STATISTICAL ANALYSIS PLAN

CRO study code CRO-MT-15-013 - Sponsor code CB-03-01/26

A PHASE 3, MULTICENTER, RANDOMIZED, DOUBLE-BLIND, VEHICLE-CONTROLLED STUDY TO EVALUATE THE SAFETY AND EFFICACY OF CORTEXOLONE 17α-PROPIONATE (CB-03-01) 1% CREAM APPLIED TWICE-DAILY FOR 12 WEEKS IN SUBJECTS WITH FACIAL ACNE VULGARIS

Multicenter, randomized, double-blind, vehicle-controlled, parallel group comparison study

IND number: 112,137

EudraCT number: 2015-002623-26

Test product:	CB-03-01 cream containing 1% cortexolone 17α-propionate (BID)
Placebo product:	Vehicle cream (BID)
Sponsor:	Cassiopea S.p.A. Via C. Colombo 1 20020 Lainate (MI), Italy
Sponsor Representative:	[redacted]
Development phase:	Phase 3
Version and date:	Final version 1.0, 29MAY2018

This study was conducted in accordance with Good Clinical Practice (GCP), ICH topic E6

This document comprises 41 pages plus appendices

STATISTICAL ANALYSIS PLAN APPROVAL

SPONSOR
Cassiopea S.p.A.

Sponsor Representative

[redacted]

CRO

[redacted]

Statistical Analysis

[redacted]

TABLE OF CONTENTS

STUDY SCHEDULE

PROCEDURES	**Visit 1 Screening/ Baseline Day 1**	**Visit 2 Week 4 (Day 29 ± 5)**	**Visit 3 Week 8 (Day 57 ± 7)**	**Visit 4 Week 12[1] (Day 85 ± 10)**
Informed Consent/Assent[2]	X[2]			
Medical/Dermatological History	X			
Demographics	X			
Review of Prior & Concomitant Therapy	X			
Inclusion/Exclusion Criteria	X			
Dermatologic & Limited Physical Exam	X			
Fitzpatrick Skin Type Assessment	X			
Urine Pregnancy Tests[3]	X	X	X	X
Photographs	X	X	X	X
Lesion Counts	X	X	X	X
Investigator's Global Assessment	X	X	X	X
Local Skin Reaction Assessment	X (pre- & post-application)	X	X	X
Electrocardiograms	X			X
Review of Concomitant Medications & Procedures	X	X	X	X
Study Medication Dispensed & Collected[4,5]	X	X	X	X
Review of Compliance		X	X	X
Adverse Events Evaluation	X	X	X	X

1. Subjects who terminate early shall complete all final visit activities designated at Week 12.
2. Consent/assent may be performed up to 45 days prior to the Baseline Visit. Subjects who require "washout" for longer than 45 days were re-consented.
3. Testing to be conducted on all WOCBP. UPT minimal sensitivity is 25 mIU/mL.
4. Instruct subject and the subject's parent/guardian on test article application and provide instruction sheet after assessing the subject's use of the test article with respect to the amount and number of doses applied since the last visit.
5. Dispense subject diary at initial and subsequent visits, as needed.

LIST OF ABBREVIATIONS

ADaM	Analysis Data Model
ADS	Analysis Datasets Specification
AE	Adverse Event
ANCOVA	Analysis of Covariance
ATC	Anatomical Therapeutic Chemical classification system
BID	Twice a day
BOCF	Baseline Observation Carried Forward
CB-03-01	Cortexolone 17α-propionate
CDISC	Clinical Data Interchange Standards Consortium
CRF	Case Report Form
CV%	Percent Coefficient of Variation
ECG	Electrocardiogram
EDC	Electronic Data Capture
FDA	Food and Drug Administration
GCP	Good Clinical Practices
HPA	Hypothalamic-Pituitary-Adrenal
ICH	International Conference of Harmonization
IGA	Investigator's Global Assessment
ILC	Inflammatory Lesions Count
IND	Investigational New Drug
ITT	Intent-to-Treat
IVRS	Interactive Voice Response System
IWRS	Interactive Web Response System
LOCF	Last Observation Carried Forward
LSR	Local Skin Reaction
MAR	Missing at Random
MWV	Missing as Worst Value
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
NILC	Non-Inflammatory Lesions Count
OR	Odds Ratio
PP	Per-Protocol
PT	Preferred Term
PTAE	Pre-Treatment Adverse Event
$\pi_{2p(0,1)}$	Proportion of subjects achieving at least a two-point reduction in IGA compared to Baseline AND an IGA score of 0 (clear) or 1 (almost clear) at Week 12
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis Software
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TLC	Total Lesions Count
UPT	Urine Pregnancy Test
WC	Worst Case
WOCBP	Women of Childbearing Potential

1. INTRODUCTION

Statistical analysis will be performed by the [redacted]. The present Statistical Analysis Plan (SAP) specifies the analyses of the data from clinical study protocol CB-03-01/26. The endpoints and methods of analysis specified in this SAP are consistent with ICH E9 guidelines (1), with the FDA draft guidance for industry on Ance Vulgaris (2) and with the study protocol (4). The SAP has been compiled by the [redacted], reviewed by the Sponsor and finalized before the EDC database lock.

1.1 Changes with respect to the study protocol

No change with respect to the study protocol.

2. STUDY DESCRIPTION

2.1 Background

CB-03-01 (cortexolone 17α-propionate) is a steroidal antiandrogen that is being developed as cream for the topical treatment of acne vulgaris, an androgen-dependent skin disorder. CB-03-01 binds to the human androgen receptor displacing the androgenic hormones and acts as an androgen antagonist. In human plasma, CB-03-01 is rapidly metabolized to parent cortexolone so that its systemic bioavailability after topical application is anticipated to be low. Cortexolone, the main by-product, is a physiological component of the pool of endogenous corticosteroids, and is an intermediate in the synthesis of glucocorticoids, but it only exhibits weak glucocorticoid properties (5).

In single-dose studies in mouse and in rat, and following repeated subcutaneous administration in rats, and dermal administration in rabbits and minipigs, the drug was well tolerated with no significant local or unexpected systemic effects. CB-03-01 is not mutagenic in the Ames test or in the human chromosome aberration test, and it is not a skin sensitizer.

To date, four Phase 1 and three Phase 2 clinical studies have been completed with CB-03-01 cream in Europe and the United States, with over 500 subjects exposed to CB-03-01 creams. All studies were conducted using CB-03-01 cream, 1%; additional concentrations of CB-03-01 cream (0.1% and 0.5%) were also used in the most-recent Phase 2 Dose-Escalating study.

The studies included:

1. A Phase 1 Single Dose Pharmacokinetics (PK) Study in 24 healthy volunteers [randomized, double-blind, vehicle-controlled, single-ascending dose],
2. A Phase 1 Repeat Dose PK Study in 24 healthy volunteers [randomized, vehicle-controlled, 14-day repeat-dose],
3. A Phase 1 Steady State PK Study in 8 subjects with acne vulgaris [single center, open-label, pilot, 6-week repeat-dose]
4. A Phase 1 Three-week Cumulative Skin Irritation Study in 36 healthy volunteers,
5. A Phase 2 Hypothalamic-Pituitary-Adrenal axis (HPA)/PK study in 42 subjects with acne vulgaris [open-label, 14-day repeat-dose],

6. A Phase 2 Comparator Proof of Concept Study in 72 adult subjects with facial acne vulgaris [randomized, double-blind, vehicle-controlled, active-comparator (Retin-A® [tretinoin] 0.05% cream), 8-week repeat-dose], and
7. A Phase 2 Dose-Escalating Study in 363 subjects with facial acne vulgaris [multi-center, randomized, double-blind, vehicle controlled, 12-week repeat-dose].

Overall, the data that resulted from the completion of the aforementioned studies has demonstrated that CB-03-01 creams selected for clinical development

1. are well tolerated locally,
2. result in minimal systemic exposure when applied topically,
3. have a clinical safety profile without any concern and similar to the placebo, and
4. at 1% concentration applied BID are significantly more effective than its vehicle at reducing inflammatory and non-inflammatory lesions counts in adolescent and adult subjects.

CB-03-01 could offer advantages over existing products for acne vulgaris due to its potential efficacy and lack of systemic antiandrogen activity. Because its pharmacological action appears to be limited to the local site of application, use of CB-03-01 may be able to mitigate the side effects related to systemic exposure to hormonal agents currently used in the treatment of acne.

2.2 Rationale

The concentration of CB-03-01 selected for the initial clinical studies (1%) was based on antiandrogenic activity in the hamster flank organ test in which the best effective concentration was 0.8% (5). The results of the Phase 2 Dose Escalation Study demonstrated that CB-03-01 cream, 1% applied BID had more favorable results than other treatments (0.1% BID, 0.5% BID, and 1% QD) and no material safety issues were noted with any of the concentrations of CB-03-01 cream (6). Based on the collective findings of the previous Phase 1 and Phase 2 studies, CB-03-01 cream, 1% BID appears to be the best treatment candidate to advance for further clinical study and development as a topical treatment for acne vulgaris. Thus, the present Phase 3 study has been designed to determine and compare the efficacy of treatment with CB-03-01 cream, 1% and the vehicle cream applied twice daily for 12 weeks in subjects with moderate to severe facial acne vulgaris.

