DS Biopharma
Protocol #: # DS107E-06

A Randomised, Double-blind, Vehicle-Controlled, Phase IIb Study to Assess the Efficacy and Safety of Topically Applied DS107 Cream to Adults with Mild to Moderate Atopic Dermatitis

Protocol Version 5.0/Amend. #4

Statistical Analysis Plan

Created by: Clinical Trial Data Services, LLC
40 Nagog Park, Suite #210
Acton, MA 01720

Lavin Consulting LLC
3 Cahill Park Drive
Framingham MA 01702

By signing the following, I agree to the contents in the Statistical Analysis Plan and its associated attachments.

Approved by:

Name	Title	Signature	Date
Phil Lavin, PhD	Senior Biostatistical Consultant (for sponsor)		
Terri Sampo, MPH	Principal Consultant, CTDS		10/11/2017

LIST OF ABBREVIATIONS

15-HETrE	15-hydroxyeicosatrienoic acid
AD	Atopic Dermatitis
AE	Adverse Event
BD	Bis Die (Twice daily)
BP	Blood Pressure
BPM	Beats Per Minute
CM	Concomitant Medication
COX	Cyclooxygenase
CRA	Clinical Research Associate
CRO	Contract Research Organisation
CRF	Case Report Form
CsA	Cyclosporin A
CDISC	Clinical Data Interchange Standards Consortium
CTA	Clinical Trial Agreement
DGLA	Dihomo-Gamma-Linolenic Acid
DLQI	Dermatology Life Quality Index
DM	Data Manager
DS	DS Biopharma
EASI	Eczema Area and Severity Index
EC	Ethics Committee
EDC	Electronic Data Capture
FAS	Full Analysis Set
FSH	Follicule Stimulating Hormone
GCP	Good Clinical Practice
GENMOD	Generalized Linear Model (binary outcomes)
GLP	Good Laboratory Practice
ICF	Informed Consent Form
ICH	International Conference on Harmonisation
IGA	Investigator's Global Assessment
IgE	Imunnoglobulin E
IMP	Investigational Medicinal Product
ISF	Investigator Site File
ITT	Intention-To-Treat
IWRS	Interactive Web Response System
MCAR	Missing Completely at Random
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model with Repeated Measures
MNAR	Missing Not at Random
NOAEL	No Observed Adverse Event Limit
NRS	Numeric Rating Scale
OTC	Over The Counter

PBIR	Probability of Being in Response
PGD1	Prostaglandin D1
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PIS	Patient Information Sheet
PK	Pharmacokinetics
POEM	Patient Orientated Eczema Measure
PPS	Per Protocol Analysis Set
PUVA	Psoralen & Ultraviolet A
PV CRO	Pharmacovigilance Contract Research Organisation
SAE	Serious Adverse Event
SAS	Statistical Analysis System
SAP	Statistical Analysis Plan
SCORAD	SCORing AD
SDV	Source Data Verification
SDTM	Study Data Tabulation Model
SOP	Standard Operating Procedure
SPC	Summary of Products Characteristics
SS	Safety Analysis Set
SUSAR	Suspected Unexpected Serious Adverse Reaction
Th	T helper cell
TNF-α	Tumor Necrosis Factor-Alpha
UV-A/B	Ultraviolet-A/B
VAS	Visual Analogue Scale

Table of Contents

1 Protocol Summary

Title	A Randomised, Double-blind, Vehicle-Controlled, Phase IIb Study to Assess the Efficacy and Safety of Topically Applied DS107 Cream to Adults with Mild to Moderate Atopic Dermatitis
Study Objectives	Efficacy Objective: To compare the efficacy of topically applied DS107 cream (1% & 5%) versus vehicle in the treatment of adult patients with mild to moderate Atopic Dermatitis (AD). Safety Objective: To compare the safety of topically applied DS107 cream (1% & 5%) versus vehicle in the treatment of adult patients with mild to moderate AD.
Primary Endpoint	*Primary Endpoint* • Change from baseline in Numeric Rating Scale (NRS) for Pruritus in the treated DS107 dose populations compared to vehicle population at Week 8/exit. *Co-Primary Endpoint* • Change from baseline in Eczema Area and Severity Index (EASI) in the treated DS107 dose populations compared to vehicle population at Week 8/exit.
Secondary Endpoint(s)	Change from baseline in Numeric Rating Scale (NRS) for Pruritus in the treated DS107 dose populations compared to vehicle population at Weeks 2, 4, 6, 10, and last treatment visit. Proportion of patients achieving a decrease of at least 2.7 points in NRS in the treated DS107 dose populations compared to vehicle population from baseline to Week 2, 4, 6, 8 and 10. Change from baseline in Eczema Area and Severity Index (EASI) in the treated DS107 dose populations compared to vehicle population at Weeks 2, 4, 6, 10, and last treatment visit. Proportion of patients achieving an IGA score of 0 (clear) or 1 (almost clear) and a decrease of at least 2 points in IGA in the treated DS107 dose populations compared to vehicle population from baseline to Week 2, 4, 6, 8, 10, and last treatment visit.

