NOVARTIS

Clinical Development

ZPL389

CZPL389A2203 and A2203E1 / NCT03948334

A randomized, double-blind, placebo-controlled multicenter dose-ranging study and extension to assess the safety and efficacy of multiple oral ZPL389 doses and long-term safety and efficacy of oral ZPL389 with concomitant or intermittent use of TCS and/or TCI in patients with moderate to severe atopic dermatitis

Statistical Analysis Plan (SAP)

Author:	Statistician, [redacted]
Document type:	SAP Documentation
Document status:	Final
Release date:	29-Sep-2020
Number of pages:	35

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
29-Sep-2020	Prior to DB lock	Creation of final version	N/A - First version	NA

Table of contents

List of abbreviations

AE	Adverse event
ATC	Anatomical Therapeutic Classification
AUC	Area Under the Curve
bid	bis in diem/twice a day
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
DMC	Data Monitoring Committee
FAS	Full Analysis Set
eCRF	Electronic Case Report Form
IVR	Interactive Voice Response
IWR	Interactive Web Response
MedDRA	Medical Dictionary for Drug Regulatory Affairs
NCI	National Cancer Institute
o.d.	Once Daily
OS	Overall Survival
PFS	Progression-Free Survival
[redacted]	[redacted]
PPS	Per-Protocol Set
PRO	Patient-reported Outcomes
qd	Qua'que di'e / once a day
[redacted]	[redacted]
RAP	Report and Analysis Process
RECIST	Response Evaluation Criteria in Solid Tumors
SAP	Statistical Analysis Plan
SOC	System Organ Class
TCI	Topical Calcineurin Inhibitors
TCS	Topical Corticosteroids
TFLs	Tables, Figures, Listings
WHO	World Health Organization

1 Introduction

This Statisticial Analysis Plan (SAP) describes the statistical analysis planned for the Clinical Study Report (CSR) for studies ZPL389A2203 and ZPL389A2203E1 (henceforth called A2203 and A2203E1, respectively). Both studies will be reported in the same CSR.

Following an interim analysis with database lock in May 2020, the decision was made to stop both study A2203 (the core study) and A2203E1 (an extension study) and this SAP reflects changes to the analysis strategy described in the protocols (V3 for A2203 and V1 for A2203E1) that were made following this decison.

1.1 Study design

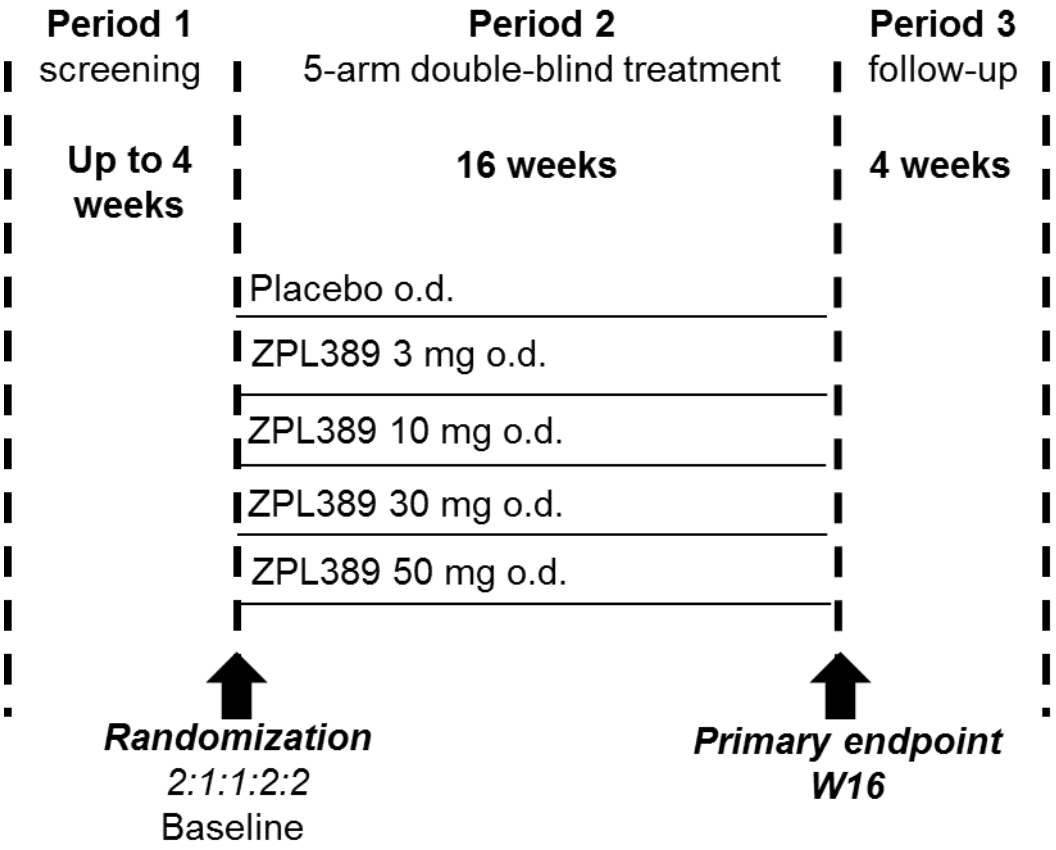
The core study (A2203) was planned as a randomized, double blind, placebo controlled, parallel group, 5-arm study in at least 360 subjects with moderate to severe Atopic Dermatitis (AD). Unbalanced randomization ratio was used, intending to result in approximately 90 subjects randomized in each of the 30 mg/day and 50 mg/day dose arms as well as in the placebo arm, and 45 subjects in each of the 3 mg/day and 10 mg/day arms. The randomization was stratified by baseline severity of atopic dermatitis (IGA score moderate or severe) and by geographical region.