2.3 Study Objectives

The primary objective of this study is to determine the safety and efficacy of CB-03-01 cream, 1% versus the vehicle cream applied twice daily for 12 weeks in subjects with facial acne vulgaris.

2.4 Study Design

This is a multicenter, randomized, double-blind, vehicle-controlled, parallel-group comparison study of safety and efficacy for CB-03-01 cream, 1% in male and female subjects 9 years or older with moderate to severe acne vulgaris on the face. Eligible subjects must have acne vulgaris of the face (which can include the nose) with an Investigator's Global Assessment (IGA) score of 3 or 4, at least 30 to a maximum of 75 inflammatory lesions

(papules, pustules, and nodules) and at least 30 to a maximum of 100 non-inflammatory lesions (open and closed comedones). 732 subjects were enrolled at 48 sites (5 sites included into the study did not enroll any subject). Subjects were randomized in a 1:1 ratio to one of two treatment groups:

1. CB-03-01 cream containing 1% cortexolone 17α-propionate
2. Vehicle cream

All subjects had to apply the assigned test article to the face twice daily for 12 weeks. Refer to section 6 of the study protocol (4).

2.5 Study Population

Male and female subjects 9 years or older with facial acne vulgaris (Grade 3 or 4 on Investigator's Global Assessment [IGA]).

732 subjects were included into the study and 53 sites participated in this study (5 sites did not enroll any subject).

2.5.1 Subject Eligibility

To be included in the study, subjects had to meet the inclusion criteria and none of the exclusion criteria reported in the section 5.1.1 and 5.1.2 of the study protocol (4).

2.5.2 Subject Withdrawal Criteria

Procedures for handling subjects who were discontinued from the study are described in section 13.2 of the study protocol (4). Subjects who were discontinued were not replaced.

2.6 Randomization Assignment

Subjects who were eligible for enrollment into the study were randomized to receive CB-03-01 cream or vehicle cream in a 1:1 ratio. The randomization scheme was blocked by investigational site with a block size of 4. At each site, subject kits were dispensed according to the kit number assigned by an IVRS/IWRS system as subjects were enrolled. At each site, the study staff added the site number (provided to each site) and the subject number (starting with 001) to each kit label. The kit number dispensed to each subject was also recorded on the source documents and CRFs. Treatment group designation will remain blinded until the final database is locked, unless unblinding is required as described in section 15 of the study protocol (4). Subjects who withdrew prematurely from the study were not replaced.

3. CLINICAL EVALUATIONS AND LABORATORY TESTS

The following clinical evaluations were performed according to the schedules indicated during the study. The same investigator should have completed the evaluations for a given subject throughout the study. If this became impossible a sub-investigator with overlapping experience with the subject and the study should have completed the evaluations.

3.1 Investigator's Global Assessment (IGA)

At every study visit, overall severity of acne using a five-point scale from 0 (clear) to 4 (severe) was assessed. This is a static morphological scale that refers to a point in time and not a comparison to Baseline. Subjects had to have an IGA score of 3 (moderate) or 4 (severe) at Baseline.

Score	Assessment	Description
0	Clear	Absence of active disease with no inflammatory or non-inflammatory lesions
1	Almost Clear	Rare non-inflammatory lesions with no more than one small inflammatory lesions
2	Mild	Some non-inflammatory lesions with no more than a few inflammatory lesions (papules/pustules only; no nodular/cystic lesions)
3	Moderate	Up to many non-inflammatory lesions and may have some inflammatory lesions but no more than one nodular/cystic lesion
4	Severe	Up to many non-inflammatory lesions and inflammatory lesions but no more than a few nodular/cystic lesions

3.2 Acne Lesions Counting

The number of inflammatory lesions (papules, pustules, and nodules) and non-inflammatory lesions (open and closed comedones) on the face vertically from the hairline to mandible rim and horizontally from ear to ear, including those on the nose were counted to determine subject eligibility at Visit 1. Subjects must have at least 30 to a maximum of 75 inflammatory lesions and 30 to a maximum of 100 non-inflammatory lesions on the face (which can include the nose) at Baseline.

Lesions are defined as follows:

1. Comedones - open (blackheads) and closed (whiteheads).
2. Papules - raised inflammatory lesions with no visible purulent material.
3. Pustules - raised inflammatory lesions with visible purulent material.
4. Nodules - Any circumscribed, inflammatory masses greater or equal to 5 mm in diameter with or without cystic changes.

3.3 Fitzpatrick Skin Type Assessment

At Visit 1, the investigator or designee had to document the subject's skin phototype (I-VI) using the Fitzpatrick Skin Type Assessment as described below:

Skin Phototype	Typical Features	Tanning ability
I	Pale white skin, blue/hazel eyes, blond/red hair	Always burns, does not tan
II	Fair skin, blue eyes	Burns easily, tans poorly
III	Darker white skin	Tans after initial burn
IV	Light brown skin	Burns minimally, tans easily
V	Brown skin	Rarely burns, tans darkly easily
VI	Dark brown or black skin	Never burns, always tans darkly

3.4 Local Skin Reactions (LSRs)

At every study visit, the investigator or designee had to document the severity of the following LSRs known to be associated with application of topical steroids:

1. Telangiectasia
2. Skin atrophy
3. Striae rubrae

A five-point ordinal scale was used to assess the severity of these reactions:

Score	Assessment
0	None
1	Trace
2	Mild
3	Moderate
4	Severe

In addition, the investigator or designee had to evaluate the severity of the following LSRs known to be associated with acne vulgaris, using the five-point ordinal scales described below:

4. Erythema
5. Edema
6. Scaling/dryness

Erythema Score	**Assessment**
0	None
1	Minimal - barely perceptible erythema
2	Mild - predominantly minimal erythema (pink) in the treated area with or without a few isolated areas of more intense erythema
3	Moderate - predominantly moderate erythema (red) in the treated area with or without a few isolated areas of intense erythema (bright red)
4	Severe - predominantly intense erythema (bright red) in the treated area with or without a few isolated areas of very intense (fiery red) erythema

Edema Score	**Assessment**
0	None
1	Minimal - scant, rare edema
2	Mild - easily seen edema, minimally palpable, involving up to 1/3 of the Treatment Area
3	Moderate - easily seen edema and typically palpable, involving between 1/3 to 2/3 of the Treatment Area
4	Severe - easily seen edema, indurated in some areas, involving over 2/3 of the Treatment Area

Scaling/dryness Score	**Assessment**
0	None
1	Minimal - barely perceptible desquamation
2	Mild - limited areas of fine desquamation in up to 1/3 of the treatment area
3	Moderate - fine desquamation involving 1/3 to 2/3 of the treatment area or limited areas of coarser scaling
4	Severe - coarser scaling involving more than 2/3 of the treatment area or limited areas of very coarse scaling

In addition, subjects were asked to rate the severity of the following LSRS that occurred in the treatment area since the last visit, using the four-point ordinal scales described below:

7. Stinging/burning
8. Pruritus

Stinging/burning Score	**Assessment**
0	None
1	Minimal, barely perceptible - tolerable and little discomfort
2	Moderate - tolerable, but causes some discomfort
3	Severe - very uncomfortable or intolerable

Pruritus (itching) Score	**Assessment**
0	None - no evidence of itching
1	Mild - only aware of itching at times, only present when relaxing, not present when focused on other activities
2	Moderate - often aware of itching, annoying, sometimes disturbs sleep and daytime activities
3	Severe - constant itching, distressing; frequent sleep disturbance, interferes with activities

These LSRs were collected independently of AEs. Only LSRs that require medical intervention (e.g., prescription medication) or require withholding the application of the test articles had to be documented as AEs. Any LSRs that are not listed above had to be recorded as AEs.

3.5 Urine Pregnancy Tests (UPTs)

The UPT had to be performed at the study site if the site is registered and conforms to Clinical Laboratory Improvement Amendments (CLIA) regulations for such testing (site possesses at a minimum a current valid CLIA Certificate of Waiver) or at an appropriately registered reference laboratory. A UPT had to be performed on all WOCBP at Visit 1 (Baseline), Visit 2, Visit 3 and Visit 4 (or end of treatment if the subject withdraws prematurely). The investigator had to report the UPT results on the CRFs, in the subject's medical records and, if applicable, in independent records maintained at the study site. The UPT used must have a minimum sensitivity of 25mIU of β-HCG/mL of urine.

4. STATISTICAL CONSIDERATIONS

4.1 Study Endpoints

4.1.1 Efficacy Endpoints

4.1.1.1 Primary Efficacy Endpoints

P1: Proportion of subjects in each treatment group with at least a two-point reduction in IGA compared to Baseline AND an IGA score of 0 (clear) or 1 (almost clear) at Week 12 ($\pi_{2p(0,1)}$);
P2: Absolute change from Baseline in non-inflammatory lesions count (NILC) in each treatment group at Week 12;
P3: Absolute change from Baseline in inflammatory lesions count (ILC) in each treatment group at Week 12.