	Proportion of patients achieving a decrease of at least 2 points in IGA in the treated DS107 dose populations compared to vehicle population from baseline to Week 2, 4, 6, 8, 10, and last treatment visit. Change from baseline in IGA score in the DS107 dose treated populations compared to vehicle population at Weeks 2, 4, 6, 8, 10, and last treatment visit.
Exploratory Endpoints	• Time to response by Week 8. • Probability of being in response (PBIR) between Weeks 2 and 8. • Sustained efficacy between Weeks 8 and 10. • Change from baseline in the Dermatology Life Quality Index (DLQI) score in the DS107 dose treated populations compared to vehicle population at Week 2, 4, 6, 8, 10, and last treatment visit and from Week 8 to Week 10. • Change from baseline in the Patient Orientated Eczema Measure (POEM) score in the DS107 dose treated populations compared to vehicle population at Week 2, 4, 6, 8, 10, and last treatment visit and from Week 8 to Week 10. • Change from baseline in the Patient Global Impression of Severity (PGI-S) score in the DS107 dose treated populations compared to vehicle population at Week 2, 4, 6, 8, 10, and last treatment visit and from Week 8 to Week 10. • Change from baseline in the Patient Global Impression of Change (PGI-C) score in the DS107 treated populations compared to vehicle population at Week 2, 4, 6, 8, and 10, and last treatment visit and from Week 8 to Week 10. • Determination of AD biomarkers in the DS107 dose treated populations compared to vehicle population at Baseline/Day 0 and Week 8/Early Termination (samples to be retained for the potential analysis at a later date).
Safety Endpoints	• Adverse event (AE) and serious adverse event (SAE) frequency and severity.

	• Safety laboratory parameters (haematology, clinical chemistry, urinalysis). • Clinical safety examinations (vital signs, physical examination).
Test Product, Dose and Mode of Administration	All study medication will be blinded. DS107 cream will be provided as a cream containing either 5% or 1% DS107. Vehicle will be provided as a matching cream. This study will involve two dose levels of DS107 for 8 weeks (5% DS107 cream BD or 1% DS107 cream BD) and a vehicle control BD
Study Design:	Approximately 300 patients with mild to moderate AD will be included in this multicenter, double-blind, vehicle controlled, 3 arm, Phase IIb study. The sample size may be increased at an interim analysis (conducted after n=150 complete Week 8/Exit) to update primary efficacy endpoint assumptions and to achieve at least 80% power and two-sided 2.6% Type I error (to allow for each DS107 dose to be separately compared to vehicle control). All patients will sign an informed consent and undergo screening for study eligibility. Patients will be randomized (1:1:1) at baseline visit to either receive 5% DS107 cream, 1% DS107 cream or vehicle cream twice daily for 8 weeks. During the 8 weeks of treatment patients will have to liberally apply their assigned treatment topically to all affected or commonly affected areas twice daily (morning and evening). Patients will come to the clinic on 7 separate visits: Screening, Baseline (Day 0), Week 2, Week 4, Week 6, Week 8 (end of treatment) and Week 10 (follow-up). All patients will exit the study at the Week 10 visit.
Inclusion Criteria	1. Patients with a clinically confirmed diagnosis of active AD according to Hanifin and Rajka criteria. 2. Patients with mild to moderate AD at baseline as defined by an IGA score of 3 or 2 at baseline visit. Patients who are classified as having moderate AD should also have an EASI score of ≥12 at the baseline visit. 3. Patients with AD covering a minimum 5% of the body surface area at baseline.

	4. Patients whose pre-study clinical laboratory findings do not interfere with their participation in the study, in the opinion of the Investigator. 5. Patients who are able and willing to stop current treatments for AD, including the use of emollients on the affected skin, throughout the study. 6. Male or female patients aged 18 years and older on the day of signing the informed consent form (ICF). 7. Female patients and male patients with female partners of child bearing potential must use adequate contraception or have a sterilized partner for the duration of the study. Adequate contraception is defined as: systemic hormonal contraceptives, intrauterine device or barrier method of contraception in conjunction with spermicide, or agree to sexual abstinence. Hormonal contraceptives must be on a stable dose for at least one month before baseline. 8. Patients who are able to communicate well with the Investigator, to understand and comply with the requirements of the study, and understand and sign the written informed consent.
Exclusion Criteria	1. Patients with other skin conditions that might interfere with AD diagnosis and/or evaluation (such as psoriasis or current active viral, bacterial and fungal topical skin infections) as assessed by the Investigator. 2. Patients who have used systemic treatments (other than biologics) that could affect AD less than 4 weeks prior to baseline visit (Day 0), e.g. retinoids, methotrexate, cyclosporine, hydroxycarbamide (hydroxyurea), azathioprine and oral/injectable corticosteroids. Intranasal corticosteroids and inhaled corticosteroids for stable medical conditions are allowed. 3. Patients who have used any topical medicated treatment for AD two weeks prior to start of treatment/baseline (Day 0) including but not limited to, topical corticosteroids, calcineurin inhibitors, tars, bleach, antimicrobials and bleach baths. 4. Patients who use topical products containing urea, ceramides or hyaluronic acid two weeks prior to Day 0. 5. Patients who use anti-histamines for AD within 3 days of baseline. Non-sedative anti-histamines for other indications may be used throughout the study provided the patient is on a stable dose for 4 weeks prior to Baseline. 6. Patients who have had excessive sun exposure, have used tanning booths or other ultraviolet (UV) light sources four weeks prior to baseline (Day 0) and/or are planning a trip to a sunny