The possibility to perform an interim analysis (IA) in order to support decision making concerning the current clinical study or project was addeded to the protocol in V3. This interim analysis was performed in May 2020 on 187 patients who had completed week 16 or discontinued before this timepoint. Based on the results from this IA, it was decided to stop both the core and extension studies for reasons of lack of efficacy and thus the final number of patients, and follow-up time per patient is less than planned in the protocol.

The core study (A2203) consisted of a screening period of up to 4 weeks (depending on current medication use and associated washout period), and a 16-week double blinded treatment period. The primary endpoint was to be assessed at the end of the 16 week treatment period.

After the end of treatment visit, subjects were offered the possibility of ongoing treatment in the extension study (A2203E1), or of entering the 4 week treatment-free follow up period.

Figure 1-1 Core study design (A2203)

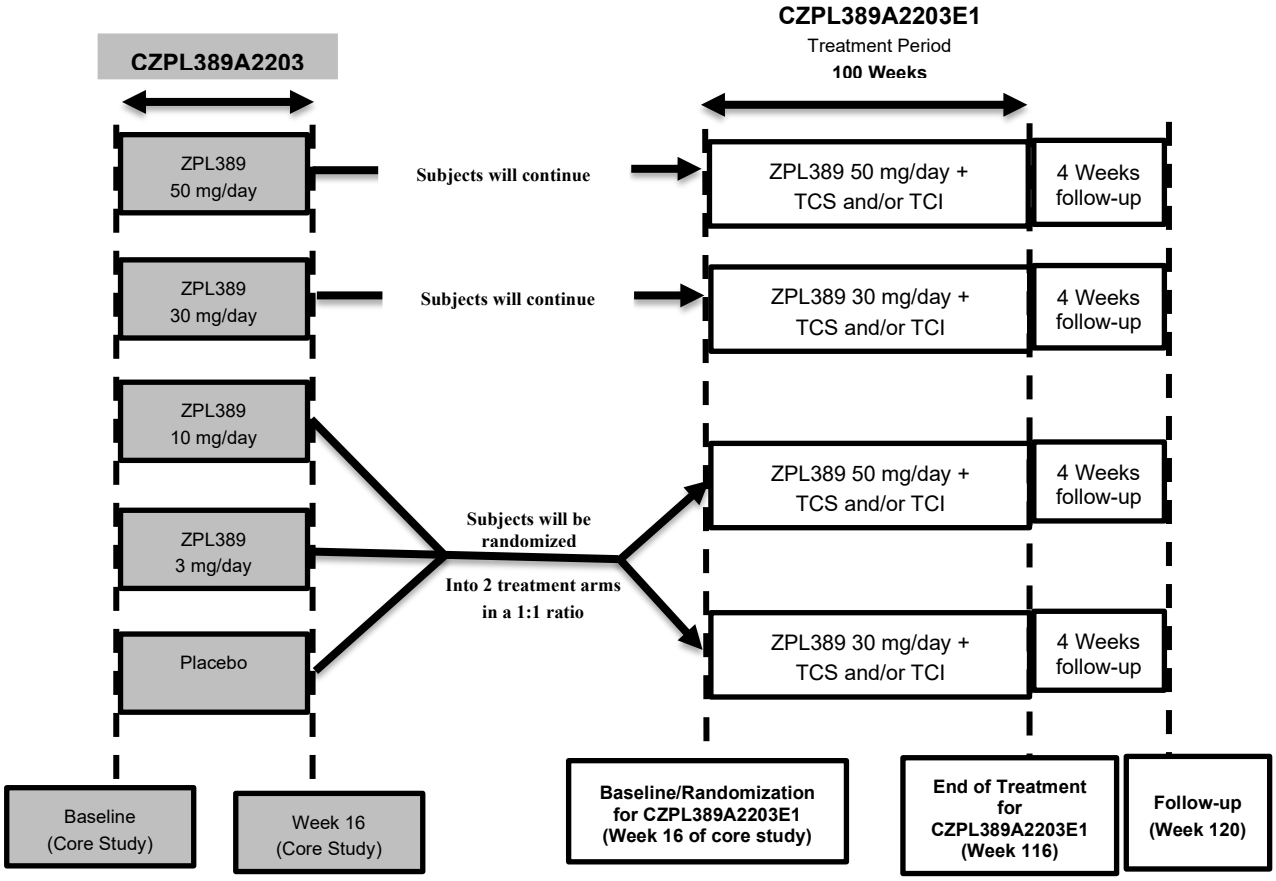


The extension study (A2203E1) was intended to be a randomized, double blind, parallel group, 2-arm study in up to approximately 202 to 230 subjects. This number assumed a discontinuation rate of 15% to 25% in the core study (A2203)

In the extension study, subjects who had received ZPL389 30 mg o.d. or 50 mg o.d. doses in the core study, would continue to receive the same doses in a double-blinded fashion. Subjects who had received ZPL389 3 mg, 10 mg or placebo in the core study, would be randomized to receive ZPL389 30 mg o.d. or 50 mg o.d. in a 1:1 ratio. All subjects would receive concomitant or intermittent topical corticosteroids (TCS) and/or topical calcineurin inhibitors (TCI) along in a standardized regimen with ZPL389.

After completing 100 weeks of study treatment during the extension study (i.e. up to Week 116 including 16 weeks of the core study), subjects were to enter the 4 weeks treatment-free follow-up period.

Figure 1-2 Extension study design (A2203E1)



1.2 Study objectives and endpoints

Following the decision to stop the studies, not all [redacted] will be evaluated in the abbreviated CSR, as indicated in the table below.

