Protocol No. MC2-01-C2
Novella Study No. NYA13480

STATISTICAL ANALYSIS PLAN

A Randomised, Multicentre, Investigator-Blind, Parallel-Group Trial to Evaluate the Efficacy and Safety of MC2-01 Cream Compared to MC2-01 Cream Vehicle and Active Comparator in Subjects with Mild-to-Moderate Psoriasis Vulgaris

Novella Clinical
365 West Passaic Street
Rochelle Park, NJ 07662, USA

05 July 2018

Johan Selmer, MD
VP Medical Affairs
MC2 Therapeutics

Signature _____ **Date** _____

Carol Udell, MS
Sr. Director, Clinical Reporting
Novella Clinical

Signature _____ **Date** _____

Table of Contents

List of Abbreviations and Definitions of Terms

Abbreviation	Definition
AE	Adverse event
ADaM	Analysis data model
ANOVA	Analysis of variance
AUC	Area under the curve
BDP	Betamethasone dipropionate
BOCF	Baseline observation carried forward
BSA	Body surface area
CAL	Calcipotriene
CI	Confidence interval
DLQI	Dermatology Life Quality Index
ECG	Electrocardiogram
EQ-5D	EuroQOL five dimensions questionnaire
EQ-VAS	EQ-5D visual analog scale
IP	Investigational product
IRB	Institutional Review Board
ITT	Intent to treat
LOCF	Last-observation-carried forward
MCMC	Markov Chain Monte Carlo
MI	Multiple imputation
mPASI	Modified Psoriasis Area and Severity Index
MedDRA	Medical Dictionary for Regulatory Activities
NI	Non-inferiority
NRI	Non-responder imputation
ODS	SAS Output Delivery System
PASI	Psoriasis Area and Severity Index
PASI 75	75% reduction in mPASI
PASI 50	50% reduction in mPASI
PGA	Physician's global assessment
PP	Per protocol
PRO	Patient-reported outcome
PT	Preferred Term
QOL	Quality of Life
RTF	Rich-text-formatted

Abbreviation	Definition
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SD	Standard deviation
SDTM	Study data tabulation model
SE	Standard error
SGA	Subject's global assessment
SUSAR	Suspected unexpected serious adverse reaction
TEAE	Treatment emergent adverse event
TESAE	Treatment emergent serious adverse event
TC	Telephone call
UPT	Urine pregnancy test

1. INTRODUCTION

This Statistical Analysis Plan (SAP) is based on the study Protocol version 3.0 dated February 27, 2018.

This document provides additional details concerning the statistical analyses outlined in the protocol and reflects any changes to the protocol from any amendments. This plan will not repeat all the definitions given in the protocol but will provide further details of the summaries and analyses planned therein.

Additional analyses will be made as part of the validation of the Psoriasis Treatment Convenience Scale. These analyses will be described and reported separately.

2. STUDY OBJECTIVES

The primary objective is to show therapeutic non-inferiority of MC2-01 cream to active comparator, as well as to characterise the safety profile of MC2-01 cream in subjects with psoriasis vulgaris.

3. STUDY DESIGN

This is a Phase 3, randomised, investigator-blind, multicentre, vehicle and comparator controlled, parallel-group, 3-arm trial designed to show therapeutic non-inferiority of MC2-01 (CAL and BDP) cream to active comparator in subjects with mild-to-moderate psoriasis vulgaris. The primary efficacy endpoint is the proportion of subjects in each treatment group with 'treatment success' at Week 8, defined as a minimum 2-point decrease from Baseline to Week 8 on the Physician's Global Assessment (PGA) of disease severity on the trunk and limbs; i.e. 0 (clear) or 1 (almost clear) disease for subjects with moderate disease at Baseline; or a score of 0 (clear) for subjects with, a score of mild disease at Baseline.

A total of 791 subjects are planned to be enrolled in order to have approximately 712 completed subjects. Qualified subjects will be randomised into MC2-01 cream, MC2-01 cream vehicle and active comparator in a 3:1:3 ratio stratifying by site and baseline severity. Subjects are healthy males or non-pregnant females, at least 18 years of age, with a clinical diagnosis of plaque psoriasis (psoriasis vulgaris) of at least 6 months duration that involves non-scalp regions of the body (trunk and/or limbs), with a PGA of disease severity of mild or moderate on the body (trunk and/or limbs) and a minimum modified Psoriasis Area and Severity Index (mPASI) score of at least 2 in at least one body region (i.e., psoriasis affecting at least 10% of trunk and/or limbs).

The maximum trial duration for each subject will be approximately 14 weeks and includes a screening period of up to 4 weeks (if washout of prohibited medications is required), an 8-week treatment period, and a follow-up period of 2 weeks. The study will include a total of 8 scheduled visits: a screening visit, a baseline visit, four on treatment visits (Visit 3 telephone call), end of

treatment visit, and a follow-up visit. Unscheduled visits may be performed at any time during the trial if judged necessary by the Investigator, such as for a severe reaction and clinically significant AE. Details of the event will be recorded in the subject's records.

During the treatment period, the investigator will score the disease severity of the subject using the PGA, modified Psoriasis Area and Severity Index (mPASI) and body surface area (BSA) involvement; and the subject will perform a disease assessment using subject's global assessment (SGA). The subjects will be asked to complete the Dermatology Life Quality Index (DLQI), EuroQOL five dimensions (EQ-5D), and the Psoriasis Treatment Convenience Scale questionnaires. Safety assessments (local skin reaction, AEs, laboratory tests, electrocardiogram (ECG), vital signs and physical examination) will be performed.

4. HARDWARE AND SOFTWARE

Statistical analysis will be performed following Novella Clinical standard operating procedures (SOP) and on the Novella Clinical computer network. All statistical analysis will be performed using SAS Version 9.3 or higher with program code prepared specifically for the project by qualified Novella Clinical statisticians and SAS programmers.

The SAS programs will generate rich-text-formatted (RTF) output with the "RTF" extension using the SAS Output Delivery System (ODS). The summary tables and listings will be formatted using the Times New Roman 9-point font. The RTF output is included in report documents prepared with Microsoft Word and converted to PDF format without typographical change.