	climate or to use tanning booths or other UV sources between screening and follow-up visits. 7. Patients who have a history of hypersensitivity to any substance in DS107 or vehicle cream. 8. Patients who have a white cell count outside of the normal reference range at screening, which cannot be justified by the investigator. 9. Patients who have any clinically significant controlled or uncontrolled medical condition or laboratory abnormality that would, in the opinion of the Investigator, put the patient at undue risk or interfere with interpretation of study results. 10. Patients who have a clinically significant impairment of renal or hepatic function. 11. Patients with significant uncontrolled cardiovascular, neurologic, malignant, psychiatric, respiratory or hypertensive disease, as well as uncontrolled diabetes and floride arthritis or any other illness that, in the opinion of the Investigator, is likely to interfere with completion of the study. 12. Patients with chronic infectious diseases (e.g., hepatitis B, hepatitis C or infection with human immunodeficiency virus). 13. Patients with a history of clinically significant drug or alcohol abuse in the opinion of the Investigator in the last year prior to baseline (Day 0). 14. Patients who have participated in any other clinical study with an investigational drug within 3 months before the first day of administration of study treatment. 15. Patients who have had treatment with biologics as follows: a. Any cell-depleting agents including but not limited to rituximab: within 6 months before the screening visit, or until lymphocyte count returns to normal, whichever is longer, b. Other biologics influencing cell proliferation: within 6 months before the screening visit. 16. Patients who are pregnant, planning pregnancy, breastfeeding and/or are unwilling to use adequate contraception (as specified in inclusion criterion 7) during the trial. 17. Patients, in the opinion of the Investigator, not suitable to participate in the study.

2 Statistical Methodology

2.1 General Considerations

The objective of this double-blind Phase IIb study is to assess the efficacy and safety of topically applied DS107 Cream to adults with mild to moderate Atopic Dermatitis. This study uses a randomized (1:1:1, Treatment group A: 1% DS107 cream, Treatment group B: 5% DS107 cream, Treatment group C: Vehicle cream), double-blind, vehicle-controlled parallel group design. Disposition, compliance, demographics, and primary and secondary efficacy results will be presented by treatment group separately for mild and moderate AD as well as for combined groups. Study medication will be applied twice daily to the designated Treatment Area for 8 weeks. An End of Therapy (EOT) visit will occur at Week 8 for subjects who complete 8 weeks of treatment or earlier if the subject discontinues from the study prematurely. An End of Study (EOS) visit will be conducted at 10 weeks post-treatment for all subjects, even if they withdrew early. For subjects who discontinue early, they will also be asked to return two weeks later for the safety assessments listed at Week 10/Visit 7.

All subjects screened into the study will be assigned a patient identifier number during screening that will be used on all patient documentation. The patient identifier number will contain the site number and the patient number assigned in numerical order at the screening visit (e.g.: 102-10 for the tenth patient screened at the site number 102). Patient numbers will be assigned in ascending order starting with 01. All data will be listed in subject data listings sorted by site and subject number unless otherwise noted. Data tabulations will be prepared as described in the sections below.

All statistical processing will be performed using Statistical Analysis System (SAS®) Version 9.3 or higher, unless otherwise noted. All analyses will be conducted in accordance with ICH E6 and E9 standards and applicable Contract Research Organisation (CRO) Standard Operating Procedures (SOPs). CDISC and SDTM data conventions will be followed; these data sets will include DLQI scoring performed by the data management group and provided to CTDS in the SDTM datasets.

The primary focus will be the separate comparisons of the two DS107 concentrations vs. vehicle control using a Dunnett's correction (two-sided p=0.026) to reach statistical significance supported by the two-sided p-values and the two-sided 97.4% confidence intervals (Dunnett's procedure). Comparisons between the two active groups will be considered to be descriptive statistics.

The following summary statistics will be reported for continuous data: number of subjects, mean, standard deviation, median, minimum and maximum. For categorical data, the number and percentage of subjects within each treatment group will be reported.

Baseline values will be defined as the last non-missing measurement prior to dosing with study drug (where baseline record flag='Y' in datasets), unless otherwise specified. Change from baseline will be defined as the post-baseline visit value minus the baseline or previous value unless otherwise specified.

2.2 Subject Population

2.2.1 Baseline Characteristics/Medical History

Demographics - gender, age, race (years), ethnicity, country and baseline BSA will be summarized by treatment group for the Full analysis set (FAS) population. Demographics will be displayed separately for the mild and moderate AD as well as overall.

P-values will be provided for the overall global test and each vehicle pair and for descriptive comparison of 5% DS107 to 1% DS107. Binary measures (gender, ethnicity, country and race—white vs non-white) will be compared using a two-sided Fisher Exact test, while continuous measures (age and baseline BSA) will be compared using ANOVA with Tukey option for the comparison of differences in means between the groups.

Medical History will be tabulated descriptively, including severity of AD, duration of AD (acute, intermediate or chronic) and other relevant medical history. The specific AD medical history reported terms and areas affected will be provided in the listings as text fields.

2.2.2 Case Disposition

All subjects enrolled into the study that are issued a subject number will be accounted for in Subject Disposition. Data will be tabulated by treatment group and the numbers and percentages of subjects in each treatment group who are enrolled, randomized and treated, who comprise the Full Analysis Set (FAS), Safety Analysis Set (SS), and the Per-Protocol Analysis Set (PPS) if there is a >10% loss relative to the FAS analysis populations, and who complete the study or withdraw prematurely along with the reason for discontinuation will be presented. The number of subjects who were seen at each specified study visit will also be tabulated by treatment group and by site. Results will be displayed separately by mild and moderate AD as well as overall.

Protocol deviations (major and minor) will be defined prospectively and will be presented as a listing by treatment group. Major protocol deviations will be tabulated by treatment group.

Reasons for any early withdrawals and the last treatment visit completed prior to discontinuation will be provided in a listing.

2.2.3 Dose Regimen

The following three formulations have identical appearance: 1% DS107 cream, 5% DS107 cream and Vehicle cream in a 1:1:1 ratio, respectively. The medication assigned by randomization will be topically applied liberally by the patients to all affected or commonly affected areas twice daily (morning and evening) for 8 weeks. Tabulations of exposure to study medication are described in Section 2.2.13.1, Extent of Exposure. Details of study medication use will be provided in a subject data listing.