Table 1-1 Objectives for the core study A2203

Objective	Endpoint	Included in CSR?
Primary objective		
To characterize the dose-response relationship of ZPL389 in subjects with moderate to severe AD assessed by IGA response after 16 weeks of treatment	IGA response at Week 16	Yes
Secondary objectives		

Objective	Endpoint	Included in CSR?
To characterize the dose-response relationship of ZPL389 in subjects with moderate to severe AD assessed using the percent change from Baseline in Eczema Area and Severity Index (EASI) score after 16 weeks of treatment	Percent change from baseline EASI score at Week 16	Yes
To evaluate the efficacy across different dose levels as assessed by EASI and IGA compared to placebo over time	At each visit: IGA score IGA response EASI score (absolute and percent change from Baseline) EASI50 response, EASI75 response	Yes (except IGA score and absolute change in EASI score from baseline)
To assess the safety and tolerability of different doses of ZPL389 as compared to placebo	Frequency of adverse events	Yes

[redacted]

Objective	Endpoint	Included in CSR?
[redacted]		

Table 1-2 Objectives for the extension study A2203E1

Objectives	Endpoints	Included in CSR?
Primary objective		

Objectives	Endpoints	Included in CSR?
To assess the short-term and long-term safety of 30 mg o.d and 50 mg o.d ZPL389 with concomitant or intermittent use of TCS and/or TCI up to total of 32 weeks and 116 weeks of treatment.	Frequency of AEs at Week 32 (short-term) and Week 116 (long-term).	Yes
Secondary objectives		
To evaluate the efficacy of 30 mg o.d and 50 mg o.d ZPL389 with concomitant or intermittent use of TCS and/or TCI as assessed by IGA response over time.	IGA score over time (absolute and relative frequencies from core study baseline).	Yes
To evaluate the efficacy of 30 mg o.d and 50 mg o.d ZPL389 with concomitant or intermittent use of TCS and/or TCI as assessed by EASI over time.	EASI score over time (absolute and percent change from core study baseline).	Yes
[redacted]		

Objectives	Endpoints	Included in CSR?
[redacted]		

2 Statistical methods

2.1 Data analysis general information

Novartis will perform the analyses for this study. Statistical software SAS verion 9.4 or later will be used.

All available data from the core and extension studies will be presented together.

Where summary statistics are presented these will include, for continuous variables N, mean, standard deviation, minimum, and maximum, and for discrete variables, counts and appropriate percentages.

If not otherwise specified, p-values will be presented for two-sided hypothesis testing, two-sided confidence intervals will be displayed and the level of significance will be set to 5% two-sided.

In general listings will be presented by treatment sequence showing the sequence of treatments in the core and extension study, i.e as follows and in this order

- Placebo – did not participate in extension
- Placebo – ZPL389 50mg
- Placebo – ZPL389 30mg

- ZPL389 3mg – did not participate in extension
- ZPL389 3mg – ZPL389 30mg
- ZPL389 3mg – ZPL389 50mg
- ZPL389 10mg – did not participate in extension
- ZPL389 10mg – ZPL389 30mg
- ZPL389 10 mg – ZPL389 50mg
- ZPL389 30mg – did not participate in extension
- ZPL389 30mg – ZPL389 30mg
- ZPL389 50mg – did not participate in extension
- ZPL389 50mg – ZPL389 50mg

2.1.1 General definitions

Study treatment is either placebo or 3, 10, 30 or 50mg ZPL389.

Day 1 is defined to be the first day of administration of any study treatment.

All other study days will be labeled relative to Day 1. For event dates on or after Day 1, studyday for a particular event date is calculated as [Date of event] – [Date of first dose]+1, i.e., Day 2, Day 3, etc., will be one day, two days, etc., after Day 1, respectively. For the dates beforeDay 1, study day for an event date is calculated as [Date of event] – [Date of first dose], i.e., Day -1, Day -2, etc., will be one day, two days, etc., before Day 1, respectively. Duration of an event will be calculated as (Event end date – Event start date + 1).

The descriptor “Day 0” will not be used.

Screening refers to any procedure (e.g., checking inclusion and exclusion criteria) performed prior to the date of first dose of study treatment. Per protocol, subject informed consent must be obtained prior to performing any study related activity. The date of signing informed consent is the start date of screening period.

Baseline is considered to be the last assessment (including those from unscheduled visits) obtained before the first dose of study treatment. All assessments obtained after first dose of study treatment are considered as post-baseline unless otherwise specified.

In general, a baseline value refers to the last measurement made prior to administration of the first dose of study treatment. However, if no pre-treatment value exists, values obtained after first dose of treatment but on that same day of dosing will be considered as the baseline.

The date of last dose will be collected via the CRF. The end of the treatment period will be defined as the last dose date or last visit in treatment period whichever occurs earlier. i.e., for subjects who discontinued or have their last visit earlier than last dose date, the end of study treatment will be the date of the last study visit in the corresponding treatment period.

The core study treatment period is defined to be the time up until the date of the first visit in the extension study whilst the patient is receiving study treatment.

The extension study treatment period is defined to be the time up from the first visit in the extension study until the end of study treatment

The follow-up period is defined to be the time from last dose of study treatment until the end of the study (latest of core and extension for that patient).

The concept of assessment windows will be used for the data that is summarized by visit. Each assessment window is based on the study evaluation schedule and comprises a set of days around the nominal visit day. For any assessment, there are the protocol defined scheduled visits around which windows were created to cover the complete range of days within the study. The visit windows are shown in Table 2-1.

These assessment windows apply to measurements taken at every visit. For assessments which are not collected at every visit, the assessment windows defined in Table 2-1 will be combined. For example, if an assessment is measured at Week 4 and Week 8 only, the Week 4 assessment window for this parameter will extend from Day 2 to Day 43 (combining Week 2 to Week 4), Week 8 will extend from Day 44 to Day 71 (combining Week 6 to Week 8).