Datasets will be created and taken as input to validated SAS programs to generate the report-ready tables, listings, and figures. Each output display will show the names of the data sets and SAS program used to produce it.

5. DATABASE CLOSURE

After completion of all data review procedures, validation of the project database, and approval of the data review document by the study sponsor, the clinical database will be closed. Any change to the clinical database after this time will require written authorization, with explanation, by the Sponsor and the Biostatistician.

6. SAMPLE SIZE DETERMINATION

The primary endpoint is the proportion of subjects who achieve a minimum 2-point decrease from Baseline in PGA at Week 8. The primary endpoint will be used for the non-inferiority evaluation.

The sample size calculation is based on the non-inferiority margin for the difference in PGA response.

<u>Non-inferiority margin:</u>

The non-inferiority margin was determined according to the draft Food and Drug Administration (FDA)-Guidance for Industry: Non-inferiority Clinical Trials as a fixed margin.

A summary of responders, based on the PGA, at Week 8 from three randomised trials of calcipotriene and betamethasone dipropionate gel is shown below to help determine the statistical margin M1 and the clinical margin M2. The clinical margin M2 is used in the determination of sample size.

Table 1. Determination of sample size

	Daivobet		Vehicle		Daivobet – Vehicle	
Trial	**N**	**PGA success %**	**N**	**PGA success %**	**%**	**99% CI**
Menter et al, 2013	482	29.0	95	6.3	22.7	(14.4, 31.1)
Langley et al, 2011	183	39.9	91	5.5	34.4	(23.2, 45.6)
Fleming et al, 2010	162	27.2	40	0	27.2	(18.2, 36.2)
Total					26.2	(19.9, 32.6)

The lower 99%-confidence limit of the difference in response rate between calcipotriene/betamethasone dipropionate gel and vehicle was calculated to be 19.9%. Therefore, the prespecified M1 margin was determined to be 19.9%, a conservative choice for the active control effect size. The largest loss of effect that would be clinically acceptable (M2) is set to 10% points.

Assumption's for the non-inferiority comparison

- The lower non-inferiority boundary for M1 is set to 19.9% for the difference between vehicle and active comparator in response rates according to PGA
- The lower non-inferiority boundary for M2 is set to 10% points for the difference between MC2-01 cream and active comparator in response rates according to PGA.

Assumptions for sample size calculation:

- The response of active comparator is assumed to be 30%.
- The absolute difference of MC2-01 cream and active comparator is assumed to favour MC2-01 cream by at least 2.5 percentage points. The treatment difference between MC2-01 cream and active comparator was estimated based on preliminary results from a plaque psoriasis trial that showed slightly better efficacy for MC2-01 cream compared with active comparator.
- The power should be at least 90%.
- The error probability is set to 0.05 for a 2-sided test.

Under these assumptions, a sample size of N=305 per active treatment group is calculated. Assuming 90% completion rate of the randomised subjects, a total sample size of N=339 per active treatment group is required to be randomised.

The comparison to vehicle to show superiority of MC2-01 cream and active comparator will need N=103 per treatment group under the assumption of a vehicle responder-rate of 11%, a response rate of at least 30% for the MC2-01 cream and active comparator, and 90% power. Following a 3:1:3 randomisation ratio, 113 subjects will be randomised to receive the MC2-01 vehicle cream (assuming a 90% completion rate).

Therefore a total enrolment of approximately 791 subjects is planned to have approximately 712 completed subjects.

7. ANALYSIS POPULATION

Three analysis populations will be defined for analysis:

- Intent-To-Treat (ITT) population: will include all subjects who were randomised and dispensed the trial medication at Randomisation. Subjects who return all the trial medication unused will be excluded.
- Per Protocol (PP) population: will include subjects in the ITT population who complete the trial without any major protocol violations. The composition of the PP population will be determined and documented in blind reviews of the database conducted prior to un-blinding the trial database. Subjects may be excluded from PP population if any of the following criteria are met.
    - Failure to meet key Inclusion/Exclusion criteria; I.e. violation of inclusion/exclusion criteria that according to medical judgement may impact the primary endpoint analysis;
    - Usage of restricted medications/treatments; I.e. medications/treatments that according to medical judgement may impact the primary endpoint analysis;
    - Nonadherence to the visit schedule at Week 8 ± 7days;
    - Noncompliance with the trial treatment regimen; adherence to the treatment schedule is defined as application of IP as per subject diary for at least 80% of the days from Week 4 to Week 8;
    - Noncompliance with the trial treatment regimen; adherence to the treatment schedule is defined as application of IP as per subject diary for at least 80% of the days from Day 1 to Week 8;

      In particular, subject treatment days that are approved by the investigator for study drug stoppage due to clear skin, i.e., those marked as N/A in subject diary, will be included in the computation of compliance as if subjects were compliant these days.

- Safety Population: will include all subjects who were randomised and dispensed the trial medication at Randomisation. Subjects who return all the trial medication unused will be excluded.

The ITT and PP Population will be used for the analyses of the primary efficacy endpoint. The PP population will be considered as the primary population for analyses of therapeutic non-inferiority of MC2-01vs active comparator. The ITT analysis will be considered as the primary population for the analysis of superiority of MC2-01 and active comparator vs. vehicle. A similar approach will be used for the secondary efficacy endpoints. The Safety Population will be used for the analyses of safety endpoints.

All efficacy analyses will be conducted according to the randomised treatment assignment; all safety analyses will be conducted according to the treatment actually received.

8. ESTIMANDS

When analyzing the confirmatory endpoints aiming to demonstrate superiority, the primary analysis will be focusing on a treatment-policy estimand. An additional hypothetical-strategy estimand will be used to assess external validity for the primary and the key secondary endpoint. The primary analyses of non-inferiority will be made based on the PP-population, using only observed cases. To assess external validity, also the analysis of non-inferiority will be made based on the two estimands, treatment-policy and hypothetical-strategy. They are described in further detail below.

Treatment-policy estimand: the effect of initiating the treatment regimen in subjects with mild-to-moderate psoriasis, with or without any other psoriasis treatment, and whether or not the subject adhered to the treatment regimen.