2.2.4 Study Populations for Analysis

2.2.4.1 Full Analysis Set (FAS)

The FAS consists of all patients who are randomized to the study and received at least one dose of study medication. FAS is the primary analysis population for efficacy endpoints. Analysis will be done according to the treatment patients were randomized to.

2.2.4.2 Safety Analysis Set (SS)

The SS consists of all patients who received at least one dose of the medication. SS is the analysis population for all safety endpoints. Analysis will be done according to the actual treatment patients received.

2.2.4.3 Per-Protocol Analysis Set (PPS)

The PPS is the subset of FAS who completed the study without any major violations. Protocol violations will be assessed for each patient in a blinded fashion prior to database lock at a Blind Data Review Meeting (BDRM), and the PPS will also be finalized during this meeting. PPS is a supportive analysis population for the primary efficacy endpoint. Analysis will be done according to the treatment that patients were randomized to.

2.2.4.4 Subgroup Analyses

Subgroup analyses will be conducted on the primary and secondary endpoints stratifying by severity of atopic dermatitis (mild or moderate), defined as an IGA score of 2=mild or 3=moderate at the Baseline/visit 2 visit, which is consistent with the AD severity used for the entrance criteria in the study. Statistical testing for this subgroup analyses will be limited to the primary efficacy endpoint but summary statistics will be provided for all secondary efficacy endpoints.

2.2.5 Study Assessment Time Points

The study consists of seven protocol-specified visits which will be assessed as nominal visits from an analysis perspective:

- Visit 1/Screening visit (Day -30 to -1): Once informed consent has been obtained, the Investigator will assign a Patient Number and the subject will undergo the screening procedures. Following completion of a successful screening visit, patients will begin the comparative treatment period (8 weeks).
- Visit 2/Baseline (Day 0): At the start of the comparative treatment period, after confirmation of continued eligibility, patients will be randomly assigned to one of the three treatment regimens. Study medication is dispensed at Visit 2. The first application of study medication will be carried out at the site once all baseline assessments have been completed. The patient will carry out their second application of study medication in the evening of Day 0.
- Visits 3, 4 and 5: Follow-up visits at Week 2 (Day 14 ± 2 days), Week 4 (Day 28 ± 2 days) and Week 6 (Day 42 ± 2 days) respectively.
- Visit 6 – End of Treatment (EOT): Week 8 (Day 56 ± 2 days) for subjects who complete 8 weeks of treatment; earlier for subjects who discontinue treatment prematurely.
- Visit 7/Week 10: End of Study visit (Day 70 ± 3 days) for all subjects.
- Unscheduled Visits: Unscheduled visits may occur when a patient needs to make a visit in between the scheduled visit dates due to an AE, difficulty complying with the study protocol requirements, or a significant change in their disease state. All procedures that are medically necessary should be followed.

2.2.6 Methods for Handling Missing Data

The effects of missing efficacy data will be incorporated through the appropriate covariance structure from the GLMM. This will apply to continuous as well as binary endpoints.

A sensitivity analysis will be performed for Week 8/Exit outcome under the missing at not random (MNAR) assumption for the primary efficacy analyses using the FAS. The primary statistical analysis assumes "Missing At Random (MAR)" when handling missing data. The treatment effect obtained under the MAR assumption is essentially that which could have been reached if all patients had fully adhered to treatment or, in other words, the effect a patient may expect if they take the medication as directed. This is sometimes known as the 'de jure' or 'efficacy' estimand. Because of the lack of perfect adherence in practice, the 'de facto' or effectiveness treatment effect will also be estimated. This estimand includes assumptions regarding the treatment effects that could be expected to occur when patients discontinue treatment. The jump to reference method described by Carpenter et al. (J Biopharm Stat, 2013; 23(6):1352-71) will be used to estimate the de facto estimand, using the vehicle arm as the reference. This is based on the assumption that patients who discontinue from study drug have no alternative oral treatment option suitable for longer-term use and so their responses are likely to revert to those of the vehicle group. The sensitivity analyses for missing data will be performed on the FAS only.

2.2.7 Safety Monitoring

No safety monitoring committees are planned for the study.

2.2.8 Statistical Hypotheses

The original null hypothesis (H_0) for each analysis is the equality of the means for the three treatment groups. A global test was to be performed across all three treatment groups. The test will still be performed

The revised primary focus will be the separate comparisons of the two DS107 concentrations vs. vehicle control using a Dunnett's correction (two-sided 97.4% confidence intervals and two-sided p=0.026) to reach statistical significance. The null hypotheses are the pairwise comparisons of the equality of the two DS107 concentrations vs. vehicle. Dunnett's procedure will be used to adjust for the multiple comparisons. The overall 5% Type I error rate is controlled for by the Dunnett's correction for multiple comparisons.

Testing between the two active groups will be considered as descriptive statistics.

All statistical testing will be two-sided; global tests will be performed at the 0.05 level while all pairwise DS107 comparisons vs vehicle will be performed at the 0.026 level.

2.2.9 Sample Size Justification

The sample size for the topical DS107 Phase IIb study was informed from the post-hoc analysis of the Phase IIa study. NRS and EASI will be the primary and co-primary efficacy endpoints. The positive efficacy data over 4 weeks showed no sign of a plateau and suggests further improvements may be observed at Week 8. Dunnettt's procedure will be used to allow simultaneous comparisons of each DS dose concentration vs. vehicle control.

The sample size was estimated from the Phase IIa results at Week 4/exit; the mean improvement and corresponding standard deviation for vehicle subjects were used to perform the following calculations. As shown below, 100 patients per treatment arm (300 patients in total) will be required to detect mean advantages of 3.03 units for NRS (10-point scale) and 6.6 units for EASI favoring DS107 vs. vehicle. This assumes a two-sided test with 80% power conducted at the 5% overall significance level with a Dunnettt's adjustment for multiple doses.