When assessment windows are used, all visits will be re-aligned, i.e., they will be mapped into one of the assessment windows. For example, if the Week 4 visit of a subject is delayed and occurs on Day 60 instead of on Day 29, the data will be considered to relate to Week 8. In the case of major deviations from the visit schedule, or due to unscheduled visits, several assessments of a subject may fall in a particular assessment window (either scheduled or unscheduled). The approaches to handle multiple assessments in a given window are specified below.

Assignment of data to the assessment windows will be done before deriving further parameters (e.g. change from baseline, responder status) from the assessments. The assessment windows will be used for summary tables and figures but listings will show the original visit at which the data were recorded, along with the study day of the assessment.

Table 2-1 Visit windows

Visit as used for reporting	Week	Scheduled study day	Assessment window
Screening	SCN	-35 to -1	Day -35 to -1
Baseline	0	1	Day 1
Week 2	2	15	Day 2 – 22
Week 4	4	29	Day 23 – 26
Week 6	6	43	Day 37 – 50
Week 8	8	57	Day 51 – 71
Week 12	12	85	Day 72 – 99
Week 16	16	113	Day 100 – 127
Week 20	20	141	Day 128 – 155
Week 28	28	197	Day 184-211
Week 32	32	225	Day 212-267
Week 44	44	309	Day 268-351
Week 56	56	393	Day 352-435
Week 68	68	477	Day 436-519
Week 80	80	561	Day 520-603
Week 92	92	645	Day 604-687

Visit as used for reporting	Week	Scheduled study day	Assessment window
Week 104	104	729	Day 688-771
Week 116	116	8130	Day 772-827
Week 120	120	841	Day >= 828

If more than one assessment falls into the interval, the rules defined below will be applied unless otherwise specified.

- for quantitative variables, the closest to the actual visit is chosen (if two assessments have the same distance, then the earlier one will be chosen);
- for qualitative variables, the worst record is selected.

Confounding medications

Any use of rescue medication or prohibited medication will be considerd as confounding therapy, where rescue medications are collected from the rescue therapy CRF and concomitant medications listed below will be considered as prohibited medications.

1. All cases of conmeds in the following ATC classes
   - D07 CORTICOSTEROIDS, DERMATOLOGICAL PREPARATIONS
   - R06A ANTIHISTAMINES for systemic use
2. Conmeds in the class L04 IMMUNOSUPPRESSANTS that are also PDs with PD identifier COMD3.

2.2 Analysis sets

Randomized Set (RAN): All subjects randomized. Subjects will be analyzed according to the treatment assigned to at randomization.

Full Analysis Set (FAS): All subjects to whom study treatment has been assigned. Subjects will be analyzed according to the treatment assigned to at randomization. Mis-randomized subjects (mis-randomized in IRT) will be excluded from FAS. Mis-randomized subjects are defined as cases where IRT contacts were made by the site either prematurely or inappropriately prior to confirmation of the subject's final randomization eligibility and no study medication was administered to the subject.

Safety Set: All subjects who received at least one dose of study treatment. Subjects will be analyzed according to treatment received.

Extension study set: All subjects in the full analysis set who have at least one visit in the extension study.

The protocol deviation codes leading to exclusion from the analysis sets defined above are presented in below.

Table 2-2 Subject classification rules

Analysis set	PD codes that cause a subject to be excluded	Non-PD criteria that cause a subject to be excluded
Randomization set	None	Not-randomized.

Analysis set	PD codes that cause a subject to be excluded	Non-PD criteria that cause a subject to be excluded
Full analysis set	None	Not-randomized, mis-randomized.
Safety set	None	No treatment taken

2.3 Patient disposition, demographics and other baseline characteristics

2.3.1 Patient disposition

Patient dispostion will be summarized based on the randomization set and will focus on showing the number of patients enrolled and the times at which they discontinued the study, including how many progressed to the extension study and how many did not. Data will be summarised by treatment group to determine whether discontinuation rates and reasons were similar across groups.

The key focus is on enrolment and discontinuation during the core study. Subject disposition in the extension is unlikely to be as intended due to the early stopping of the trial.

Since part of the study was conducted during the COVID-19 pandemic summary tables will also indicate how many patients were enrolled and completed the trial during the pandemic time period.

2.3.2 Demographics and baseline characteristics

The demography and baseline characteristics summaries are intended to answer the questions:

- Is the trial population the one we intended to recruit?
- Are the characteristics of the trial population similar across treatment groups?

Contrary to the protocol, demographics and baseline characteristics will be summarized for the full analysis set instead of the randomized set. This will enable a comparison of the population for whom efficacy data is presented, with other studies.

Only key demographic and baseline characteristics that would be important to understand the patient population and put efficacy results into context of other trials will be presented.

Specifically these are:

- Age and age category (<65 years, 65 years and older)
- Weight
- Body mass index
- Gender
- Race
- Smoking status
- Baseline IGA score categories (moderate, severe)
- Baseline EASI
- [redacted]

- Baseline total body surface area (BSA) affected by AD
- [redacted]
- Time since diagnosis of AD
- Age at the time of diagnosis of AD
- Previous exposure to systemic corticosteroids(yes, no)
- Previous exposure to topical corticosteroids (TCS) therapy (yes, no)
- Previous exposure to topical calcineurin inhibitors (TCI) therapy (yes, no)
- Previous exposure immunosuppressant therapy for atopic dermatitis (yes, no)
- Previous exposure biologic systemic therapy for atopic dermatitis (yes, no)
- Previous exposure to other therapy for atopic dermatitis (including phototherapy and tanning booths) (yes, no)

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

2.4.1 Study treatment / compliance

The summaries of study treatment intake are intended to answer the question:

- Was there a difference in study treatment intake across the dose groups?

Other questions related to the intake of study treatment, such as whether this affected the efficacy outcomes are considered in later sections.

This summary will be based on the safety set to include all patients who received study treatment according to the treatment they received.