Thus, for the treatment-policy estimand all available data will be used. Missing data will be imputed

Hypothetical-strategy estimand: the expected effect in subjects with mild-to-moderate psoriasis if all subjects adhered to the treatment regimen and did not take any other treatment that according to medical judgement may impact the primary efficacy endpoint.

This is interpreted in the sense that subjects did not initiate other treatment that according to medical judgement may improve psoriasis, and for visits after baseline adhered to the treatment schedule defined as application of IP as specified by the investigator at the previous site visit for at least 70% of the days from the Baseline visit to Week 4, Week 4 to Week 6 and Week 6 to Week 8. For example, applying IP for 10 out of 14 days between Week 4 and Week 6 is considered adherence, while 9 out of 14 days is not. Computations to determine adherence are

done using actual visit dates and compliance data from the compliance page in the CRF. For the hypothetical-strategy estimand, only data at or prior to use of other treatment that may improve psoriasis according to medical judgement, and prior to non-adherence to treatment schedule is used. Missing data, data after use of other psoriasis treatment, and data after non-adherence is imputed.

In addition, estimands corresponding to a composite-strategy estimand and a while-on-treatment estimand will also be applied.

9. HANDLING OF MISSING DATA

For superiority analyses based on the ITT set, a variety of methods will be used to impute missing data, including multiple imputation (MI), last-observation-carried forward (LOCF), baseline observation carried forward (BOCF) and non-responder imputation (NRI). MI will be the primary imputation method and sensitivity analyses using LOCF, BOCF and NRI will be performed to assess the robustness of imputation assumptions. The primary non-inferiority evaluation will be on the PP population using an Observed Cases (OC) analysis, i.e., there will be no imputation for missing data at any time point.

The imputation procedures for post-baseline missing data are described below.

- MI: Multiple imputations is a simulation based approach where missing values are replaced by multiple Bayesian draws from the conditional distribution of missing data given the observed data and covariates, creating multiple completed data sets. These completed datasets can then be analysed using standard analysis methods. Rubin (1987) presented rules for how to combine the multiple sets of estimates to produce overall estimates, confidence intervals, and tests that adequately incorporate missing data uncertainty.

  Missing values for PGA data will be imputed using the SAS MI procedure. Intermittent missing values of PGA will be imputed separately for each treatment group using the Markov Chain Monte Carlo (MCMC) method. 100 copies of the dataset with monotonic missing pattern will be generated. Assuming normality for the PGA score is motivated by Schaefer (1997), where it was shown that the multivariate normal approximation for the imputation of incomplete categorical and binary data is robust.

  For each of the 100 datasets, missing values at scheduled visits will then be imputed separately for each treatment group using a monotone regression method that includes: prior systemic biologic use for psoriasis (yes/no) and baseline PGA severity (mild/moderate) as categorical variables; duration (years) of psoriasis (calculated as year of baseline minus year of first psoriasis diagnosis) and PGA scores at weeks 1, 4, 6, and 8 as continuous variables.

The primary efficacy endpoint will be calculated from these imputed datasets. Note that imputed PGA scores can take fractional values (eg, 1.4, 2.2, etc.). In order to account for this, PGA score will be rounded to the nearest positive integer number and negative values will assume to be 0. PGA Treatment Success status (Yes/No) will then be derived based on the imputed datasets.

Missing values for secondary efficacy endpoints, mPASI and itch by numerical rating scale (NRS), will be imputed by MI using similar method as described above for the primary efficacy endpoint, with replacement of PGA by each respective endpoint in the imputation models.

A pre-specified seed number of 213213 will be used in all imputation procedures as described above.

This method will be used to impute missing data corresponding to the treatment-policy estimand as well as the hypothetical-strategy estimand.

- LOCF: The last observed value will be carried forward for any subsequent missing values. Baseline values will not be carried forward.
- BOCF: The baseline value will be used for any missing post-baseline values.
- NRI: Any missing value will be assumed to be a 'non-responder' with respect to the given analysis variable. This method corresponds to a composite strategy estimand.
- Imputation of missing or invalid Psoriasis Treatment Convenience Scale (PTCS) score

  PTCS score will be calculated as sum of questions 1 to 5. The score is to be calculated if at most two of the questions have missing answers, in which case the missing answers will be imputed from the mean of the answered questions. If more than two questions have missing answers, the PTCS score will be considered as missing. A PTCS score is defined as valid, if subjects have used the study medication at some point within 7 days prior to the day of the assessment. A PTCS score is considered as invalid if subjects have discontinued study medication for more than 7 days prior to the day of assessment. Invalid measurements will not be applied in the analysis of that visit. Missing and invalid PTCS scores will be imputed using the last valid measure prior to the visit. This approach corresponds to a 'PCTS while on treatment' estimand.

10. DATA CONVENTIONS FOR ANALYSIS

9.1 General Statistical Principles

All statistical processing will be performed using the SAS system (Version 9.3 or higher).

All observed and derived variables (e.g., change from baseline) used in the summaries of analyses will be presented in by-treatment and by-subject listings. Descriptive statistics will be used to

provide an overview of the efficacy and safety results. For categorical parameters, the number and percentage of subjects in each category will be presented. The denominator for percentage will be based on the number of subjects appropriate for the purpose of analysis. For continuous parameters, descriptive statistics will include number of observations, mean, standard deviation (SD), median, Q1, Q3, and range.

Except where noted, all statistical tests will be 2-sided and will be performed at the 0.05 level of significance.

No interim analyses are planned.

9.2 Study Day

For purpose of the SAP, Day 1 is defined as the date of first application of study drug, which corresponds to Day 0 of the study protocol. Study day is calculated relative to the date of Day 1.

9.3 Baseline, Change from Baseline and Reduction from Baseline

Baseline is defined as the last available non-missing value prior to first application of study drug. Change from baseline will be defined as the post-baseline value minus the baseline value, unless otherwise stated. Reduction from baseline will be defined as the baseline value minus the post-baseline value, unless otherwise stated. Percent reduction from baseline is calculated as the baseline value minus post-baseline value divided by the baseline value, expressed as a percentage, unless otherwise stated.