4.1.1.2 Secondary Efficacy Endpoints

S1: Absolute change from Baseline in total lesions count (TLC, including inflammatory and non-inflammatory lesions) in each treatment group at Week 12;
S2: Percent change from Baseline in total lesions count (TLC, including inflammatory and non-inflammatory lesions) in each treatment group at Week 12;
S3: Percent change from Baseline in non-inflammatory lesions count (NILC) in each treatment group at Week 12;
S4: Percent change from Baseline in inflammatory lesions count (ILC) in each treatment group at Week 12.

4.1.2 Safety Endpoints

SF1: Local and systemic AEs at every visit (Baseline, Weeks 4, 8, and 12);
SF2: Local Skin Reactions (telangiectasia, skin atrophy, striae rubrae, erythema, edema, scaling/dryness, stinging/burning and pruritus) scored by frequency and severity at every visit (Baseline, Weeks 4, 8 and 12);
SF3: UPTs in all WOCBP at Visit 1 (Baseline), Visit 2, Visit 3 and Visit 4 (or end of treatment if the subject withdraws prematurely);
SF4: Material changes from Baseline in ECGs at Week 12.

4.2 Hypothesis Tests

Hypothesis tests on the primary endpoints will be conducted according to the following hierarchical order:

HT1: Superiority of TEST vs. VEHICLE in terms of the proportion of subjects in each treatment group with at least a two-point reduction in IGA compared to Baseline AND an IGA score of 0 (clear) or 1 (almost clear) at Week 12 ($\pi_{2p(0,1)}$);
HT2: Superiority of TEST vs. VEHICLE in terms of absolute change from Baseline in non-inflammatory lesions count (NILC) at Week 12;
HT3: Superiority of TEST vs. VEHICLE in terms of absolute change from Baseline in inflammatory lesions count (ILC) at Week 12.

The null hypotheses of the three hypothesis tests on the primary endpoints have all to be rejected in order to claim superiority of of TEST vs. VEHICLE.

In order to keep an overall Type I error α=0.05 two-sided for the whole study, all the hypothesis tests have to be performed at the same significance level α=0.05 two-sided and the failure to reject a null hypothesis will imply the failure of rejecting of all the subsequent null hypotheses.

This approach inflates the overall Type II error β, which in the worst case scenario is equal to the sum of the Type II error β connected with the individual hypothesis tests. This inflation must be taken into account for a proper estimation of the sample size (3).

Hypothesis tests on the secondary endpoints will be conducted according to the following hierarchical order:

HT4: Superiority of TEST vs. VEHICLE in terms of absolute change from Baseline in total lesions count (TLC) at Week 12;
HT5: Superiority of TEST vs. VEHICLE in terms of percent change from Baseline in total lesions count (TLC) at Week 12;
HT6: Superiority of TEST vs. VEHICLE in terms of percent change from Baseline in non-inflammatory lesions count (NILC) at Week 12;
HT7: Superiority of TEST vs. VEHICLE in terms of percent change from Baseline in inflammatory lesions count (ILC) at Week 12.

The hypothesis test for the secondary endpoints will be performed at the same significance level α=0.05 two-sided of the primary endpoints if and only if all the null hypotheses on the primary endpoints are rejected and the failure to reject a null hypothesis will imply the failure of rejecting of all the subsequent null hypotheses.

4.2.1 Superiority of TEST vs. VEHICLE in terms of the proportion of subjects with at least a two-point reduction in IGA compared to Baseline AND an IGA score of 0 (clear) or 1 (almost clear) at Week 12

$\boldsymbol{\pi_{2p(0,1),TEST}}$: Proportion of subjects in TEST treatment group achieving "success" at Week 12, with "success" defined as at least a two-point reduction in IGA compared to Baseline AND an IGA score of 0 (clear) or 1 (almost clear)

$\boldsymbol{\pi_{2p(0,1),VEHICLE}}$: Proportion of subjects in VEHICLE treatment group achieving "success" at Week 12, with "success" defined as at least a two-point reduction in IGA compared to Baseline AND an IGA score of 0 (clear) or 1 (almost clear)

$$\mathbf{OR_{2p(0,1),TV}} = \frac{\boldsymbol{\pi_{2p(0,1),TEST}}\left(\mathbf{1} - \boldsymbol{\pi_{2p(0,1),VEHICLE}}\right)}{\boldsymbol{\pi_{2p(0,1),VEHICLE}}\left(\mathbf{1} - \boldsymbol{\pi_{2p(0,1),TEST}}\right)}$$

$$\mathbf{HT1}: \begin{cases} \mathbf{H_0}: \mathbf{OR_{2p(0,1),TV}} \leq \mathbf{1} \\ \mathbf{H_a}: \mathbf{OR_{2p(0,1),TV}} > \mathbf{1} \end{cases}$$

4.2.2 Superiority of TEST vs. VEHICLE in terms of absolute change from Baseline in NILC at Week 12

$\mu_{NILC,TEST}$: Mean absolute change from baseline of NILC at Week 12 in TEST treatment group

$\mu_{NILC,VEHICLE}$: Mean absolute change from baseline of NILC at Week 12 in VEHICLE treatment group

$$\mathbf{HT2}: \begin{cases} \mathbf{H_0}: \mu_{NILC,TEST} - \mu_{NILC,VEHICLE} \leq \mathbf{0} \\ \mathbf{H_a}: \mu_{NILC,TEST} - \mu_{NILC,VEHICLE} > \mathbf{0} \end{cases}$$

4.2.3 Superiority of TEST vs. VEHICLE in terms of absolute change from Baseline in ILC at Week 12

$\mu_{ILC,TEST}$: Mean absolute change from baseline of ILC at Week 12 in TEST treatment group

$\mu_{ILC,VEHICLE}$: Mean absolute change from baseline of ILC at Week 12 in VEHICLE treatment group

$$\mathbf{HT3}: \begin{cases} \mathbf{H_0}: \mu_{ILC,TEST} - \mu_{ILC,VEHICLE} \leq \mathbf{0} \\ \mathbf{H_a}: \mu_{ILC,TEST} - \mu_{ILC,VEHICLE} > \mathbf{0} \end{cases}$$

4.2.4 Superiority of TEST vs. VEHICLE in terms of absolute change from Baseline in TLC at Week 12

$\mu_{TLC,TEST}$: Mean absolute change from baseline of TLC at Week 12 in TEST treatment group

$\mu_{TLC,VEHICLE}$: Mean absolute change from baseline of TLC at Week 12 in VEHICLE treatment group

$$\mathbf{HT4}: \begin{cases} \mathbf{H_0}: \mu_{TLC,TEST} - \mu_{TLC,VEHICLE} \leq \mathbf{0} \\ \mathbf{H_a}: \mu_{TLC,TEST} - \mu_{TLC,VEHICLE} > \mathbf{0} \end{cases}$$

4.2.5 Superiority of TEST vs. VEHICLE in terms of percent change from Baseline in TLC at Week 12

$\mathbf{P}_{TLC,TEST}$: Mean percent change from baseline of TLC at Week 12 in TEST treatment group

$\mathbf{P}_{TLC,VEHICLE}$: Mean percent change from baseline of TLC at Week 12 in VEHICLE treatment group

$$\mathbf{HT5}: \begin{cases} \mathbf{H_0}: \mathbf{P}_{TLC,TEST} - \mathbf{P}_{TLC,VEHICLE} \leq \mathbf{0} \\ \mathbf{H_a}: \mathbf{P}_{TLC,TEST} - \mathbf{P}_{TLC,VEHICLE} > \mathbf{0} \end{cases}$$

4.2.6 Superiority of TEST vs. VEHICLE in terms of percent change from Baseline in NILC at Week 12

$\mathbf{P_{NILC,TEST}}$: Mean percent change from baseline of NILC at Week 12 in TEST treatment group

$\mathbf{P_{NILC,VEHICLE}}$: Mean percent change from baseline of NILC at Week 12 in VEHICLE treatment group

$$\mathbf{HT6}: \begin{cases} \mathbf{H_0}: \mathbf{P_{NILC,TEST}} - \mathbf{P_{NILC,VEHICLE}} \leq \mathbf{0} \\ \mathbf{H_a}: \mathbf{P_{NILC,TEST}} - \mathbf{P_{NILC,VEHICLE}} > \mathbf{0} \end{cases}$$

4.2.7 Superiority of TEST vs. VEHICLE in terms of percent change from Baseline in ILC at Week 12

$\mathbf{P_{TLC,TEST}}$: Mean percent change from baseline of ILC at Week 12 in TEST treatment group

$\mathbf{P_{TLC,VEHICLE}}$: Mean percent change from baseline of ILC at Week 12 in VEHICLE treatment group

$$\mathbf{HT7}: \begin{cases} \mathbf{H_0}: \mathbf{P_{ILC,TEST}} - \mathbf{P_{ILC,VEHICLE}} \leq \mathbf{0} \\ \mathbf{H_a}: \mathbf{P_{ILC,TEST}} - \mathbf{P_{ILC,VEHICLE}} > \mathbf{0} \end{cases}$$

4.3 Sample Size Calculations and Power Considerations

Assumptions used for sample size calculation were estimated from Phase 2 Dose Escalation Study 171-7151-201 (6).

Sample size calculations were performed on the basis of the formulas for testing superiority in term of means difference and odds ratio of Wang and Chow (7, 8).