Although not expected based on the interim results, in case there are large differences between study analysis sets, further summaries may be produced based on other analysis sets.

Duration of exposure is defined as the time from first dose of study medication to the last dose, i.e.

- duration of exposure (days) = last dose date – first dose date +1
- duration of exposure (years) = duration of exposure (days) / 365.25

In addition to summaries of the total duration of exposure, the number of subjects with exposure of particular durations will be summarized, i.e. the number and percentage of patients with exposure of

- ≥2 weeks
- ≥4 weeks
- ≥8 weeks
- ≥12 weeks
- ≥16 weeks
- ≥20 weeks
- ≥24 weeks
- ≥28 weeks

- ≥32 weeks
- ≥40 weeks
- ≥44 weeks
- ≥1 year
- ≥56 weeks
- ≥1.5y
- ≥2y

2.4.2 Prior, concomitant and post therapies

The summaries of study treatment intake are intended to answer the question:

- What medications were used by patients before start of this trial?
- What medications were used by patients whilst on treatment in this trial?
- What medications were used by patients after study treatment in this trial

Other questions related to the intake of confounding treatment (treatment that may be considered to affect the efficacy outcomes) are considered in later sections.

This summary will be based on the safety set to include all patients who received study treatment according to the treatment they received.

Medications will be presented by ATC code and preferred term. ATC codes will be displayed in alphabetical order, and grouped by anatomical main group (the first level of the ATC code).

Medical procedures and significant non-drug therapies as coded in MedDRA will be summarized along with medications.

Prior medications

Prior medications are defined as drugs taken and stopped prior to the first dose of study treatment in the core study.

These will be displayed for the baseline period by core study treatment group

Concomitant medication

Concomitant medications are defined as drugs taken at least once between start and end of study treatment (inclusive)

Concomitant treatments will be displayed for:

- the core study treatment period by the relevant treatment group for that period,
- the extension study treatment period by the relevant treatment group for that period

Follow-up medications

Follow-up medications are defined as drugs started after the last dose of study treatment.

These will be displayed for the follow-up period by the last treatment received prior to entering the follow-up period.

2.5 Analysis of the primary objective

2.5.1 Primary endpoint

The primary objective for the extension study is to assess safety of ZPL389. This is described in section 2.8. This section focuses on the primary objective of the core study which is to characterize the dose-response relationship of ZPL389 in subjects with moderate to severe AD assessed by IGA response after 16 weeks of treatment.

The primary clinical question of interest is: What is the effect of ZPL389 versus placebo on IGA response after the full 16 weeks of treatment in patients with moderate or severe atopic dermatitis, who do not use confounding therapy that may affect their AD or the assessment of it, regardless of their use of protocol-allowed concomitant medicatons?

This estimand will enable assessment of the effect of ZPL389 in the absence of confounding medication (such as rescue medication) and after the intended 16 weeks of treatment.

The primary estimand is described by the following attributes:

Population: Adult moderate to severe atopic dermatitis patients as defined by the study inclusion and exclusion criteria who were randomized to treatment, and not mis-randomized in the IRT. This implies that analyses will be based on the full analysis set.

Endpoint: IGA response at week 16, where IGA response is defined as an IGA score of 0 or 1 with at least a 2 point improvement compared to baseline and no use of confounding therapy before the assessment time point.

Treatment of interest: the randomized treatment (ZPL389 or placebo) plus medications as allowed by the protocol but with no further treatment for patients who discontinue study treatment due an AE or lack of efficacy.

Handling of remaining intercurrent events:

- Treatment discontinuations for lack of efficacy or adverse event: consider as non-responder
- Treatment discontinuations for any other reason: set data after discontinuation to be missing

The summary measure: difference in response rates between ZPL389 and placebo.

This estimand differs from that intended by the protocol in that the definition of the endpoint of IGA response has changed to include any use of confounding medication as a non-response, rather than only confounding medication that was adjudicated as having an effect on the IGA response, as specified in the protocol. This change was made following the interim analysis where it was established that handling use of confounding medication did not affect the overall conclusions and so it was decided not to perform the adjudication.

2.5.2 Statistical hypothesis, model, and method of analysis

The key questions to address this objective are

- Is there a dose-response relationship for IGA response after 16 weeks of treatment?
- What is the form of this dose-response?

- What doses deliver sufficient efficacy in terms of IGA response after 16 weeks of treatment?

Results from the interim analysis indicate that there is no dose-response relationship and therefore the second and third questions are no longer relevant. Thus the analysis of the primary variable will focus only on the logistic regression analysis described in the protocol, establishing the presence or absence of a dose-response relationship. The MCP-Mod procedure as outlined in the protocol will not be carried out.

The primary objective will be achieved if the null hypothesis of no difference between treatment groups is rejected at 5% one-sided level.

A logistic regression model will be fitted with IGA response as the outcome variable and treatment (dose as categorical variable) and baseline IGA as covariates.

The response Y_{ij} for subject j in dose group i receiving dose d_i is modelled as $Y_{ij} \sim Bin(1, p_{ij})$ with $p_{ij} = logit^{-1}(\delta_i + \mathbf{x}_{ij}'\boldsymbol{\beta})$, i=1,…, 5, j=1, …, n_i

where Bin(1,p) denotes the binomial distribution with parameter 1 and p, δ_i is the effect of dose d_i, $\boldsymbol{\beta}$ refers to the vector of regression coefficients corresponding to additional covariates x_{ij} which represent the baseline IGA category. Finally, n_i is the number of subjects allocated to each treatment group. If the covariate baseline IGA prevents fitting of a model, it will be dropped.

2.5.3 Handling of missing values/censoring/discontinuations

Where data for IGA response are missing, either because the IGA was not recorded or because the IGA response was set to missing as part of the handling of intercurrent events, the IGA score will be imputed using multiple impuation methods and the IGA response derived from this imputed IGA score.