9.4 Analysis Visit Windows

Efficacy and safety endpoints will be analysed according to their windowed visits defined by actual study day. If more than one valid evaluation occurs within a single visit window, then the analysis will take the one closest to the target day. If the two visits are equidistant from the target day, the later will be used. In particular, for safety assessments with repeats, the scheduled (non-repeat) values will be included for summary.

The following analysis visit windows will apply:

Visit	Week	Target Day	Assessment Window
2	1	8	Post-dose – Day 11
4	4	29	23 — 36
5	6	43	37 — 50
6	8	57	51 — 64
7 (Follow-up)	10	71	Nominal (Post last-dose)

9.5 Analysis Sites

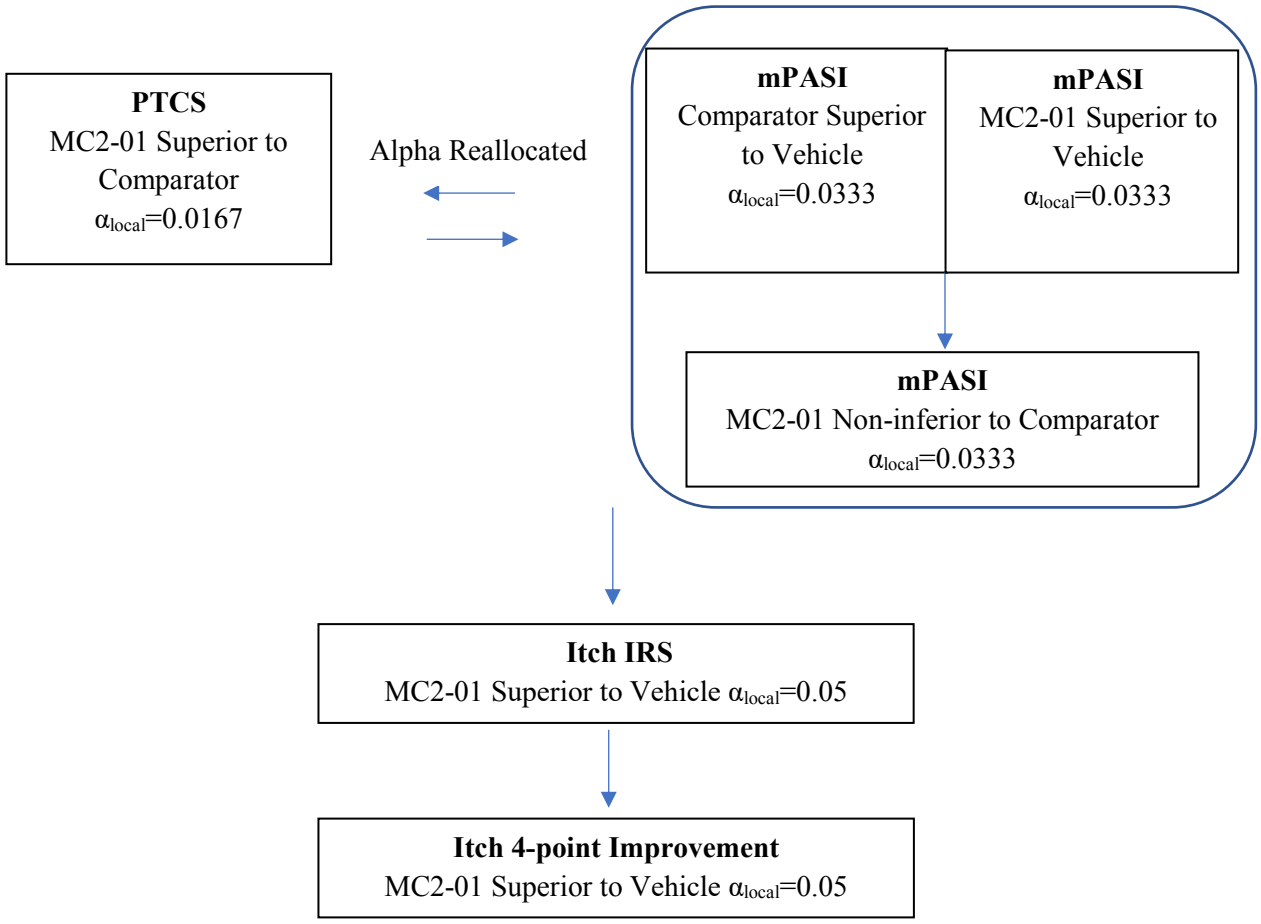
There will be approximately 57 trial sites in the United States. The goal is to randomise at least 14 subjects at each site. Sites that do not randomise at least 14 subjects will be combined in order of geographical proximity. The exact composition of these "analysis sites" will be determined and documented prior to breaking the trial blind. Statistical test investigating the statistical homogeneity of the primary outcome across "analysis sites" will be performed. Further detail is provided in section 10.9.1.

9.6 Multiple Comparisons

The global Type-I error for the secondary endpoints will be controlled by the following testing procedure:

1. Secondary endpoints will be tested only if the primary endpoint will reject the Null-hypothesis;
2. The two secondary endpoints, percentage reduction from baseline in the mPASI score at Week 8, and subject assessment of PTCS at Week 8 between MC2-01 cream and active comparator will be tested in a hierarchical way with a loop back procedure with initial error probability weights 2/3 for H1 (change from baseline in mPASI) and 1/3 for H2 (subject assessment of treatment convenience), resulting in $\alpha 1=0.0333$ and $\alpha 2=0.0167$. If any of the null-hypotheses can be rejected, the respective error probability will be shuffled to the other hypothesis such that the global Type-I error does not exceed 5%.
3. If both non-inferiority of MC2-01 to active comparator with respect to percentage reduction from Baseline in mPASI and superiority of MC2-01 vs. active comparator with respect to PTCS has been demonstrated, superiority of MC2-01 vs vehicle with respect to the below endpoints will be tested in a hierarchical manner based on a significance level of 5%:
    - Change in itch intensity assessed on a numerical rating scale (Itch by NRS) from baseline to Week 4
    - Percentage of subjects with four-point improvement on NRS from baseline to Week 4 (in the subgroup of subjects with an NRS score ≥ 4 at baseline)

The testing strategy for the secondary endpoints is illustrated below in the case where the primary null-hypotheses have been rejected. If superiority/non-inferiority is demonstrated for one of the secondary endpoints, the corresponding unspent α-level will be reallocated according to the direction of the edges.