All the hypothesis tests for the primary endpoints have to be performed at the same significance level α=0.05 two-sided and the failure to reject a null hypothesis will imply the failure of rejecting of all the subsequent null hypotheses.

In order to keep the overall Type II error β=0.1 for the whole study, the sample sizes for the three hypothesis tests on the primary endpoints were calculated considering an individual Type II error β=0.033 (3).

The hypothesis tests for the secondary endpoint will be performed at the same significance level α=0.05 two-sided of the primary endpoints if and only if all the null hypotheses on the primary endpoints are rejected and the failure to reject a null hypothesis will imply the failure of rejecting of all the subsequent null hypotheses. The sample sizes for the four hypothesis tests on the secondary endpoints were calculated considering an individual Type II error β=0.1.

Based on the previous assumptions, at least 350 subjects in each treatment group (CB-03-01 1% BID and vehicle BID) were included into each Phase 3 study to provide sufficient power (90%) with the chosen primary endpoints as this approach would:

1. allow a reasonable allowance for non-completers and to provide a more meaningful number of subjects exposed to CB-03-01 cream in the safety population of the program and
2. be adequate for investigating with a sufficient power the secondary endpoint.

Details of the calculations are reported in the following sections.

4.3.1 Proportion of subjects in each treatment group with at least a two-point reduction in IGA compared to Baseline AND an IGA score of 0 (clear) or 1 (almost clear) at Week 12

$$\mathbf{N_{TEST} = N_{VEHICLE}} = \frac{\left(\mathbf{z_{\alpha/2} + z_{\beta}}\right)^2}{\left(\mathbf{\log\left(OR_{2p(0,1)}\right)}\right)^2}\left(\frac{\mathbf{1}}{\mathbf{\pi_{2p(0,1),TEST}\left(1 - \pi_{2p(0,1),TEST}\right)}} + \frac{\mathbf{1}}{\mathbf{\pi_{2p(0,1)VEHICLE}\left(1 - \pi_{2p(0,1),VEHICLE}\right)}}\right)$$

A post-hoc analysis of the data collected in the Phase 2 Dose Escalation Study 171-7151-201 (6) performed only for the subjects with 30-75 inflammatory lesions and 30-100 non-inflammatory lesions at baseline provided the following results for the proportion of success and the related sample size estimation.

	TEST	**VEHICLE**
Success	3 (12.50%)	0 (0.00%)
Failure	21 (87.50%)	22 (100.00%)

Since the number of success is zero for VEHICLE, a continuity correction (Haldane correction) was applied to each cell in the previous table in order to calculate the odds ratio:

	TEST	**VEHICLE**
Success	3.5 (14.00%)	0.5 (2.17%)
Failure	21.5 (86.00%)	22.5 (97.83%)

$$\mathbf{OR_{2p(0,1),TV}} = \frac{\mathbf{\pi_{2p(0,1),TEST}\left(1 - \pi_{2p(0,1),VEHICLE}\right)}}{\mathbf{\pi_{2p(0,1),VEHICLE}\left(1 - \pi_{2p(0,1),TEST}\right)}} = \mathbf{7.3256}$$

$$\boldsymbol{\alpha = 0.05\ two - sided, \beta = 0.033 \Rightarrow N_{TEST} = N_{VEHICLE} = 202}$$

202 subjects per group should be enrolled in order to have 202 subjects per group included into the ITT Set.

4.3.2 Absolute change from baseline in NILC in each treatment group at Week 12

$$\mathbf{N_{TEST} = N_{VEHICLE}} = \frac{\mathbf{2\sigma^2\left(z_{\alpha/2} + z_{\beta}\right)^2}}{\left(\mu_{NILC,TEST} - \mu_{NILC,VEHICLE}\right)^2}$$

A post-hoc analysis of the data collected in the Phase 2 Dose Escalation Study 171-7151-201 (6) performed only for the subjects with 30-75 inflammatory lesions and 30-100 non-inflammatory lesions at baseline provided the following results for the least square means (i.e. the means obtained after having controlled for the fixed effects and the covariate of the ANCOVA) of the absolute change from baseline in non-inflammatory lesions count and for the related sample size estimation.

	Number of subjects	**Least Square Mean**
TEST	24	-22.398746
VEHICLE	22	-0.924803

Considering a mean absolute change from baseline in non-inflammatory lesions count of **-22** for TEST treatment group and **-1** for VEHICLE treatment group and a common standard deviation of **17**, we obtain the following estimate of the sample size.

$$\boldsymbol{\alpha = 0.05\ two - sided, \beta = 0.033, \mu_{ILC,TEST} = -22, \mu_{ILC,VEHICLE} = -1, \sigma = 17}$$

$$\boldsymbol{N_{TEST} = N_{VEHICLE} = 19}$$

19 subjects per group should be enrolled in order to have 19 subjects per group included into the ITT Set.

4.3.3 Absolute change from Baseline in ILC in each treatment group at Week 12

$$\boldsymbol{N_{TEST} = N_{VEHICLE} = \frac{2\sigma^2\left(z_{\alpha/2} + z_{\beta}\right)^2}{\left(\mu_{ILC,TEST} - \mu_{ILC,VEHICLE}\right)^2}}$$

A post-hoc analysis of the data collected in the Phase 2 Dose Escalation Study 171-7151-201 (6) performed only for the subjects with 30-75 inflammatory lesions and 30-100 non-inflammatory lesions at baseline provided the following results for the least square means (i.e. the means obtained after having controlled for the fixed effects and the covariate of the ANCOVA) of the absolute change from baseline in inflammatory lesions count and for the related sample size estimation.

	Number of subjects	**Least Square Mean**
TEST	24	-14.473098
VEHICLE	22	-10.621723

Considering a mean absolute change from baseline in inflammatory lesions count of **-15** for TEST treatment group and **-10** for VEHICLE treatment group and a common standard deviation of **17**, we obtain the following estimate of the sample size.

$$\boldsymbol{\alpha = 0.05\ two - sided, \beta = 0.033, \mu_{ILC,TEST} = -15, \mu_{ILC,VEHICLE} = -10, \sigma = 17}$$

$$\boldsymbol{N_{TEST} = N_{VEHICLE} = 334}$$

334 subjects per group should be enrolled in order to have 334 subjects per group included into the ITT Set.

4.3.4 Absolute change from Baseline in TLC at Week 12

$$\mathbf{N_{TEST}} = \mathbf{N_{VEHICLE}} = \frac{\mathbf{2\sigma^2(z_{\alpha/2} + z_{\beta})^2}}{(\mu_{\mathbf{TLC,TEST}} - \mu_{\mathbf{TLC,VEHICLE}})^2}$$

A post-hoc analysis of the data collected in the Phase 2 Dose Escalation Study 171-7151-201 (6) performed only for the subjects with 30-75 inflammatory lesions and 30-100 non-inflammatory lesions at baseline provided the following results for the least square means (i.e. the means obtained after having controlled for the fixed effects and the covariate of the ANCOVA) of the absolute change from baseline in total lesions count and for the related sample size estimation.

	Number of subjects	**Least Square Mean**
TEST	24	-35.7127562
VEHICLE	22	-13.0667959

Considering a mean absolute change from baseline in total lesions count of **-35** for TEST treatment group and **-13** for VEHICLE treatment group and a common standard deviation of **17**, we obtain the following estimate of the sample size.

$$\boldsymbol{\alpha = 0.05\ two - sided, \beta = 0.1, \mu_{TLC,TEST} = -35, \mu_{TLC,VEHICLE} = -13, \sigma = 17}$$

$$\mathbf{N_{TEST} = N_{VEHICLE} = 13}$$

13 subjects per group should be enrolled in order to have 13 subjects per group included into the ITT Set.

4.3.5 Percent change from Baseline in TLC at Week 12

$$\mathbf{N_{TEST}} = \mathbf{N_{VEHICLE}} = \frac{\mathbf{2\sigma^2(z_{\alpha/2} + z_{\beta})^2}}{(\mathbf{P_{TLC,TEST}} - \mathbf{P_{TLC,VEHICLE}})^2}$$

A post-hoc analysis of the data collected in the Phase 2 Dose Escalation Study 171-7151-201 (6) performed only for the subjects with 30-75 inflammatory lesions and 30-100 non-inflammatory lesions at baseline provided the following results for the least square means (i.e. the means obtained after having controlled for the fixed effects and the covariate of the ANCOVA) of the percent change from baseline in total lesions count and for the related sample size estimation.

	Number of subjects	**Least Square Mean**
TEST	24	-35.455503
VEHICLE	22	-14.144333

Considering a mean percent change from baseline in total lesions count of **-35** for TEST treatment group and **-14** for VEHICLE treatment group and a common standard deviation of **40**, we obtain the following estimate of the sample size.

$$\boldsymbol{\alpha = 0.05 \text{ two} - \text{sided}, \beta = 0.1, \mu_{TLC,TEST} = -35, \mu_{TLC,VEHICLE} = -14, \sigma = 40}$$

$$\boldsymbol{N_{TEST} = N_{VEHICLE} = 77}$$

77 subjects per group should be enrolled in order to have 77 subjects per group included into the ITT Set.