Table 1. Sample Size Calculation

Two group Satterthwaite t-test: No Difference (unequal variances) (equal n's)		
	NRS	**EASI**
Overall Dunnett's Type I Error	0.026	0.026
1 or 2 sided test?	2	2
Target DS107 Mean Reduction, μ1	3.030	6.600
Expected Vehicle Mean Reduction, μ2	1.810	4.000
Mean Difference Goal, μ1 - μ2	1.220	2.600
Phase 2a DS107 Standard Deviation, σ1	2.490	4.450
Phase 2a Vehicle Standard Deviation, σ2	3.040	7.110
Power (%)	80	80
n per group	100	100

2.2.10 Interim Analyses

An interim analysis is planned to assess the primary and co-primary efficacy endpoint futility, to assess the original primary efficacy endpoint, and to adjust

sample size. The sample size may be increased at the interim analysis (conducted after n=150 complete Week 8/Exit) to update primary efficacy endpoint assumptions and to achieve at least 80% power and two-sided 2.6% Type I error (to allow for each DS107 dose to be separately compared to vehicle control).

2.2.11 Randomization/Unblinding

2.2.11.1 Randomization Methodology

Approximately 300 patients will be randomized to the three blinded treatment groups in a 1:1:1 ratio as follows:

- Treatment group A: 1% DS107 cream applied topically to all affected or commonly affected areas twice-daily for 8 weeks
- Treatment group B: 5% DS107 cream applied topically to all affected or commonly affected areas twice-daily for 8 weeks
- Treatment group C: Vehicle cream applied topically to all affected or commonly affected areas twice-daily for 8 weeks.

A randomization list, permuted blocks and stratified by site was generated by HMD Clinical (25 Balmuir Avenue, Bathgate, EH48 4BW, U.K.). The randomization schedule with study drug assignments was generated prior to the start of the study and will be known only to the individuals responsible for labelling the study drug, the statisticians generating the schedule and the IWRS team responsible for implementing the schedule. The IWRS will assign a medication kit number to each patient and the contents will be based on the randomization code.

2.2.11.2 Unblinding

All study site personnel, as well as the personnel involved in the monitoring or conduct of the study, will be blinded to the individual patient treatment assignments. Randomisation details will be kept strictly confidential, accessible only in an emergency to authorized persons, until the time of formal unblinding. The blinded code for the trial will be broken only after all patient data has been recorded and verified and the database locked.

Emergency unblinding will be carried out through the IWRS system with relevant site personnel and pharmacovigilance monitors provided with the required system access to carry out unblinding.

Any unblinding (and reason for unblinding) prior to database lock will be reported in the final CSR.

2.2.12 Efficacy Analyses

Summary statistics by visit will be provided for all efficacy endpoints. For continuous efficacy endpoints, summary statistics will also be provided for change from baseline at each post-baseline visit. Results for the primary and secondary efficacy endpoints will be presented separately for mild and moderate AD as well as combined across mild and moderate AD strata. Models will include a baseline covariate for mild/moderate for all primary and secondary efficacy analyses. Models will be performed for the primary efficacy endpoint for the separate AD strata but not for the secondary efficacy endpoints.

2.2.12.1 Primary Endpoint

The primary efficacy variable is the change from baseline in Numeric Rating Scale (NRS) in the DS107 dose treated populations compared to vehicle population at week 8/exit.

The change from baseline in NRS will be analyzed for the FAS with SAS PROC MIXED using code similar to the following, with CFB_NRS and B_NRS representing the change from baseline and the baseline NRS, respectively.

```
proc mixed data= FAS;
class trt  AD visit;
model CFB_NRS= trt AD B_ NRS/CL alpha=0.052;
lsmeans trt*visit / diff CL;
run;
```

where fas= Full Analysis Set (FAS). The cl and alpha= option specifies that two-sided exact 97.4% confidence intervals will be provided. The between-treatment effects are determined from the diff option of the LSMEANS statement. Results will be presented separately for mild and moderate AD. The primary analysis will be based on the FAS and repeated for the PPS as a supportive analysis.

Co-Primary Endpoint

Change from baseline in Eczema Area and Severity Index (EASI) in the DS107 dose treated population compared to vehicle population at Week

8/exit. The change from baseline in EASI will be analyzed the same way as the primary endpoint.

The study drug is to be considered effective when both of the primary endpoint and co-primary endpoint demonstrate an effect, i.e. reject the null hypothesis using the separate comparisons vs. vehicle control.

2.2.12.2 Secondary Endpoints

The secondary endpoints are:

- Change from baseline in Numeric Rating Scale (NRS) in treated population compared to vehicle population at Weeks 2, 4, 6, 10 and subject's last treatment. These will be analyzed using PROC MIXED as described in the primary endpoints section. Three separate models will be done using the same PROC MIXED code, one for the change from baseline to Week 2, 4 and 6, one for the change from baseline to Week 10 and the third for the change from baseline to last treatment visit.

As NRS is a daily assessment, the analysis of NRS will be performed using the daily average at each visit. At each visit, the average of the NRS values since the previous visit will be computed. Missing values for any day will be ignored and the between-visit daily average will be computed using non-missing values only. The baseline value will be the average computed during the screening period. This analysis of change from baseline in NRS will be performed the same as the analysis of change from baseline in EASI, with CFB_NRS and B_NRS representing the change from baseline in NRS and the baseline NRS, respectively.