10. STATISTICAL EVALUATION

10.1 Subject Disposition

The number and percentage of subjects who were screened, randomised, included in each analysis population, completed the study, discontinued from the study (overall and by reason for discontinuation), and who were excluded from the PP population (overall and by reason for exclusion) will be summarized, overall, by treatment group and by sites.

A by-subject enrolment and disposition listing will be presented for all randomised subjects. Subjects who are screen failures and subjects who are not randomised will also be presented in a separate listing.

10.3 Protocol Deviation

Protocol deviations will be presented in a by-subject listing.

10.4 Demographic and Baseline Characteristics

Demographics and baseline characteristics will be summarized overall and by treatment for the ITT, PP, and Safety populations. The following demographic and baseline variables will be included:

- Age (years)
- Sex
- Race
- Ethnicity
- Fitzpatrick Skin Type
- Height (cm)
- Weight (kg)
- BMI (kg/m2)
- Baseline PGA (Trunk and/or Limbs)
- Baseline mPASI (Trunk and/or Limbs)
- Baseline SGA
- Duration of psoriasis in years (calculated as year of baseline minus year of first psoriasis diagnosis)
- Prior systemic biologic use for psoriasis (yes/no)

10.5 Visit Compliance

A by-subject visit list will be provided for all randomised subjects.

10.6 Study Medication Exposure and Compliance

Subjects will record the date of each trial medication application in the subject diary. The kit number, and corresponding date dispensed and returned will be collected. The weight of the returned kits will also be collected.

The following parameters of study medication exposure will be summarized by treatment group for the ITT and PP Population:

- Total number of days of exposure to the study drug, defined as the date of last application of study drug minus date of first application plus one
- Total number of missed doses collected in eCRF

- Total number of doses applied, defined as the total number of doses required minus total number of doses missed. Total number of doses required is defined as total number of days of exposure minus any days of investigator approved stop (if applicable)
- Total amount of drug used (g), defined as the summation of amount of drug used per kit for all dispensed kits. Amount of drug used per kit is defined as the difference in weight between the returned and dispensed kits, where weight of dispensed kits is a unit value. Amount of drug used for sealed and unreturned kits will be assumed to be 0.0 g.
- Amount of drug used per week (g/per week), defined as the total amount of drug used divided by total number of days of exposure to the study drug multiplied by 7.
- Percentage of compliance in each of the 2 periods (from Day 1 to Week 8 and from Week 4 to Week 8), defined as (the total number of doses applied in the given period + number of days where no dose is to be taken due to clear skin as agreed with the investigator in the given period) divided by total days of exposure of each period. Total days of exposure is calculated as last dose date – first dose date + 1 for Day 1 to Week 8; and last dose date – Week 4 visit date + 1 for Week 4 to Week 8 period. If Week 4 visit is missing, the duration of exposure is calculated as total days of exposure from Day 1 to Week 8 – 29.
- Subject compliance (Yes, No), defined as at least 80% in percentage compliance from Week 4 to Week 8 and from Day 1 to Week 8

10.7 Prior and Concomitant Medications

Prior (with stop dates prior to Day 1) and concomitant medications (ongoing or with stop dates on or after Day 1) for all randomised subjects will be provided in a by-subject listing. Prior and concomitant medications will be summarized by WHO-DDE Anatomical-Therapeutic-Chemical (ATC) classification and preferred term (PT) for each treatment and overall for all randomised subjects.

For the determination of prior vs. concomitant medications, the following rules regarding the stop date will be applied:

- If only year was recorded, and it is before Day 1, it is a prior medication; if year is same or after Day 1, it is assumed to be a concomitant medication.
- If day is missing, but month and year are before Day 1, it is a prior medication; if month and year are the same or after Day 1, it is assumed to be a concomitant medication.
- If start date is after Day 1, it is a concomitant medication regardless.

Psoriasis treatment history will be summarized by frequency counts and percentages, and a by-subject listing.

10.8 Medical/Surgical History

Past and current medical conditions will be coded using MedDRA dictionary and summarized by System Organ Class (SOC) and PT for PP and ITT. Medical history for all randomised subjects will be provided in a by-subject listing.

10.9 Efficacy Analysis

10.9.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the proportion of subjects in each treatment group with 'treatment success' at Week 8, defined as a minimum 2-point decrease from Baseline to Week 8 on the PGA of disease severity on the trunk and limbs; i.e. 0 (clear) or 1 (almost clear) disease for subjects with moderate disease at Baseline; or a score of 0 (clear) for subjects with, a score of mild disease at Baseline.

Superiority analysis:

The superiority evaluation will be based on the ITT population using MI.

The PGA response rate for MC2-01 cream and active comparator at Week 8 will first be compared to MC2-01 cream vehicle at Week 8 for a superiority evaluation using the treatment-policy estimand for the ITT population and imputing missing data using multiple imputation.

The analysis will be done using a logistic model with treatment, baseline PGA severity (mild/moderate) and analysis site as independent variables. Superiority will be tested at the 5% significance level ($p<0.05$; two-sided). The estimate and standard error of the log odds ratio will be evaluated for each of the multiple imputed datasets and combined in PROC MIANALYZE.

Sensitivity analyses:

Internal validity:

- The primary analysis using the treatment-policy estimand is repeated for the ITT population using LOCF, BOCF, and NRI respectively.

External validity:

- The primary analysis is repeated using the hypothetical-strategy estimand based on the ITT population and imputing missing data using MI.
- OC analysis based on the PP population.

Non-inferiority (NI) analysis:

If and only if superiority to MC2-01 cream vehicle for both MC2-01 cream and for active comparator is claimed at the 5% significance level, MC2-01 cream will be compared to active comparator for a non-inferiority evaluation on the PP-population using OC analysis.

The PGA success rate of MC2-01 cream at Week 8 will be compared with that of active comparator at Week 8, using a therapeutic non-inferiority margin of 10% points.