4.3.6 Percent change from Baseline in NILC at Week 12

$$\boldsymbol{N_{TEST} = N_{VEHICLE} = \frac{2\sigma^2\left(z_{\alpha/2} + z_{\beta}\right)^2}{\left(P_{NILC,TEST} - P_{NILC,VEHICLE}\right)^2}}$$

A post-hoc analysis of the data collected in the Phase 2 Dose Escalation Study 171-7151-201 (6) performed only for the subjects with 30-75 inflammatory lesions and 30-100 non-inflammatory lesions at baseline provided the following results for the least square means (i.e. the means obtained after having controlled for the fixed effects and the covariate of the ANCOVA) of the percent change from baseline in non-inflammatory lesions count and for the related sample size estimation.

	Number of subjects	**Least Square Mean**
TEST	24	-33.718297
VEHICLE	22	-5.133111

Considering a mean percent change from baseline in non-inflammatory lesions count of **-33** for TEST treatment group and **-5** for VEHICLE treatment group and a common standard deviation of **40**, we obtain the following estimate of the sample size.

$$\boldsymbol{\alpha = 0.05 \text{ two} - \text{sided}, \beta = 0.1, \mu_{TLC,TEST} = -33, \mu_{TLC,VEHICLE} = -5, \sigma = 40}$$

$$\boldsymbol{N_{TEST} = N_{VEHICLE} = 43}$$

43 subjects per group should be enrolled in order to have 43 subjects per group included into the ITT Set.

4.3.7 Percent change from Baseline in ILC at Week 12

$$\boldsymbol{N_{TEST} = N_{VEHICLE} = \frac{2\sigma^2\left(z_{\alpha/2} + z_{\beta}\right)^2}{\left(P_{ILC,TEST} - P_{ILC,VEHICLE}\right)^2}}$$

A post-hoc analysis of the data collected in the Phase 2 Dose Escalation Study 171-7151-201 (6) performed only for the subjects with 30-75 inflammatory lesions and 30-100 inflammatory lesions at baseline provided the following results for the least square means (i.e. the means obtained after having controlled for the fixed effects and the covariate of the ANCOVA) of

the percent change from baseline in inflammatory lesions count and for the related sample size estimation.

	Number of subjects	**Least Square Mean**
TEST	24	-37.657976
VEHICLE	22	-23.450604

Considering a mean percent change from baseline in inflammatory lesions count of **-37** for TEST treatment group and **-23** for VEHICLE treatment group and a common standard deviation of **40**, we obtain the following estimate of the sample size.

$$\boldsymbol{\alpha = 0.05\ two - sided, \beta = 0.1, \mu_{TLC,TEST} = -37, \mu_{TLC,VEHICLE} = -23, \sigma = 40}$$

$$\boldsymbol{N_{TEST} = N_{VEHICLE} = 172}$$

172 subjects per group should be enrolled in order to have 172 subjects per group included into the ITT Set.

5. BLINDING/UNBLINDING AND BLIND REVIEW

Blinding is important for the integrity of this clinical drug trial. However, the blind may be broken in the event of a medical emergency in a subject, in which knowledge of the test article identity is critical to the subject's management. Before breaking the blind for a subject, the investigator should determine that the information is necessary (i.e., that it alters the subject's immediate management). In many cases, particularly when the emergency is clearly not test article related, the problem may be effectively managed by assuming that the subject is receiving active product without the need for unblinding. The need to break the blind had first to be discussed with the responsible Medical Monitor and the best method to do this was be determined.

Data collected during the study will be revised in blind, the assignment of subjects to the analysis sets (ITT, PP and Safety) and the definition of the analysis centers (see section 6.5) will be performed before the database lock during the blind review meeting/conference.

The review team will be composed only of people not directly involved in the clinical phase of the trial in order to safeguard their blindness to the administered treatments.

6. STATISTICAL METHODS

All statistical analysis will be performed using SAS for Windows Version 9.3 (TS1M1) (9) (or higher).
The data documented in this trial and the parameters measured will be summarised using classic descriptive statistics: mean, standard deviation, percent coefficient of variation (CV%), minimum, median and maximum values for quantitative variables. Frequency counts and percentage of subjects within each category are provided for categorical data. Summaries will be provided for each treatment group. Data not available will be managed according to the procedure described in the paragraph 6.4.

6.1 Tables, Listings and Figures layout

Tables and Listings will be provided according to the following settings:

- Background: White
- Foreground: Black
- Font face: Times
- Font style: Roman
- Font size: 10 pt
- Font weight: Medium (data, footers and notes), Bold (titles and headers)
- Font width: Normal
- Layout: Landscape
- Top Margin: 0.8 cm
- Bottom Margin: 0.8 cm
- Left Margin: 0.8 cm
- Right Margin: 0.8 cm
- Test label: CB-03-01
- Placebo label: Vehicle
- Date format: ddMMMyyyy

- Means, standard deviations, percent coefficient of variations, medians, lower confidence limits and upper confidence limits will be rounded to one digit more than the original data
- Minima and maxima will keep the same number of decimal digits as the source values
- p-values will be rounded to the fourth decimal digit and will be flagged by an asterisk (*) in case of statistical significance (i.e. p-value < 0.05 or, in case of center by treatment interaction, p-value < 0.10). p-values lower than 0.0001 will be reported as "<.0001 *".

The data and results of CB-03-01 will be presented before the data and results of Vehicle in all listings and tables.

6.2 Analysis sets

The following analysis sets will be defined:

1. Intent-to-Treat Set (ITT): all randomised subjects. This analysis set will be used for primary analyses and sensitivity analyses.

2. Per Protocol Set (PP): all randomised subjects who will fulfil the study protocol requirements in terms of investigational medicinal product intake and collection of primary efficacy data, will not have any inclusion/exclusion criteria violation and with no major deviations that may affect study results. This analysis set will be used for sensitivity analyses.

3. Safety Set: all subjects who will receive at least one dose of the test article. This analysis set will be used for safety analyses.

6.2.1 Exclusion from the Per Protocol Set

Reasons for the exclusion from the Per Protocol Set include the following:

1. lack of compliance to the test article (subjects with an overall compliance not evaluable due to missing data or less than 80% as defined in paragraph 6.3);
2. exposure to an test article different from the one assigned to the subject;
3. lack of the primary variables assessments at Baseline or at Week 12;
4. failure to satisfy any inclusion/exclusion criteria (eligibility violations);
5. intake of prohibited medications.

The precise reasons for excluding subjects from the per protocol set will be fully defined and documented in the blind review meeting/conference report before breaking the blind.

6.3 Evaluation of Treatment Compliance

Compliance to the test article will be evaluated at each visit and overall according to the following formula:

$$\textbf{Compliance} = \textbf{100} \times \frac{\textbf{Number of actual applications}}{\textbf{Number of scheduled applications}}$$

Non-compliance will be defined as a compliance value less than 80% and subjects with an overall compliance not evaluable due to missing data or less than 80% will be excluded from the Per Protocol Set (see paragraph 6.2.1).

6.4 Handling of missing data

6.4.1 Methods for replacing missing data

1. Multiple imputation (MI) under missing at random (MAR) assumption: multiple imputation is a general approach to the problem of missing data that aims to allow for the uncertainty about the missing data by creating several different plausible imputed data sets and appropriately combining results obtained from each of them. Missing at random (MAR) approach assumes that any systematic difference between the missing values and the observed values can be explained by differences in observed data. The use of multiple imputation under MAR assumption provides unbiased estimations of the parameters and allows to evaluate the uncertainty of parameters' estimation due to the presence of missing data.

   A Markov Chain Monte Carlo (MCMC) method that assumes multivariate normality will be used to impute all the missing values. In order to prevent a power falloff due to choosing a number of imputations too small, 10 imputations will be performed (10).

   The multiple imputation procedure will be implemented by using the MI procedure of the SAS system with the following options:

   - `NIMPUTE=10`
   - `SEED=1`
   - `ROUND=0`
   - `MINMAXITER=10000000`

   and the following MCMC statement:

   - `MCMC IMPUTE=FULL CHAIN=SINGLE INITIAL=EM (BOOTSTRAP=0.9 CONVERGE=1E-6 ITPRINT MAXITER=1000)`

   A unique imputation process will be performed without taking into account the treatment arm.

   The imputed data sets will be then analysed with the MIANALYZE procedure of the SAS system with the following options:

   - `ALPHA=0.05`
   - `EDF=number of subjects of the ITT set - number of estimated parameters (including intercept)]`

Type-III analyses obtained from multiple imputations will be combined according to the method proposed by Binhuan Wang and Yixin Fang (11).

2. Missing as Worst Value (MWV): for each subject, missing values of a variable will be replaced with the worst value (failure in case of dichotomous variables or the higher value collected in the study for lesions counting) of that variable regardless of the treatment arm.

3. Worst Case (WC): for each subject, missing values of a variable will be replaced with the worst value (failure in case of dichotomous variables or the higher value collected in the study for lesions counting) of that variable if the subject is assigned to the CB-03-01 arm and with the best value of that variable (success in case of dichotomous variables or the lower value collected in the study for lesions counting) if the subject is assigned to the Vehicle arm.

4. Last Observation Carried Forward (LOCF): for each subject, missing values of a variable are replaced by the last observed value of that variable.