- Proportion of patients achieving at a decrease of at least 2.7 points in NRS in treated population compared to vehicle population from baseline to Week 2, 4, 6, 8, 10, and last treatment visit. These will be analyzed using PROC GENMOD using code similar to the following, with Outcome and B_NRS representing the outcome (whether or not achieved a decrease of at least 2 points) and the baseline NRS, respectively.

```
PROC GENMOD data=FAS descending;
      Class trt AD visit;
      Outcome= trt AD B_NRS/dist=bin cl alpha=0.052;
      Repeated subject=trt*visit/corr=un;
      slice trt*visit / sliceby=visit ilink cl diff ;
run;
```

Three separate models will be done using the same PROC GENMOD code, one for the change from baseline to Week 2, 4, 6 and 8, one for the change from baseline to Week 10 and the third for the change from baseline to last treatment visit.

- Change from baseline in Eczema Area and Severity Index (EASI) in treated population compared to vehicle population at Weeks 2, 4, 6 and 10 and subject's last treatment visit. These will be analyzed using the similar PROC MIXED code as described in the primary endpoint section. Three separate models will be done using the same PROC MIXED code for the change from baseline to Week 2, 4 and 6, for the change from baseline to Week 10, and for the change from baseline to last treatment visit.

- Proportion of patients achieving an IGA score of 0 (clear) or 1 (almost clear) and a decrease of at least 2 points in IGA in treated population compared to vehicle population from baseline to Week 2, 4, 6, 8 and 10 and change from baseline to subject's last treatment visit. These analyses will be performed using the same PROC GENMOD code described above. Three separate models will be done for the change from baseline to Week 2, 4, 6 and 8, for the change from baseline to Week 10, and for the change from baseline to last treatment visit.

• Proportion of patients achieving a decrease of at least 2 points in IGA in treated population compared to vehicle population from baseline to Week 2, 4, 6, 8, 10, and change from baseline to subject's last treatment visit. These analyses will be performed using the same PROC GENMOD code described above. Three separate models will be done using the same PROC GENMOD code for the change from baseline to Week 2, 4, 6 and 8, for the change from baseline to Week 10 and for the change from baseline to last treatment visit.

• Change from baseline IGA in treated population compared to vehicle population at Weeks 2, 4, 6, 8 and 10 and subject's last treatment visit. These will be analyzed using the same PROC MIXED using code similar to the following, with CFB_IGA and B_IGA representing the change from baseline and the baseline IGA, respectively. Method=ml is the maximum likelihood estimate.

```
proc mixed method =ml data= FAS;
class trt visit AD;
model CFB_IGA= trt AD B_ IGA/CL alpha=0.052;
lsmeans trt*visit / diff CL;
run;
```

Three separate models will be done using the same PROC MIXED code for the change from baseline to Week 2, 4, 6 and 8, for the change from baseline to Week 10 and for the change from baseline to last treatment visit.

The method for controlling the Type I error rate is provided in Section 2.2.12.4.

2.2.12.3 Exploratory Endpoints

• Time to response by Week 8. A Kaplan-Meier lifetable will be constructed using nominal visits. A Wilcoxon-Gehan test will be used to compare all treatments; p-values will also be computed for each of the three treatment pairs for the primary endpoint using the FAS population. This will be a graph.

• Probability of being in response (PBIR) between Weeks 2 and 8. The PBIR will be computed as a descriptive statistic for the primary endpoint using the FAS population. Each visit will be equally weighted in this analysis. This will be a graph.

• Sustained efficacy between Weeks 8 and 10. This analysis will be endpoint-specific. Binary endpoints will be evaluated using logistic regression (PROC LOGISTIC) controlling for baseline IGA score and baseline AD severity while continuous endpoints will be analyzed using PROC GLM controlling for the same factors. A shift table will be created presenting the number and frequencies of subjects in response and the shift from Week 8 to Week 10 for the primary endpoint analysis for both the FAS and PPS populations.

• Change from baseline in the Dermatology Life Quality Index (DLQI) score in treated population compared to vehicle population at Week 2, 4, 6, 8 and 10 and from Week 8 to Week 10 and change from baseline to subject's last treatment visit. These analyses will be done using PROC MIXED similar to the code described above.

DLQI is a set of 10 questions, scored 0=Not at all, 1=A little, 2=A lot, or 3=Very much, or 0=Not relevant, except for Questions 1 and 2 that do not include a response of 'Not Relevant'. Missing responses are assigned '0'. The DLQI total score is calculated by summing the score of each question resulting in a maximum of 30 and a minimum of 0. The higher the score, the more quality of life is impaired. This scoring will be done by the data management group and provided to CTDS in the SDTM datasets.

Mean DLQI will be computed as the Total DLQI divided by the number of questions with non-missing responses. The scoring instructions for DLQI specify that if 2 or more questions are not answered, then the questionnaire will not be scored; this approach is not taken here with Mean DLQI computed instead. Summary statistics for Total DLQI and Mean DLQI will be presented. The analysis of change from baseline in Mean DLQI will be performed the same as the analysis of change from baseline in EASI, with appropriate changes in the variables for change from baseline and Mean DLQI at baseline.

Summary statistics as continuous data and change from baseline will be presented by treatment group and visit for the totals for the 4 sub-scales:

Symptoms and feelings	Questions 1 and 2
Daily Activities	Questions 3 and 4
Leisure	Questions 5 and 6
Personal relationships	Questions 8 and 9

• Change from baseline in the Patient Orientated Eczema Measure (POEM) score in treated population compared to vehicle population at Week 2, 4, 6, 8 and 10 and from Week 8 to Week 10 and change from baseline to subject's last treatment visit. These analyses will be done using PROC MIXED similar to the code described above.

