recorded number of missed doses at any date after their treatment end date, use the latest date available with missed doses recorded. Otherwise, use treatment end date.
3. Calculate the starting number of doses as end date minus treatment start date plus 1.
4. From that, the total number of all missed doses on the face will be subtracted..

The calculation of total applications on the trunk will be done following the same steps, but using PGA scores and counts of missing doses on the trunk.

Expected Applications

A subject's expected number of applications will assume that subjects take one application per day of treatment duration. Thus, a subject's expected applications is equal to their treatment duration.

Compliance

Compliance percentage will be calculated as a subject's total applications divided by their expected applications multiplied by 100.

7.2. Medication Usage

A subject's average daily medication usage (g/day) will be calculated as their total medication usage divided by their overall treatment duration. Overall treatment duration is calculated as the subject's treatment end date minus their treatment start date plus one.

As per Galderma's clinical supply process, the weight of each dispensed tube is assigned based on the average weight of a sample of kits from each batch. Hence, it is possible that some subjects are recorded as having returned more grams of drug than they were dispensed. If a subject's total medication usage is negative, it will be set to zero.

7.3. Adverse Events

A treatment-emergent adverse event is defined as an adverse event that occurred, or increased in severity, on or after the date of the first dose of study drug. Events with missing severity will be considered severe; events with missing relatedness data will be considered related to the study drug.

Adverse events of special interest will be summarized, and are defined as follows:

- Out-of-range laboratory results that are identified as clinically significant and related to the study drug,
- Related cutaneous AEs which lead to permanent treatment discontinuation, and
- Suspicion of allergic reaction suspected to be related to the study drug.

Severe adverse events are defined as those with severity marked as "Severe" on the CRF. Adverse events related to study drug or protocol procedure are defined as those marked as having a "reasonable possibility" of being related to study drug or protocol procedure, respectively, on the CRF.

A listing of pre-treatment adverse events will be included. Pre-treatment adverse events are those with an onset date prior to the date of the first application of study drug.

7.4. Local Tolerability

Local tolerability will be summarized by visit on the face and trunk separately. Summaries will present descriptive statistics of local tolerability scores when treated as a continuous measure, as well as counts and percentages of subjects with each score treated as categorical. Categorical summaries of scores that are worsened from baseline will also be presented.

Worsened from baseline is defined as having a local tolerability assessment at the summarized time point that is greater than the local tolerability assessment at baseline.

Subgroup analysis of local tolerability scores will be presented separately for the face and trunk. These summaries will treat the scores as categorical, and will present scores that have worsened from baseline for both the final and worst scores per subject.

Local tolerability scores will also be presented treating the score as dichotomous. Subjects will be summarized into two categories based on their score: "None (0)" or "Mild/Moderate/Severe (1-3)," and presented at Baseline, Final visit, and their Worst score.

Graphical assessment of local tolerability will be done on all subjects, not just those with a local tolerability assessment worsened from baseline. Subgroup analysis of local tolerability will not be graphed.

7.5. Laboratory Values

Laboratory values will be presented using the International System of Units (SI units).

For summary purposes, laboratory values that are listed as above or below particular thresholds will be numerically listed as above or below that threshold, respectively, by the minimum measured amount for that parameter. For example, if a parameter is measured to two decimal places, and has a result of “> 5” then, for summary purposes, the value of 5.01 will be used. Values with “<” or “>” will be classified as Low or High, respectively, unless such classifications aren’t applicable for that parameter, in which case they will be classified as Normal.

Shifts in laboratory assessments will be presented from the baseline visit to the last post-baseline visit; subjects missing baseline or post-baseline laboratory data will not be included.

7.6. Vital Signs

Pulse rate, systolic blood pressure, and diastolic blood pressure will be summarized by visit. Shifts in overall vital sign assessments will be presented from the baseline visit to the last post-baseline visit; subjects with no post-baseline vital sign data will not be included. Shifts in normality classification of vital sign parameters based on normal reference ranges will also be presented.