The multiple imputation model will include IGA categories from all visits from randomization up to Week 16.

For the estimation of the IGA response rates at Week 16 for each treatment group, each imputed dataset for subjects with missing data above will be combined with the non-missing responses for subjects without missing data. For each imputed dataset the logistic regression will be performed and results combined for estimation.

2.5.4 Supportive analyses

Supportive analyses for the primary estimand will focus on different methods of handling data in the cases of intercurrent events.

The following supportive estimands will be considered:

Supportive estimand 1:

Population: Same as for the primary estimand

Endpoint: IGA response at week 16, where IGA response is defined as an IGA score of 0 or 1 with at least a 2 point improvement compared to baseline and no use of confounding therapy before the assessment time point.

Treatment of interest: the randomized treatment (ZPL389 or placebo) plus medications as allowed by the protocol but with no further treatment for patients who discontinue study treatment due an AE or lack of efficacy.

Handling of remaining intercurrent events:

- Treatment discontinuations for any reason: set data after discontinuation to be missing

The summary measure: difference in response rates between ZPL389 and placebo.

This estimand is considered since the classification of all patients who discontinue due to lack of efficacy or AE as non-responders in terms of IGA response may be too extreme. It is possible that patients experience improvement in their skin symptoms, but no improvement in another symptoms of their disease (such as itching) and for that reason discontinue study treatment. In this case, setting IGA response to non-response may not reflect what was happening to the patient. This supportive estimand will therefore help to assess whether the assumption that discontinuing the study due to lack of efficacy or AE implies lack of response on IGA is a reasonable assumption.

Supportive estimand 2:

Population: Same as for the primary estimand

Endpoint: IGA response at week 16, where IGA response is defined as an IGA score of 0 or 1 with at least a 2 point improvement compared to baseline. Any use of confounding treatment is ignored.

Treatment of interest: the randomized treatment (ZPL389 or placebo) plus medications as allowed by the protocol and further treatment taken to treat atopic dermatitis.

Handling of remaining intercurrent events:

- Treatment discontinuations for any reason: set data after discontinuation to be missing

The summary measure: difference in response rates between ZPL389 and placebo.

This estimand is considered to address the question of whether ZPL389 plus other medication as needed is more effective that placebo plus other medication as needed.

Supportive estimand 3:

Population: Adult moderate to severe atopic dermatitis patients as defined by the study inclusion and exclusion criteria who were randomized to treatment, and not mis-randomized in the IRT. This implies that analyses will be based on the full analysis set.

Endpoint: IGA response at week 16, where IGA response is defined as an IGA score of 0 or 1 with at least a 2 point improvement compared to baseline and no use of confounding therapy before the assessment time point.

Treatment of interest: the randomized treatment (ZPL389 or placebo) plus medications as allowed by the protocol but with no further treatment for patients who discontinue study treatment due an AE or lack of efficacy.

Handling of remaining intercurrent events:

- Treatment discontinuations for any reason: consider as non-responder at week 16

The summary measure: difference in response rates between ZPL389 and placebo.

This estimand is intended to assess the "worst case" situation where any discontinuation of study treatment is considered to represent lack of efficacy regardless of the reason entered on the CRF for the discontinuation. Given that a larger than expected proportion of patients gave "subject decision" as the reason for discontinuation this estimand will also help to address the question of whether ZPL389 is effective under the worst case assumption about what response would have been if patients had remained on treatment.

2.6 Analysis of secondary objectives

Secondary objectives for the core study are to characterize the dose-response in terms of EASI and to evaluate efficacy over time.

[redacted]

For the extension study the secondary objectives are to evaluate the efficacy of ZPL389 when used with concomitant or intermittent TCS and/or TCI.

2.6.1 Dose-response for EASI50 response and EASI75 response at week 16

The dose-response analyses for EASI50 response and EASI75 response will follow similar principles as those for the IGA response. The primary estimand and the three supportive estimands will be considered in order to investigate the effect of different methods of handling data in the cases of intercurrent events.

2.6.2 Dose-response for percentage change from baseline in EASI at week 16

A secondary objective for core study is to characterize the dose-response relationship in terms of percentage change from baseline EASI score after 16 weeks of treatment.

The primary clinical question of interest is: What is the effect of ZPL389 versus placebo on EASI after the full 16 weeks of treatment in patients with moderate or severe atopic dermatitis, who do not use confounding therapy that may affect their AD or the assessment of it, regardless of their use of protocol-allowed concomitant medicatons?

This estimand will enable assessment of the effect of ZPL389 in the absence of confounding medication (such as rescue medication) and after the intended 16 weeks of treatment.

The main estimand is described by the following attributes:

Population: Adult moderate to severe atopic dermatitis patients as defined by the study inclusion and exclusion criteria who were randomized to treatment, and not mis-randomized in the IRT. This implies that analyses will be based on the full analysis set.

Endpoint: Percentage change from baseline in EASI scores at week 16.

Treatment of interest: the randomized treatment (ZPL389 or placebo) plus medications as allowed by the protocol but with no further treatment for patients who discontinue study treatment due an AE or lack of efficacy.

Handling of remaining intercurrent events:

- Use of confounding medication: set data after use of confounding medication to missing
- Treatment discontinuations for any reason: set data after discontinuation to be missing

The summary measure: difference in means between ZPL389 and placebo.

This estimand differs from that intended by the protocol in that the definition data recorded after the intercurrent events of use of confounding medication or treatment discontinuation are ignored in the analysis. The protocol had specified inclusion of all data as recorded regardless of use of confounding medication or discontinuation of treatment. This change was made to allow for an estimand for the analysis of dose-response for EASI that followed similar principles as that for the IGA response.

The following supportive estimand will also be considered to address the question of what would be the effect of ZPL389 versus placebo after treatment for up to 16 weeks and when confounding medication is allowed.