POEM is a set of 7 questions regarding the frequency of conditions related to eczema, with responses of 0 = no days. 1 = 1-2 days, 2 = 3-4 days, 3 = 5-6 days, or 4 = every day. The sum of the scores for the 7 questions is the Total POEM.

Mean POEM is the Total POEM divided by the number of questions with non-missing responses. The scoring instructions for POEM specify that if response to one question is missing, the response for that question will be equated to 0, and if the responses to more than one question are missing, then the questionnaire is not scored; this approach is not taken here with Mean POEM for questions with non-missing responses computed instead. The analysis of change from baseline in Mean POEM will be performed the same as the analysis of change from baseline in EASI, with appropriate changes in the variables for change from baseline and Mean POEM at baseline.

Summary statistics (counts and frequencies) will be presented by treatment group and visit for categories based on Total POEM of:

Clear or almost clear = 0 to 2;
Mild eczema = 3 to 7;
Moderate eczema = 8 to 16;
Severe eczema = 17 to 24;
Very severe = 25 to 28.

Shift tables presenting the number and frequency of subjects in each eczema severity category at Week 8 and last treatment visit and Week 8 to Week 10 will be presented by treatment group.

• Change from baseline in the Patient Global Impression of Severity (PGI-S) score in treated population compared to vehicle population at Week 2, 4, 6, 8 and 10 and from Week 8 to Week 10 and change from baseline to subject's last treatment visit. These analyses will be done using PROC MIXED similar to the code described above.

The categories for the PGI-S are: normal, mild, moderate or severe. A Shift table will be created presenting the number and frequencies of

subjects in each category and the shift from the follow-up visit to baseline.

• Change from baseline in the Patient Global Impression of Change (PGI-C) score in treated population compared to vehicle population at Week 2, 4, 6, 8 and 10 and from Week 8 to Week 10 and change from baseline to subject's last treatment visit. PGI-C is captured at the follow-up visits and is a 7 –point scale assessing the overall status of the patient since the start of the study with responses of 1=very much improved, 2=much improved, 3=minimally improved, 4= no change, 5=minimally worse, 6= much worse and 7=very much worse. Summary statistics (count and frequencies) will be presented by treatment group and visit for the status categories. A shift table presenting the number and frequency of subjects in each category at Week 8 to Week 10 will be presented by treatment group.

• Determination of AD biomarkers in treated population compared to vehicle population at Baseline/Day 0 and Week8/Early Termination (samples to be retained for the potential analysis at a later date).
Details of these analyses will be described in a separate document and the analyses will not be included with the analyses described herein.
The method for controlling the Type I error rate is provided in Section 2.2.12.4.

2.2.12.4 Controlling Type I error

Dunnett's test will be used to provide for simultaneous comparisons of each DS107 dose vs. vehicle by allowing two-sided p=0.026 per dose comparison.

No multiplicity adjustment is necessary for the primary and co-primary endpoints as the Type I error rate will not be affected because both must be met for study success. A hierarchical testing sequence methodology will be applied to the second and exploratory endpoints in order to control the overall type 1 error. Table 2 provides the order of testing for each active dose group compared to the vehicle group for the secondary and exploratory efficacy endpoints; these will each be global tests across all post-baseline outcomes excluding Week 10. Each of these tests will be at the 0.05 significance level. A test will be considered statistically significant only if the corresponding test has a p-value <0.05 and all previous tests (as determined by the testing order defined in Table 2) have a p-value <0.05.

Table 2. Hierarchical Testing Sequence for Secondary and Exploratory Endpoints

Testing Order	Endpoint	Efficacy Endpoint
1	Secondary	Change in NRS from baseline
2	Secondary	Proportion achieving a reduction of at least 2.7 points in NRS
3	Secondary	Change in EASI from baseline
4	Secondary	Proportion achieving an IGA score of 0 or 1 and a decrease of at least 2 points in IGA
5	Secondary	Proportion achieving a decrease of at least 2 points in IGA
6	Secondary	Change in IGA from baseline
7	Exploratory	Change in DLQI from baseline
8	Exploratory	Change in POEM from baseline
9	Exploratory	Change in PGI-S from baseline
10	Exploratory	Change in PGI-C from baseline

2.2.13 Safety Analyses

All subjects enrolled in the study who received at least one dose of the medication will be included in the safety analysis (i.e. Safety Analysis Set).

2.2.13.1 Extent of Exposure

Treatment duration is defined as 8 weeks (Day 56 ± 2 days). The date of the first and last applications of study medication will be collected from information recorded on the study drug admin page of the eCRFs. Study medication is to be applied twice daily.

Extent of exposure to study medication will be tabulated by treatment group by 2-week intervals. The number of applications of study medication will be summarized by treatment group by each 2-week interval only for those subjects who returned to the clinic for each interval. The expected number of applications per 2-week intervals is 14, based on twice daily dosing. Compliance will be measured by calculating the number of applications/expected number of applications *100 and presented by treatment group for each interval.

2.2.13.2 Safety Endpoint(s)

The safety objective is to compare the safety of topically applied DS107 cream (1% & 5%) versus vehicle, in the treatment of adult patients with mild to moderate AD. Results will be pooled for 1% and 5% as well.

2.2.13.3 Adverse Events

Counts and percentages of subjects who experienced any of the following will be presented by treatment group, as well as a total DS107 treatment group that includes both the 1% and 5% DS107 subjects: any AE, any IMP related AE, any AE requiring modification of study medication dosing (dose reduced, drug interrupted), discontinuations of study participation due to an AE (drug withdrawn), any serious adverse events (SAE), deaths, or IMP related SAEs. Only those adverse events determined to be treatment emergent (TEAE) will be presented (where Qlabel ="Treatment Emergent" from the SUPPAE SDTM dataset). Treatment emergent adverse events are those events with the date of AE onset being greater than or equal to the date of the first application date of study drug. A two-sided Fisher Exact test will be utilized for an overall global test and for the pair-wise comparisons of each treatment compared to vehicle.