The percentage of subjects in each group with treatment success will be calculated along with its 95% confidence interval (CI) using normal approximation. A 95%, 2-sided CI for PM-PA, where PM and PA is the PGA treatment success rate at Week 8 for MC2-01 cream and active comparator respectively, will be computed using normal approximation using Proc FREQ with RISKDIFF option. MC2-01 will be considered non-inferior to active comparator if the lower bound of the 2-sided 95% CI is ≥ -10% points.

Sensitivity analyses:

- The primary analysis is repeated using the hypothetical-strategy estimand based on the ITT population and imputing missing data using MI.

The primary analysis is repeated using the treatment-policy estimand for the ITT population using MI, LOCF, BOCF, and NRI respectively. For the sensitivity non-inferiority analysis using MI, the treatment difference will be evaluated for each of the multiple imputed datasets using PROC FREQ with RISKDIFF option. The estimates and standard errors of the response rate difference based on the imputed datasets will combined by applying Rubin's rules in PROC MIANALYZE and the 95% CI of the difference will be computed.

To investigate the homogeneity of the primary efficacy outcome across analysis sites, a Forest plot will be prepared showing the PGA treatment success rate difference of each analysis site. Analysis sites with outlying results will be investigated and sensitivity analysis excluding outliers may be performed if deemed necessary.

10.9.2 Secondary Efficacy Analyses

Secondary endpoints are the following:

- Percentage reduction from Baseline in the mPASI score at Week 8, defined as Baseline score minus Week 8 score divided by Baseline score.
- Subject assessment of treatment convenience at Week 8 using the PTCS between MC2-01 cream and active comparator; PTCS consists of 6 disease-specific, self-report questions with a recall period of 1 week and rated on a 1-10 scale:
    1. How easy was the treatment to apply to the skin?
    2. How greasy was the treatment when applying it to the skin?
    3. How moisturised did your skin feel after applying the treatment?
    4. How greasy did your skin feel after applying the treatment?
    5. How much did treating your skin disrupt your daily routine?

6. Overall, how satisfied were you with the medical treatment?

The PTCS score calculated as the sum of the first 5 questions will be used for purpose of the secondary endpoint analysis. Question number 6 will not contribute to the PTCS score but will be included in the validation of the PTCS

- Change in itch intensity assessed on a numerical rating scale (Itch by NRS) from baseline to Week 4 (MC2-01 cream vs MC2-01 cream vehicle)
- Percentage of subjects with four-point improvement on the itch by NRS from baseline to Week 4 (MC2-01 cream vs MC2-01 cream vehicle). The endpoint will be analyzed in the subgroup of subjects with Itch (assessed by NRS) ≥ 4 at baseline.

The secondary efficacy endpoint analyses will be based on the PP population and ITT population. For the analysis of non-inferiority, the primary population will be the PP-population and the analysis will be based on an OC analysis. For the analyses of superiority, the primary population will be the ITT population, using imputation methods as described in Section 8 above.

Secondary endpoints will only be tested if the primary endpoint rejects the null-hypothesis.

Percentage reduction from Baseline in mPASI score at Week 8

The analyses are done in analogy with those of the primary endpoint.

Superiority analysis:

The percent reduction in mPASI score at Week 8 of MC2-01 cream and active comparator will first be compared with that of MC2-01 cream vehicle for a superiority evaluation using the treatment-policy estimand for the ITT population and imputing missing data using multiple imputations. Superiority will be assessed using an analysis of covariance (ANCOVA) model with treatment, baseline PGA severity (mild/moderate), baseline mPASI and analysis site as independent variables.

Internal validity:

- The primary analysis using the treatment-policy estimand is repeated for the ITT population using LOCF and BOCF respectively.

External validity:

- The primary analysis is repeated using the hypothetical-strategy estimand based on the ITT population and imputing missing data using MI.
- OC analysis based on the PP population.

Non-inferiority (NI) analysis:

If and only if superiority to MC2-01 cream vehicle for both MC2-01 cream and for active comparator can be claimed at the significance level determined by the loop-back procedure described in section 9.6 above, the NI analysis of the mPASI will be conducted.

The primary analysis of NI will be based on the PP-population with an OC analysis using a similar ANCOVA model. A lower confidence limit will be calculated for the difference in mean percentage reduction (MC2-01 cream – active control), using the LSMEANS function. The alpha level corresponding to the confidence interval will be determined by the loop-back procedure described above. Non-inferiority will be claimed if the lower limit of the CI is ≥ −10%.

Sensitivity analyses:

- The primary analysis is repeated using the hypothetical-strategy estimand based on the ITT population and imputing missing data using MI.
- The primary analysis is repeated using the treatment-policy estimand for the ITT population using MI, LOCF, and BOCF respectively.

Subject assessment of treatment convenience at Week 8

Superiority between MC2-01 cream and active comparator will be assessed by using an ANOVA model with treatment, baseline PGA severity (mild/moderate), and analysis site as independent variables. The primary analysis will be based on the ITT population, where missing and invalid PTCS scores are imputed using the last valid measure prior to the visit corresponding to a PTCS while-on-treatment estimand.

Superiority will be claimed if the lower limit of the CI, derived from the difference (MC2-01 cream – active comparator) in LSMEANS, is >0. The alpha level corresponding to the CI will be determined by the loop-back procedure.

Change in Itch by NRS from baseline to Week 4

Superiority between MC2-01 cream and vehicle will be assessed by using an ANCOVA model with treatment, baseline PGA severity (mild/moderate), baseline Itch by NRS value and analysis site as independent variables. The primary analysis will be based on the ITT population, using imputation methods as described in Section 8 above.

Percentage of subjects with 4-point improvement in NRS from baseline to Week 4

Superiority between MC2-01 cream and vehicle will be assessed by using a logistic regression model with treatment, baseline PGA severity (mild/moderate), baseline Itch by NRS value and analysis site as independent variables. The primary analysis will be based on the subgroup of the ITT population with baseline itch ≥ 4, using imputation methods as described in Section 8 above.