Supportive estimand 4:

Population: Adult moderate to severe atopic dermatitis patients as defined by the study inclusion and exclusion criteria who were randomized to treatment, and not mis-randomized in the IRT. This implies that analyses will be based on the full analysis set.

Endpoint: Percentage change from baseline in EASI scores at week 16.

Treatment of interest: the randomized treatment (ZPL389 or placebo) plus medications as allowed by the protocol but with no further treatment for patients who discontinue study treatment.

Handling of remaining intercurrent events:

- No intercurrent events are considered.

The summary measure: difference in means between ZPL389 and placebo.

[redacted]

2.6.5 Efficacy over time up to week 16

The principles laid out for the primary estimand will be followed in calculating the summary measures for each endpoint at each time point.

The underlying clinical question in general is: What is the effect of the study treatment versus placebo on the endpoint at each week, in patients with moderate or severe atopic dermatitis, who do not use confounding therapy that may affect their AD or the assessment of it, regardless of their use of protocol-allowed concomitant medicatons, and who have received the full intended duration of treatment at that timepoint?

Other questions that may have been of interest in the case of an effective treatment, such as when is a treatment effect first seen, or how long is the treatment effect maintained for, will not be addressed.

The main secondary estimands are described by the following attributes:

Population: Adult moderate to severe atopic dermatitis patients as defined by the study inclusion and exclusion criteria who were randomized to treatment, and not mis-randomized in the IRT. This implies that analyses will be based on the full analysis set.

Endpoints:

IGA response at each week, where IGA response is defined as an IGA score of 0 or 1 with at least a 2 point improvement compared to baseline and no use of confounding therapy before the assessment time point.

Percentage change from baseline in EASI at each week

EASI50 response at each week, where EASI50 response is defined as a reduction from baseline of ≥ 50% in EASI score and no use of confounding therapy before the assessment time point.

EASI75 response at each week, where EASI75 response is defined as a reduction from baseline of ≥ 75% in EASI score and no use of confounding therapy before the assessment time point.

[redacted]

Change from baseline in peak pruritus score at each week

Treatment of interest: the randomized treatment (ZPL389 or placebo) plus medications as allowed by the protocol but with no further treatment for patients who discontinue study treatment due an AE or lack of efficacy.

Handling of remaining intercurrent events:

- Treatment discontinuations for lack of efficacy or adverse event: consider as non-responder
- Treatment discontinuations for any other reason: set data after discontinuation to be missing

The summary measure: difference in response rates or means between ZPL389 and placebo.

As for the primary estimands, three further supportive estimands will be considered in order to investigate the effect of different methods of handling data in the cases of intercurrent events.

2.6.6 Efficacy over time in the presence of TCS and/or TCI use after week 16

The underlying clinical question in general is: What is the effect of the study treatment plus use of TCS and/or TCI as needed on endpoint at each week after week 16, in patients with moderate or severe atopic dermatitis, and who have received the full intended duration of treatment at that timepoint?

The secondary estimands are described by the following attributes:

Population: Adult moderate to severe atopic dermatitis patients as defined by the study inclusion and exclusion criteria who were randomized to treatment, and not mis-randomized in the IRT, and who had received 16 weeks of ZPL389 or placebo treatment and who entered the extension study. This implies that analyses will be based on the subset of the extension study set.

Endpoints:

IGA response at each week after week 16, where IGA response is defined as an IGA score of 0 or 1 with at least a 2 point improvement compared to baseline.

Percentage change from baseline in EASI at each week after week 16

EASI50 response at each week, where EASI50 response is defined as a reduction from baseline of ≥ 50% in EASI score.

EASI75 response at each week, where EASI75 response is defined as a reduction from baseline of ≥ 75% in EASI score.

[redacted]

Change from baseline in peak pruritus score at each week after week 16

Treatment of interest: the randomized treatment (ZPL389 or placebo) plus medications as allowed by the protocol including TCS or TCI and including rescue medication but with no further treatment for patients who discontinue study treatment due an AE or lack of efficacy.

Handling of remaining intercurrent events:

- Treatment discontinuations for lack of efficacy or adverse event: consider as non-responder
- Treatment discontinuations for any other reason: set data after discontinuation to be missing

The summary measure: difference in response rates or means between ZPL389 and placebo.

This estimand is different to that implied in the protocol for the extension study which planned to consider patients who use rescue medication (TCS of higher potency than that included in the standardized regimen of intermittent of concomitant TCS and/or TCI use) as non-responders.

This change is justified given that the study has been stopped and detailed consideration of the impact of rescue medication use on efficacy after week 16 is no longer required.

2.6.7 Statistical hypothesis, model, and method of analysis

For all estimands based on binary response rates (e.g. IGA response, EASI50 response, EASI75 response) the simple response rates for each treatment group will be calculated (i.e. number of responders (after imputation as necessary) / the number of subjects in that treatment group.

Differences in response rates between ZPL389 dose groups and placebo will be calculated, and the respective 95% confidence intervals derived using a normal approximation.

For estimands based on continuous data (e.g. percentage change from baseline in EASI, [redacted]) a mixed effect repeated measures model (MMRM) approach will be used. The model will contain treatment group, baseline EASI score, geographical region if possible, baseline IGA score, visit, baseline value*visit and treatment group*visit as covariates, and will use an unstructured covariance matrix. Estimates of the mean responses for each visit and treatment group, and of the differences between treatment groups at week 16 will be obtained from this model, together with 95% confience intervals.

2.6.8 Handling of missing values/censoring/discontinuations

Where data for a binary response endpoint are are missing, either because the underlying score was not recorded or because the response was set to missing as part of the handling of intercurrent events, the underlying score (IGA, or EASI) will be imputed using multiple impuation methods and the response derived from this imputed score.