Verbatim terms on the case report forms will be coded to preferred term (PT) and related system organ class (SOC) using MedDRA Coding Dictionary Version 19.1. The number and percentage of subjects who experience treatment-emergent adverse events (TEAE) will be tabulated by System Organ Class (SOC) and Preferred Term (PT) within each treatment group. The number of events will be tabulated by System Organ Class (SOC) and Preferred Term (PT) within each treatment group. In addition, a tabulation of adverse events that occur at a frequency of 2% or greater at the SOC or PT level will be prepared. A two-sided Fisher Exact test will be performed on system organ class (SOC) with a 2% or greater frequency to facilitate evaluation of potential safety differences among treatments.

Adverse events will be counted only once for a subject within each Preferred Term and System Organ Class; thus, since a subject may have more than one Preferred Term within an SOC, percentages of PT may not sum to the percentage in the SOC. If a Preferred Term is reported multiple times with differing severities (mild, moderate or severe), only the most severe is counted but all events will be counted. If a Preferred Term is reported multiple times with differing relationships to IMP, only the one that is related to IMP is counted.

2.2.13.3.1 Serious Adverse Events

The number and percentage of subjects experiencing a Serious Adverse Event (SAE) will be tabulated by SOC and PT within each treatment group and overall. SAEs will be counted only once for a subject within each PT and SOC.

2.2.13.3.2 Deaths

A listing of subjects who die while on study will be prepared along with the adverse event associated with the cause of death.

2.2.13.3.3 Interruptions or Discontinuations of Study Medication Due to an AE

A listing of subjects who have changes to their study medication dosing (dose reduced, interruptions or discontinuations) will be prepared.

2.2.13.4 Laboratory Data

Any laboratory test results considered to be clinically significant by the Investigator will be captured as an adverse event. All clinical laboratory values (haematology, biochemistry and urinalysis) will not be tabulated but will be listed for each patient by treatment group and visit. Values outside the laboratory normal ranges will be listed separately by patient and treatment group with associated comments as to their clinical significance.

2.2.13.5 Vital Signs, Physical Findings, and Other Observations Related to Safety

Vital signs are collected at the screening, visit 2/baseline and the follow-up visits. Temperature (^{o}C), systolic and diastolic blood pressure (mmHg) and heart rate (BPM) will be tabulated by treatment group using the baseline value (where baseline record flag= 'Y') and end of study (week 10). Change from end of study to baseline will be presented as descriptive statistics.

Physical Examination Findings will be presented in the listings but not tabulated.

Concomitant Medications will be coded per the WHO Drug dictionary (version effective March 2017 and provided to CTDS in the SDTM CM dataset. Concomitant Medications will be tabulated by treatment group,

Drug Class (pharmacological level, ATC3) and Drug Name (chemical substance level, ATC5). These data will be provided in subject data listings along with the verbatim drug term and usage details.

3 Document Version Control

Revision History:

REVISION	RELEASE DATE	AUTHOR	SUMMARY OF CHANGES
A	17Jul2017	CTDS/P. Lavin	Initial Version
B	11Oct2017	CTDS/P. Lavin	Updated SAP to reflect updated protocol version 5.0

Appendix A - Programming Specifications for Tables and Listings

The following specifications will be used in the production of tables and listings.

1. Page Setup

   Unless otherwise noted, tables and listings will use landscape orientation. Margins will be at least 1.5 inches on the bound side and at least 1 inch on the other three sides.
   The following header information should be included:
   - Upper left: Sponsor name and protocol number
   - Center: CONFIDENTIAL
   - Upper right: Page number shown as Page n of N. Page numbers should be sequential within a table or listing.

   The footer should include:
   - The name of the SAS program used to generate the output along with the run date/time and the words "by CTDS".
   - For tables, the corresponding listing number(s).

2. Footnotes

   Unless otherwise specified, footnotes should appear on all pages within the table.

3. Font

   Font will be 8-point Arial, or smaller if needed for space constraints. If possible, small tables should appear on one page. If tables continue on to multiple pages, there should be a page break after an assessment so that all the statistics for an assessment appear on the same page.

4. Tables

   Table titles should reflect the content of the table. Under the main title, in parentheses, the name of the analysis population being summarized should appear.

   4.1 Summary Statistics - Continuous Data

   Unless otherwise noted, the mean and median and confidence interval (CI) of a set of values should be printed out to one decimal place more than the original value. The standard deviation should be printed out to 2 decimal places more than the original value. The number of subjects on whom the parameter is assessed should appear. Minimum and maximum should be consistent with the original value. P-values will be

expressed as 4 decimal places. Any p-values that get calculated as 0.0000 should be expressed as p-value <0.0001.

4.2 Summary Statistics - Categorical Data

Numbers of subjects are reported as whole numbers. Null counts are represented as 0. Table percentages should be reported to one decimal unless otherwise noted. Null percentages should be reported as 0.0%.

For all categories, the total number of subjects with data will be presented as N and the number of subjects with non-missing data will be used as the denominator for the calculation of the percentages, unless otherwise noted.

5. Subjects Included in Listings

In general, subject data listings should include all subjects who signed the informed consent. If a listing includes a subset of subjects who meet a certain condition (e.g, subjects with SAEs) then this should be clear from the title of the listing. If there are no subjects who meet the condition (e.g. no subjects with SAEs) for either the tables or the listings, then a page marker should appear stating that no subjects met the criteria.