5. Baseline Observation Carried Forward (BOCF): for each subject, missing values of a variable are replaced by the baseline observed value of that variable.

6.4.2 Replacement rules for each analysis set

1. Intent-to-Treat Set (ITT): all missing values of the primary endpoints will be replaced according to the MI under MAR assumption method for the primary analysis and according to the MWV, WC, LOCF and BOCF methods in order to perform four distinct sensitivity analyses. Missing values of the secondary endpoints will be replaced according to the MI under MAR assumption method.

2. Per Protocol Set (PP): no replacement of missing data is required for the primary and secondary endpoints.

3. Safety Set: no replacement of missing data is required for the safety endpoints.

6.5 Analysis centers

Each center conducted the clinical study under a common protocol. Consistency in study execution at each center was emphasized. The study had to be conducted in such a manner as to have a minimum of 8 ITT subjects enrolled in each of the treatment groups at each center (i.e. 16 ITT subjects enrolled overall). In the event that there are too few subjects in a treatment arm for a center, then this center's data will be combined with another center's data in order to achieve the desired minimum number of subjects per treatment arm. The combining of data across centers will be accomplished by combining data from two or more centers based on geographic and climatic similarities until the minimum number of subjects per treatment arm is achieved. The sequence of combination will be based on the total number of subjects enrolled in each center (combining lowest to 2nd lowest and so on). Combining center data to achieve sufficient subjects per treatment arm will result in the combined centers. These combined centers, as well as individual centers with sufficient subjects per treatment arm, will be referred to as "analysis centers" in the statistical analyses.

Supposing to have N geographic and climatic areas with $M_{i\ (i=1..N)}$ centers for each area, we can write the set of all centers as

$$\{C_{ij}\}_{i=1..N,j=1..M_i}$$

and the set of ordered centers according to the ascending number of enrolled subjects within each geographic and climatic area as

$$\{C_{ij(k)}\}_{i=1..N,k=1..M_i} \mid \forall\, i\,, m < n \Rightarrow \#(C_{ij(m)}) \leq \#(C_{ij(n)})$$

Using this notation, each geographic and climatic area i can have up to A_i ($\forall\ i=1..N,\ A_i \leq M_i$) analysis centers

$$\{AC_{ij}\}_{i=1..N,j=1..A_i}$$

For example

$$AC_{i1} = C_{ij(1)}\ U\ C_{ij(2)}\ U\ C_{ij(3)}$$
$$AC_{i2} = C_{ij(4)}\ U\ C_{ij(5)}$$
$$AC_{i3} = C_{ij(6)}$$
$$\dots$$
$$AC_{iA_i} = C_{ij(M_i)}$$

The definition of the analysis centers will be performed before the database lock during the blind review meeting/conference.

6.6 Demographic, baseline and background characteristics and other study information

Demographic, baseline and background characteristics will be reported for all the enrolled subjects and analyses will be performed according to the treatment the subjects actually received (PP and Safety) and according to the treatment they were assigned (ITT).

6.6.1 Subjects' disposition

The disposition of all subjects enrolled in the study will be listed and summarised by treatment group (CB-03-01 and Vehicle). The number and percentage of subjects completing the study, the number and percentage of withdrawals and the reasons for withdrawal will be presented (Listing 16.2.4.1, Table 14.1.1.1).

6.6.2 Analysis sets

The subjects included in each analysis sets will be listed and summarised by treatment group (CB-03-01 and Vehicle) (Listing 16.2.4.2, Table 14.1.1.2).

6.6.3 Subjects excluded from efficacy and/or safety analysis

All subjects excluded from the efficacy and/or safety analysis will be listed and the reasons for exclusion will be reported (Listing 16.2.3.1).

6.6.4 Discontinued subjects

All subjects who discontinued the clinical trial will be listed. Sex, age, last visit performed before discontinuation, type of discontinuation, date of premature discontinuation and primary reason for subject premature discontinuation will be reported (Listing 16.2.1.1).

6.6.5 Protocol deviations

All the protocol deviations reported during the clinical trial will be listed by subject and summarised by treatment group (CB-03-01 and Vehicle); the number and percentage of subjects for each deviation will be reported (Listing 16.2.2.1, Table 14.1.1.8, Table 14.1.1.9).

6.6.6 Blind breaking

In case of any blind breaking, whether the unblinding was confirmed or not and the reason for unblinding will be listed by treatment group (Listing 16.2.2.2).

6.6.7 Mismatches of assigned and received IMPs

In case of any mismatches of assigned and received IMPs, the assigned and actually received IMPs will be listed (Listing 16.2.2.3).

6.6.8 Demography

Demographic data will be listed and summarised by treatment group (CB-03-01 and Vehicle). The number and percentage of subjects in each category of categorical variables (e.g. sex) and the descriptive statistics (mean, standard deviation, CV%, minimum, median and maximum) of continous variables (e.g. age) will be presented (Listing 16.2.4.3, Table 14.1.1.3, Table 14.1.1.4, Table 14.1.1.5).

6.6.9 Inclusion/Exclusion criteria

All the unmet inclusion/exclusion criteria will be listed and summarised by treatment group (CB-03-01 and Vehicle). The number and percentage of subjects for each skin type will be reported (Listing 16.2.4.4, Table 14.1.1.6, Table 14.1.1.7).

6.6.10 Fitzpatrick Skin Type Assessment

The Fitzpatrick skin type assessments will be listed and summarised by treatment group (CB-03-01 and Vehicle). The number and percentage of subjects for each skin type will be reported (Listing 16.2.10.1, Table 14.1.1.10, Table 14.1.1.11).

6.6.11 Medical and dermatological history

All the findings of medical history of all subjects enrolled in the study will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) and listed (Listing 16.2.10.2). The version of the dictionary used for coding will be reported in the clinical study report.

The findings of medical history will be summarised by treatment group (CB-03-01 and Vehicle). The number and percentage of subjects with any findings will be presented by PT and SOC (Table 14.1.1.12).

6.6.12 Dermatological examination and brief physical examination

Whether the dermatological examination was performed or not and, if performed, the date of examination and the presence of facial acne vulgaris, hypopigmentation, hyperpigmentation and other dermatological findings on face and whether the physical examination was performed or not and, if performed, the date of examination will be listed (Listing 16.2.10.3).

6.6.13 Prior and concomitant medications

All prior and concomitant medications will be coded using the World Health Organization Drug Dictionary Enhanced (WHO DDE) and listed (Listing 16.2.10.4). The version of the dictionary used for coding will be reported in the clinical study report.

Prior and concomitant medications will be listed and summarized by treatment group (CB-03-01 and Vehicle). The number and percentage of subjects with any concomitant medication will be presented by and ATC 4th level (or the higher available ATC level if 4th level is missing) and standardised drug name (Table 14.1.1.13, Table 14.1.1.14).

6.6.14 Concurrent Therapies and Procedures

Non medication concurrent therapies and procedures will be listed (Listing 16.2.10.5) by treatment group (CB-03-01 and Vehicle).

6.6.15 Photographs

Whether photographs were taken or not and, as applicable, the reason for not collecting them or the date of collection will be listed (Listing 16.2.10.6) by treatment group (CB-03-01 and Vehicle).

6.6.16 Subjects study visits

The dates of all subjects study visits will be listed (Listing 16.2.10.7) by treatment group (CB-03-01 and Vehicle).

6.7 Efficacy analysis

Subjects will be analysed according to the analysis center they will be assigned to and to the treatment they were assigned to (ITT, PP).

The original clinical center of the subjects will be used only for the the investigation of center-to-center variability.

The investigator's global assessment scores will be listed (Listing 16.2.6.1) and summarised by treatment group (CB-03-01 and Vehicle) using contingency tables (Table 14.2.1.1, Table 14.2.1.2).

The proportion of subjects with at least a two-point reduction in IGA AND an IGA score of 0 (clear) or 1 (almost clear) will be summarised by treatment group (CB-03-01 and Vehicle) using contingency tables (Table 14.2.1.3, Table 14.2.1.4).

The acne lesion counts will be listed by treatment group (Listing 16.2.6.2); the values of NILC and ILC and their changes from baseline will be summarised by treatment group (CB-03-01 and Vehicle) using descriptive statistics (Table 14.2.1.5, Table 14.2.1.6).

6.7.1 Primary efficacy analysis

6.7.1.1 Proportion of subjects in each treatment group with at least a two-point reduction in IGA compared to Baseline AND an IGA score of 0 (clear) or 1 (almost clear) at Week 12 - ITT

A logistic regression model with treatment and analysis center as fixed effects will be used to compare the proportion of subjects in each treatment group with at least a two-point reduction in IGA AND an IGA score of 0 (clear) or 1 (almost clear) at Week 12 ($\pi_{2p(0,1)}$). Subjects will be analysed according to the treatment they were assigned to (ITT) and missing data will be replaced using a multiple imputation (MI) under missing at random (MAR) assumption.

The adjusted odds ratio of the comparison of $\pi_{2p(0,1)}$ between CB-03-01 and Vehicle and its 95% confidence interval will be derived from the regression model (Table 14.2.2.1).