All secondary endpoints will be described using summary statistics. For mPASI and NRS, absolute values, absolute reduction, and percentage reduction from baseline will be summarized using

mean, median, SD, quartiles and ranges for each treatment group by visit. Frequency counts and percentages of NRS 4-point improvement responders will be provided for each treatment group by visit. For PTCS score, absolute values will be summarized using descriptive statistics for each treatment group by visit.

10.9.3 Other Endpoints

Analyses of the other endpoints will primarily be based on the PP population using an OC approach. Supportive analyses will be made based on the ITT population; OC approach will be used for all patient reported outcomes; MI will be used for efficacy endpoints based on the ITT population.

Efficacy

- PGA success rate at Week 4;
- Percentage reduction from Baseline in the mPASI score at Week 4;
- Percentage of subjects with PASI 50 (at least 50% reduction in mPASI from Baseline) at Week 4 and Week 8;
- Percentage of subjects with PASI 75 (at least 75% reduction in mPASI from Baseline) at Week 4 and Week 8.
- Reduction from Baseline in SGA at Week 4 and Week 8;

Patient reported outcomes

- Change in total DLQI score at Week 4 and Week 8 from baseline; The DLQI will be scored according to the developer instructions. The scoring of each question is as follows:

Response	**Score**
Very much	scored 3
A lot	scored 2
A little	scored 1
Not at all	scored 0
Not relevant	scored 0
Question unanswered	scored 0
Question 7: “prevented work or studying”	scored 3

The DLQI will also be summarized by the following domains: symptoms and feelings, daily activities, leisure, work and school, personal relationships, and treatment.

The scoring of DLQI will follow the developer's manual[1]. Specifically:

- If one question is left unanswered this is scored 0 and the scores are summed and expressed as usual out of a maximum of 30.
- If two or more questions are left unanswered the questionnaire is not scored.
- If question 7 is answered 'yes' this is scored 3. If question 7 is answered 'no' but then either 'a lot' or 'a little' is ticked this is then scored 2 or 1. If "Not relevant" is ticked, the score for Question 7 is 0. If it is answered 'no', but the second half is left incomplete, the score will remain 0.
- If two or more response options are ticked, the response option with the highest score should be recorded.
- If there is a response between two tick boxes, the lower of the two score options should be recorded.
- The DLQI can be analysed by calculating the score for each of its six sub-scales (see above). When using sub-scales, if the answer to one question in a sub-scale is missing, that sub-scale should not be scored.

- Change in EQ-5D score at Week 4 and Week 8 from baseline. The EQ-5D consists of a descriptive system (based on the five first questions) and a VAS scale (How good or bad is your health today: Range of 0 (worst health) to 100 (best health)). Change from baseline in EQ-5D will be described using the below endpoints
  - Change in EQ-5D VAS from baseline. The score of the VAS scale is the answer from 0 to 100. From this the change from baseline is derived.
  - Change from baseline based on the descriptive system. Each of the five questions have 5 reply options. Level '1' (No bother) is coded as '1', level '2' as '2' and so forth. Hence, after scoring each patient will have a health state consisting of 5 digits e.g. ´13425' at each visit. A post-baseline assessment is characterised according to whether there is "No change/improved health/worse health/'mixed' change" compared to baseline. Here, an EQ-5D health state is better if it is better on at least one other dimension and no worse in any other dimension. And a health state is worse than another if it is worse in at least one dimension and no better in any other dimension.

[1] DLQI Instructions for use and scoring. A Y Finlay, G K Khan April 1992. Available at: http://sites.cardiff.ac.uk/dermatology/quality-of-life/dermatology-quality-of-life-index-dlqi/dlqi-instructions-for-use-and-scoring/

- Change from baseline in EQ-5D index value. The descriptive health states described above will be transformed into an index value by use of the UK crosswalk value set.

- Change in itch intensity assessed on a numerical rating scale (Itch by NRS, 0-10) from baseline to Week 1 as well as from baseline to week 8 (MC2-01 cream vs MC2-01 cream vehicle)
- Subject assessment of each individual question at Week 8 in the Psoriasis Treatment Convenience Scale

PGA success, PASI 50, and PASI 75 will be analysed following the same method as described for the primary efficacy analysis, except that no formal non-inferiority test will be conducted. The endpoints PASI 50 and PASI 75 will be summarized using frequency counts by treatment and visit.

The endpoint change from Baseline in SGA at Week 4 and Week 8 will be analysed using an ANCOVA model with treatment, baseline PGA severity (mild/moderate), baseline SGA and analysis site as independent variable. Absolute values and absolute changes from baseline will be summarized using descriptive statistics for each treatment group by visit.

The Patient reported outcomes DLQI, EQ-5D (Vas and index value), and Itch by NRS will be analysed using an ANCOVA model with treatment, baseline PGA severity (mild/moderate), analysis site and respective baseline patient reported outcome as independent variables. Absolute values and absolute changes from baseline will be summarized using descriptive statistics for each treatment group by visit. Change from baseline in the descriptive system of EQ-5D will be described using summary statistics.

The individual questions of the PTCS will be analysed similarly to the PTCS scale. Statistical analyses for the other endpoints will be exploratory only and not for statistical inference.

10.9.4 Investigator's Assessment of the Body Surface Area Involvement (BSA)

The total psoriatic involvement on the trunk and limbs (excluding genital and intertriginous areas) will be recorded as a percentage of the total BSA. Results will be presented in a by-treatment and by-subject listing.

10.10 Safety Analysis

All safety analyses will be based on the Safety Population according to the treatment received. The safety end points are:

- Adverse events (AEs)
- Local Skin Reaction Assessment

- Laboratory parameters
- Vital signs
- Standard 12-lead electrocardiogram (ECG)
- Physical examination (PE)

10.10.1 Adverse Events

AE terms will be coded using MedDRA dictionary. A treatment-emergent AE (TEAE) is defined as an AE with a start date on or after the first application to trial medication. If relationship to study medication is missing, the event will be conservatively summarized as being related to study drug. If intensity is missing, a separate category of missing intensity will be included in the summary table, and no imputation of severity will be performed. Through the data cleaning process, all attempts will be made to avoid missing values for relationship and severity.