6.7.1.2 Absolute change from Baseline in NILC at Week 12 - ITT

An analysis of covariance (ANCOVA) will be used to compare the absolute change from Baseline in NILC in each treatment group at Week 12, with treatment and analysis center as fixed effects and the Baseline NILC as covariate. Subjects will be analysed according to the treatment they were assigned to (ITT) and missing data will be replaced using a multiple imputation (MI) under missing at random (MAR) assumption.

The adjusted means difference of the comparison of NILC between CB-03-01 and Vehicle and its 95% confidence interval will be derived from the regression model (Table 14.2.2.2).

6.7.1.3 Absolute change from Baseline in ILC at Week 12 - ITT

An ANCOVA will be used to compare the absolute change from Baseline in ILC in each treatment group at Week 12, with treatment and analysis center as fixed effects and the Baseline ILC as covariate. Subjects will be analysed according to the treatment they were assigned to (ITT) and missing data will be replaced using a multiple imputation (MI) under missing at random (MAR) assumption.

The adjusted means difference of the comparison of ILC between CB-03-01 and Vehicle and its 95% confidence interval will be derived from the regression model (Table 14.2.2.3).

6.7.2 Secondary efficacy analysis

6.7.2.1 Absolute change from Baseline in TLC at Week 12 - ITT

An analysis of covariance (ANCOVA) will be used to compare the absolute change from Baseline in TLC in each treatment group at Week 12, with treatment and analysis center as fixed effects and the Baseline TLC as covariate. Subjects will be analysed according to the treatment they were assigned to (ITT) and missing data will be replaced using a multiple imputation (MI) under missing at random (MAR) assumption.

The adjusted means difference of the comparison of TLC between CB-03-01 and Vehicle and its 95% confidence interval will be derived from the regression model (Table 14.2.2.4).

6.7.2.2 Percent change from Baseline in NILC at Week 12 - ITT

An analysis of covariance (ANCOVA) will be used to compare the percent change from Baseline in NILC in each treatment group at Week 12, with treatment and analysis center as fixed effects and the Baseline NILC as covariate. Subjects will be analysed according to the treatment they were assigned to (ITT) and missing percent changes will be calculated using the values obtained from the multiple imputation (MI) under missing at random (MAR) assumption performed for the NILC (see paragraph 6.7.1.2).

The adjusted percent means difference of the comparison of NILC between CB-03-01 and Vehicle and its 95% confidence interval will be derived from the regression model (Table 14.2.2.5).

6.7.2.3 Percent change from Baseline in ILC at Week 12 - ITT

An analysis of covariance (ANCOVA) will be used to compare the percent change from Baseline in ILC in each treatment group at Week 12, with treatment and analysis center as fixed effects and the Baseline ILC as covariate. Subjects will be analysed according to the treatment they were assigned to (ITT) and missing percent changes will be calculated using the values obtained from the multiple imputation (MI) under missing at random (MAR) assumption performed for the ILC (see paragraph 6.7.1.3).

The adjusted percent means difference of the comparison of ILC between CB-03-01 and Vehicle and its 95% confidence interval will be derived from the regression model (Table 14.2.2.6).

6.7.2.4 Percent change from Baseline in TLC at Week 12 - ITT

An analysis of covariance (ANCOVA) will be used to compare the percent change from Baseline in TLC in each treatment group at Week 12, with treatment and analysis center as fixed effects and the Baseline TLC as covariate. Subjects will be analysed according to the treatment they were assigned to (ITT) and missing percent changes will be calculated using the values obtained from the multiple imputation (MI) under missing at random (MAR) assumption performed for the TLC (see paragraph 6.7.2.1).

The adjusted percent means difference of the comparison of TLC between CB-03-01 and Vehicle and its 95% confidence interval will be derived from the regression model (Table 14.2.2.7).

6.7.3 Sensitivity analyses

The robustness of the result obtained on the ITT set for the primary endpoints will be tested by performing the following sensitivity analyses.

6.7.3.1 Per Protocol analysis

6.7.3.1.1 Proportion of subjects in each treatment group with at least a two-point reduction in IGA compared to Baseline AND an IGA score of 0 (clear) or 1 (almost clear) at Week 12 - PP

A logistic regression model with treatment and analysis center as fixed effects will be used to compare the proportion of subjects in each treatment group with at least a two-point reduction in IGA AND an IGA score of 0 (clear) or 1 (almost clear) at Week 12 ($\pi_{2p(0,1)}$). Subjects will be analysed according to the treatment they were assigned to (PP).

The adjusted odds ratio of the comparison of $\pi_{2p(0,1)}$ between CB-03-01 and Vehicle and its 95% confidence interval will be derived from the regression model (Table 14.2.3.1).

6.7.3.1.2 Absolute change from Baseline in NILC at Week 12 - PP

An analysis of covariance (ANCOVA) will be used to compare the absolute change from Baseline in NILC in each treatment group at Week 12, with treatment and analysis center as fixed effects and the Baseline NILC as covariate. Subjects will be analysed according to the treatment they were assigned to (PP).

The adjusted means difference of the comparison of NILC between CB-03-01 and Vehicle and its 95% confidence interval will be derived from the regression model (Table 14.2.3.2).

6.7.3.1.3 Absolute change from Baseline in ILC at Week 12 - PP

An ANCOVA will be used to compare the absolute change from Baseline in ILC in each treatment group at Week 12, with treatment and analysis center as fixed effects and the Baseline ILC as covariate. Subjects will be analysed according to the treatment they were assigned to (PP).

The adjusted means difference of the comparison of ILC between CB-03-01 and Vehicle and its 95% confidence interval will be derived from the regression model (Table 14.2.3.3).

6.7.3.2 Missing as worst value analysis

The analyses of sections 6.7.1.1, 6.7.1.2 and 6.7.1.3 will be repeated by replacing missing data according to the Missing as Worst Value (MWV) approach. Subjects will be analysed according to the treatment they were assigned to (ITT) (Table 14.2.3.4, Table 14.2.3.5, Table 14.2.3.6).

6.7.3.3 Worst case analysis

The analyses of sections 6.7.1.1, 6.7.1.2 and 6.7.1.3 will be repeated by replacing missing data according to the Worst Case (WC) approach. Subjects will be analysed according to the treatment they were assigned to (ITT) (Table 14.2.3.7, Table 14.2.3.8, Table 14.2.3.9).

6.7.3.4 LOCF analysis

The analyses of sections 6.7.1.1, 6.7.1.2 and 6.7.1.3 will be repeated by replacing missing data according to the Last Observation Carried Forward (LOCF) approach. Subjects will be analysed according to the treatment they were assigned to (ITT) (Table 14.2.3.10, Table 14.2.3.11, Table 14.2.3.12).

6.7.3.5 BOCF analysis

The analyses of sections 6.7.1.1, 6.7.1.2 and 6.7.1.3 will be repeated by replacing missing data according to the Baseline Observation Carried Forward (BOCF) approach. Subjects will be analysed according to the treatment they were assigned to (ITT) (Table 14.2.3.13, Table 14.2.3.14, Table 14.2.3.15).

6.7.4 Center by treatment interaction

An analysis of the center by treatment interaction will be performed in order to evaluate the presence of a significant interaction and, in case of a significant interaction is observed, in order to ensure that the primary efficacy results are not driven by extreme centers. Two different analysis of the center by treatment interaction for the primary efficacy results will be performed, one for the analysis centers and one for the original centers.

6.7.4.1 Analysis center by treatment interaction

The analyses of sections 6.7.1.1, 6.7.1.2 and 6.7.1.3 will be repeated by adding the analysis center by treatment interaction as fixed effect. The reference levels for the effects of the analysis model will be the analysis center with the highest number of subjects and the vehicle treatment, respectively (Table 14.2.3.16, Table 14.2.3.17, Table 14.2.3.18).

If the analysis center by treatment interaction factor is significant (Wald p-value < 0.10) a sensitivity analysis that excludes analysis centers with the extreme efficacy results will be performed to determine the robustness of the treatment effect. Otherwise, the data will be considered to be free of the impact of extreme analysis centers.

The identification of the extreme analysis centers will be based on the Wald p-values of the analysis center by treatment interaction terms.

Analysis centers will be excluded one by one according to the significance of the interaction term (starting from the one with the most significant interaction term). After each exclusion, the analysis will be repeated on the remaining centers until the analysis center by treatment interaction factor is not significant (Wald p-value ≥ 0.10).

Upon identification of the extreme analysis center(s), the analyses of sections 6.7.1.1, 6.7.1.2 and 6.7.1.3 will be repeated considering only the remaining analysis centers. Inferences will be based on this analyses in addition to observations regarding the extreme analysis center(s). Conclusions will be presented as appropriate to the findings of the sensitivity analysis.

6.7.4.2 Original center by treatment interaction

The analyses of sections 6.7.1.1, 6.7.1.2 and 6.7.1.3 will be repeated considering the original center as fixed effect and by adding the original center by treatment interaction as fixed effect. The reference levels for the effects of the analysis model will be the center with the highest number of subjects and the vehicle treatment, respectively (Table 14.2.3.19, Table 14.2.3.20, Table 14.2.3.21).

If the original center by treatment interaction factor is significant (Wald p-value < 0.10) a sensitivity analysis that excludes original centers with the extreme efficacy results will be performed to determine the robustness of the treatment effect. Otherwise, the data will be considered to be free of the impact of extreme centers.

The identification of the extreme centers will be based on the Wald p-values of the center by treatment interaction terms.

Centers will be excluded one by one according to the significance of the interaction term (starting from the one with the most significant interaction term). After each exclusion, the

analysis will be repeated on the remaining original centers until the original center by treatment interaction factor is not significant (Wald p-value ≥ 0.10).

Upon identification of the extreme original center(s), the analyses of sections 6.7.1.1, 6.7.1.2 and 6.7.1.3 will be repeated considering only the remaining original centers. Inferences will be based on this analyses in addition to observations regarding the extreme original center(s). Conclusions will be presented as appropriate to the findings of the sensitivity analysis.

6.8 Subgroup Analyses

No subgroup analysis is planned.

6.9 Treatment compliance and extent of exposure analyses

6.9.1 Treatment compliance

The dispensation date, the weight of the dispensed article, the return date, the weight of the returned article, the number of applied doses, the number of expected doses and compliance rate will be listed (Listing 16.2.5.1) by treatment group (CB-03-01 and Vehicle).

Descriptive statistics will be used to summarize the test article compliance at each visit and overall by treatment group (CB-03-01 and Vehicle) for the ITT and PP sets. The proportion of compliant (i.e. with a compliance ≥ 80%) and non-compliant (i.e. with a compliance < 80%) subjects will be summarized by table of frequencies at each visit and overall by treatment group (CB-03-01 and Vehicle) for the ITT and PP sets (Table 14.2.4.1, Table 14.2.4.2). Subjects with an overall compliance not evaluable (due to missing data) or less than 80% will be excluded from the PP set (see section 6.2.1).

6.9.2 Extent of exposure

The date/time of first and last application, the total amount of test article used (calculated as number of grams applied for each subject from the weights of the returned test articles) and the mean daily amount of test article applied (calculated as the total amount of test article used divided by the number of days of treatment) will be listed (Listing 16.2.5.2) by treatment group (CB-03-01 and Vehicle).

Descriptive statistics will be used to summarize exposure to the test article at each visit and overall by treatment group (CB-03-01 and Vehicle) for the ITT, PP and Safety sets (Table 14.2.4.3, Table 14.2.4.4, Table 14.2.4.5).

6.10 Safety analysis

All safety analyses will be performed on subjects of the Safety set. Subjects will be analysed according to the treatment they actually received.

6.10.1 Local Skin Reactions (LSRs)

LSRs (telangiectasia, skin atrophy, striae rubrae, erythema, edema, scaling/dryness, stinging/burning, and pruritus) will be listed (Listing 16.2.9.1) and summarized by the frequency of each individual LSR by treatment group (CB-03-01 and Vehicle) and severity at each visit (Table 14.3.5.1).

6.10.2 Adverse events

All AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) and listed. The version of the dictionary used for coding will be reported in the clinical study report.

AEs will be classified as Pre-Treatment AEs (PTAEs) and Treatment Emergent AEs (TEAEs), according to the period of occurrence, as follows:

- PTAEs: all AEs occurring before the first dose of the test article and not worsening after the first dose of the test article;
- TEAEs: all AEs occurring or worsening after the first dose of the test article.

Individual TEAEs and PTAEs will be listed by treatment group (Listing 16.2.7.1, Listing 16.2.7.2).

No summary table will be provided for PTAEs.

TEAEs will be summarised by treatment group and overall.

- The number and percentage of subjects with any TEAE and the number of TEAEs will be presented overall (Table 14.3.1.1) and by SOC and PT (Table 14.3.1.2);
- The number and percentage of subjects with any TEAE by relationship and the number of TEAEs by relationship will be presented (Table 14.3.1.1);
- The number and percentage of subjects with any TEAE related to the IMP and the number of TEAEs related to the IMP will be presented by SOC and PT (Table 14.3.1.3);
- The number and percentage of subjects with any TEAE by severity and the number of TEAEs by severity will be presented overall (Table 14.3.1.1) and by SOC and PT (Table 14.3.1.4);
- The number and percentage of subjects with any TEAE leading to discontinuation and the number of TEAEs leading to discontinuation will be presented (Table 14.3.1.1).

Serious TEAEs will be summarised by treatment group and overall.

- The number and percentage of subjects with any Serious TEAE and the number of Serious TEAEs will be presented overall (Table 14.3.1.1) and by SOC and PT (Table 14.3.1.5);

- The number and percentage of subjects with any Serious TEAE by relationship and the number of Serious TEAEs by relationship will be presented (Table 14.3.1.1);
- The number and percentage of subjects with any Serious TEAE related to the IMP and the number of Serious TEAEs related to the IMP will be presented by SOC and PT (Table 14.3.1.6);
- The number and percentage of subjects with any Serious TEAE leading to discontinuation and the number of Serious TEAEs leading to discontinuation will be presented (Table 14.3.1.1).
- The number and percentage of subjects with any Serious TEAE leading to death and the number of Serious TEAEs leading to death will be presented (Table 14.3.1.1).

All Serious TEAEs will be listed and all TEAEs leading to discontinuation will be listed (Table 14.3.2.1).

6.10.3 Electrocardiograms (ECGs)

ECGs will be evaluated for any material change during the study period. Overall interpretation (normal/borderline/abnormal) of the ECG will be listed (Listing 16.2.9.2) and summarised by treatment group (CB-03-01 and Vehicle) at each visit using contingency tables (Table 14.3.5.2). Values for heart rate, RR duration, PR duration, QRS duration, QT duration including Bazett's and Fridericia's corrections, P axis and their changes from baseline will be listed and summarised by treatment group (CB-03-01 and Vehicle) at each visit (Table 14.3.5.2) using descriptive statistics (mean, standard deviation, CV%, minimum, median and maximum). Changes in overall interpretation (normal/borderline/abnormal) of the ECG from Baseline to Visit 4 (Day 85 ± 10) will be examined using shift tables (Table 14.3.5.3).

6.10.4 Vital signs

Vital signs will be listed (Listing 16.2.9.3) and summarised by treatment group (CB-03-01 and Vehicle). Descriptive statistics (mean, standard deviation, CV%, minimum, median and maximum) will be presented (Table 14.3.5.4).

6.10.5 Pregnancy tests

The results of pregnancy test at Visit 1 (Baseline), Visit 2 (Day 29 ± 5), Visit 3 (Day 57 ± 7) and Visit 4 (Day 85 ± 10) or end of treatment if the subject withdraws prematurely will be listed (Listing 16.2.8.1).

6.11 Interim Analyses

No interim statistical analysis is planned.

7. ANALYSIS DATASETS

Analysis datasets will be created according to the ADaM model of CDISC (12).

8. REFERENCES

1. ICH Topic E9: Statistical principles for clinical trials. CPMP/ICH/363/96
2. Guidance for Industry - Acne Vulgaris: Developing Drugs Treatment - Draft version, September 2005
3. EMA - Points to consider in multiplicity issues in clinical trials - 19 September 2002
4. CB-03-01/26. A Phase 3, Multicenter, Randomized, Double-Blind, Vehicle-Controlled Study to Evaluate the Safety and Efficacy of Cortexolone 17α-Propionate (CB-03-01) 1% Cream Applied Twice-Daily for 12 Weeks in Subjects with Facial Acne Vulgaris. Cassiopea SpA, June 11, 2015
5. Celasco G, Moro, L, Bozzella R, Ferraboschi P, Bartorelli L, Quattrocchi C, Nicoletti F. Biological profile of cortexolone 17α-propionate (CB-03-01), a new topical and peripherally selective androgen antagonist. Arzneim.-Forsch. 2004; 54, 881-886
6. 171-7151-201. A Phase 2, Multicenter, Randomized, Double-Blind, Vehicle-Controlled, Dose Escalating Study to Evaluate the Safety and Efficacy of Cortexolone 17α-Propionate (CB-03-01) Cream Applied Once or Twice-Daily for 12 Weeks in Subjects with Facial Acne Vulgaris. Intrepid Therapeutics, Inc., 2014
7. Hansheng Wang, Shein-Chung Chow, "SAMPLE SIZE CALCULATION FOR COMPARING MEANS", Wiley Encyclopedia of Clinical Trials, Copyright © 2007, John Wiley & Sons, Inc.
8. Hansheng Wang, Shein-Chung Chow, "SAMPLE SIZE CALCULATION FOR COMPARING PROPORTIONS", Wiley Encyclopedia of Clinical Trials, Copyright © 2007, John Wiley & Sons, Inc.
9. SAS/STAT® 9.3 User's Guide
10. John W. Graham, Allison E. Olchowski, Tamika D. Gilreath, "How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory", Prev Sci (2007) 8:206-213 DOI 10.1007/s11121-007-0070-9
11. Binhuan Wang, New York University School of Medicine; Yixin Fang, New York University School of Medicine; Man Jin, Forest Research Institute; "Combining Type-III Analyses from Multiple Imputations"; Paper 1543-2014
12. Analysis Data Model (ADaM) - Implementation Guide - Final version 1.0, December 17, 2009

9. APPENDICES

Appendix 1. Section 14 - Tables Shells

Appendix 2. Section 16.2 - Individual Subject Data Listings Shells