All AEs will be presented in a by-subject listing for all screened subjects, detailing the verbatim term given by the investigator, the PT, SOC, location of AE to treatment area, start date, stop date, intensity, outcome, relationship to study medication (definitely related, probably related, possibly related, and not related), action taken with study medication, action taken to treat the AE, seriousness and criteria for seriousness. Serious AEs will also be presented in a separate listing.

An overall summary of AEs will be presented by treatment and overall. The summary will include the total number of events, frequency counts and percentages with:

- Any AE
- Any TEAE
- Any serious TEAE
- Any treatment-related TEAE
- Any TEAE leading to study drug discontinuation

Summaries of the incidence of TEAEs, and SAEs will be displayed by treatment according to the following:

- All TEAEs by PT in descending order of frequency (the combined frequency of both treatments)
- All TEAEs by SOC, PT, and maximum severity (mild, moderate, or severe)
- All TEAEs by SOC, PT, and maximum causality (not related, related) to the study drug

At each level of summarization, a subject will be counted once if he/she reported one or more events. The severity of TEAEs and relationship to study drug will be summarized in a similar manner. For summaries of relationship to study drug, a subject will be classified according to the

closest relationship. For summaries of TEAE severity, a subject will be classified according to the highest severity.

10.10.2 Investigator Assessed Local Skin Reaction

Assessments of local skin reaction will be performed at all post-screening visits for all subjects in the trial by the investigator. The treatment area and the immediate surrounding area will be assessed.

The LSR intensities, scored as 0=none, 1=mild, 2=moderate, or 3=severe, will be summarized by visit for each of the LSR categories.

The investigator LSR sum score will be calculated as the summation of the perilesional signs and lesional signs. The perilesional signs will include erthema, scaling, edema, atrpoy, vesicles and erosin/ulceration. The lesional signs will include vesicles and erosin/ulceration. Missing values will be imputed as the mean of the remaining categories of perilesional and lesional signs. If all categories have missing values no imputation will be made.

For the LSR sum score, the Area under the Curve (AUC) will be calculated for each subject for four time intervals using the trapezoidal rule; 1) from Baseline to end of Week 1; 2) from Baseline to end of Week 4; and 3) from Week 4 to End of Trial/last assessment and 4) from Baseline to End of Trial/last assessment. These parameters will be compared between treatment groups using ANOVA.

For each LSR category, the most intense reaction over the course of the trial will be determined for each subject, and the frequency distributions of these scores will be tabulated.

10.10.3 Subject Assessed Local Skin Reaction

Subjects will assess burning and/or pain after application. Results will be summarized and presented in a by-treatment and by-subject listing.

10.10.4 Safety Laboratory Parameters

Clinical laboratory (serum biochemistry and urinalysis) results will be presented in a by-subject listing.

Absolute values, changes from baseline, and relative changes from baseline (defined as ratio of the post-baseline vs baseline value) will be summarized using descriptive statistics (means, medians, standard deviations, ranges, and coefficient of variations) by treatment and visit. Serum biochemistry will include serum calcium (albumin corrected), serum albumin, serum alkaline phosphatase, serum phosphate, plasma parathyroid hormone and. Urinalysis parameters will include urinary calcium, urinary phosphate and urinary calcium:creatinine. Shift tables using

extended normal ranges (baseline to most extreme post-baseline value) will be summarized. Out of range laboratory values will be flagged.

Serum and urine pregnancy test will be presented in a by-subject listing with subject childbearing potential by visit.

10.10.5 Vital Signs

Vital signs data will be presented in a by-treatment and by-subject listing with indication of the arm the vital signs were collected from for the following tests:

- Blood pressure (mmHg) - systolic and diastolic
- Pulse rate (beats per min)

Absolute values and changes from baseline will be summarized using descriptive statistics by treatment and visit.

10.10.6 Standard 12-lead electrocardiogram (ECG)

A standard 12-lead electro cardiogram (ECG) will be evaluated by the Investigator or a designee and will be sent to a central laboratory for interpretation. Frequency counts will be used to summarize the ECG findings (normal, abnormal not clinically significant, abnormal clinically significant) by treatment and visit. Shift tables showing changes in ECG findings from baseline will be summarized. All results will also present in a by-subject listing.

10.10.7 Physical Examination (PE)

Physical examinations results will be presented in a by-subject listing.

11. CHANGES FROM THE PROTOCOL AND PLANNED ANALYSES

Not applicable.

12. HEADINGS

Each page of the analysis will show the sponsor’s name, the investigational product, the protocol and number. Report tables will be embedded in the MS Word report document from SAS program output without change. The footer of each table will show the name of the SAS program module which generated it, the names of all data sets providing input data in the program and the date and time the table was generated.

13. ARCHIVING AND RETENTION OF DOCUMENTS

After finalization of the analysis, the following will be archived at Novella Clinical and/or with the study sponsor:

- SAP and any amendments
- All SAS code used in the project for statistical analysis, report tables generation, and analysis data set creation
- Tables, listings and figures as included in the clinical study report
- SAS study data tabulation model (SDTM) and analysis dataset model (ADaM) datasets

14. REFERENCE

Fleming C, Ganslandt C, Guenther L et al. Calcipotriol plus betamethasone dipropionate gel compared with its active components in the same vehicle and the vehicle alone in the treatment of psoriasis vulgaris: a randomised, parallel group, double-blind, exploratory study. *Eur J Dermatol* 2010;20:465-471.

Langley R, Gupta A, Papp K, Wexler D, Osterdal ML, Curcic D. Calcipotriol plus betamethasone dipropionate gel compared with tacalcitol ointment and the gel vehicle alone in patients with psoriasis vulgaris: a randomised, controlled clinical trial. *Dermatology* 2011;222:148-156.

Menter A, Gold LS, Bukhalo M et al. Calcipotriene plus betamethasone dipropionate topical suspension for the treatment of mild to moderate psoriasis vulgaris on the body: a randomised, double-blind, vehicle-controlled trial. *J Drugs Dermatol* 2013;12:92-98.

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: Wiley.

Schaefer, J.L. (1997). Analysis of Incomplete Multivariate Data, Chapman&Hall.

15. OUTLINE OF PROPOSED TABLES, LISTINGS AND FIGURES