The multiple imputation model will include scores from all visits from randomization up to the time point of imputation.

For endpoints that are analyzed as continuous variables (for example EASI score % change from baseline), missing values will not be imputed. In this cases the MMRM analysis approach is used to handle the missing data.

If a baseline value is missing and required to derive an endpoint, the baseline value will be imputed by the mean of the respective treatment group in the respective randomization stratum.

2.7 Safety analyses

Safety will be assessed based on the safety set of patients.

2.7.1 Adverse events (AEs)

Adverse events will be coded according to the MedDRA dictionary. The actual MedDRA version used for reporting the adverse events will be described in a footnote of the CSR tables.

Only treatment emergent AEs will be presented. Treatment emergent adverse events are those events that

- started after the first dose of study treatment, or
- were present prior to the first dose of study treatment but increased in severity based on preferred term.

AEs observed four weeks or more after last study treatment intake will not be considered as treatment-emergent.

The purpose of the AE summaries are to understand what AEs are reported within different time periods relative to start of treatment. Thus AEs will be summarized for different time periods to cover the core and extension studies. Since AEs reported in a period may also be affected by AEs in any previous periods of treatment, the treatment groups for periods after 16 weeks will separate subjects who received placebo and the lower ZPL389 doses for the first 16 weeks of their treatment from those who had already received 30mg or 50mg ZPL389 for 16 weeks.

Therefore AEs will be summarized across the following time periods

- AEs with onset date during the first 16 weeks of treatment – where treatment group is the treatment group in the core study. This summary will be based on the full safety set.
- AEs with onset date after week 16 but before week 32 of study treatment. This summary will be based on the subset of the safety set comprised of patients who have on study treatment data after week 16. Treatment groups for this summary are
    1. Placebo and low dose ZPL389 followed by 30mg ZPL389 (for subjects who received placebo 3mg or 10mg ZPL389 in the core study)
    2. Placebo and low dose ZPL389 followed by 50mg ZPL389 (for subjects who received placebo 3mg or 10mg ZPL389 in the core study)
    3. 30mg ZPL389 (for subjects who received 30mg ZPL389 in the core study)
    4. 50mg ZPL389 (for subjects who received 50mg ZPL389 in the core study)
- AEs with onset date after week 32 of study treatment. This summary will be based on the subset of the safety set comprising patients who have on study treatment data after week 32. By week 32 it is expected that the treatment received in the core study will have less direct influence and therefore treatment groups for this summary are
    1. 30mg ZPL389
    2. 50mg ZPL389

AE summaries will show, for each treatment group, the number and percentage of subjects

- having any AE,
- having an AE in each primary system organ class
- having each individual AE (preferred term).

Summaries will also be presented for AEs by severity and for study treatment related AEs.

If a subject reported more than one adverse event with the same preferred term, the adverse event with the greatest severity will be presented. If a subject reported more than one adverse event within the same primary system organ class, the subject will be counted only once with the greatest severity at the system organ class level, where applicable.

AEs leading to discontinuation of the study drug will be summarized in a similar way.

Serious adverse events will also be summarized.

Separate summaries will be provided for deaths, serious adverse events, and other significant adverse events leading to discontinuation.

2.7.1.1 Adverse events of special interest / grouping of AEs

Summary tables showing the effects of ZPL389 in the core study on QTc and liver events will be produced following standard tables.

Additionally since part of the study was conducted during the COVID-19 pandemic, a listing of patients with suspected or confirmed SARS-COV-2 infection will be produced.

2.7.2 Deaths

Should any deaths be recorded, these will be listed

2.7.3 Laboratory data

Abnormal laboratory results that meet the criteria for adverse events will be summarized as part of the AE reporting.

[redacted]

2.12 Interim analysis

An interim analysis of the core study was performed with database lock in May 2020 based on approximately 187 patients who had completed to week 16 or discontinued the study. This IA

was planned to support decision making concerning future trials in the ZPL389 development program and focused on efficacy in terms of IGA response, EASI and pruritus. It was not the intention to stop the trial. The interim analysis charter described the steps takes to restrict the knowledge of the results of the study so as not to bias the trial conduct post interim analysis. Once the results were seen, the decision made to stop the core and extension studies.

3 Sample size calculation

The planned sample size for the core study was 360 subjects. Since the study was stopped early the final sample size will be less than this.

The sample size was determined with the software ADDPLAN DF, version 3.1.8, with settings for minimum power function and model based contrasts.

The sample size was derived to detect a dose-response with 90% power and a one-sided alpha of 2.5%. A placebo response rate of 10% was assumed. A maximum treatment effect for ZPL389 of 20%-points higher than placebo wass assumed. An unbalanced allocation ratio of 2:1:1:2:2 corresponding to treatment arms placebo: 3 mg: 10 mg: 30 mg: 50 mg, with more subjects allocated to placebo and the higher doses, was chosen to increase the precision of the estimate for the treatment effect at doses in the expected efficacious dose range.

Under these assumptions, a sample size of 360 subjects in total will be required, randomized to 90:45:45:90:90 subjects corresponding to placebo: 3 mg: 10 mg: 30 mg: 50 mg treatment arms.

4 Change to protocol specified analyses

A change was made to the primary estimand regarding adjuidcation of confounding medication use. This change was made following the interim analysis where it was established that handling use of confounding medication did not affect the overall conclusions and so it was decided not to perform the adjudication.

Following the interim analysis results which showed no treatment effect of ZPL389 on efficacy and the subsequent early stopping of the trial, the dose-response analysis using MCP-Mod methods as described in the protocol will not be performed.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

Any partial dates will be imputed as follows:

Take the earlier day of

- The last day in the month and
- The end day of the corresponding epoch.

5.1.2 Concomitant medication date imputation

Impute concomitant medication (CM) end date:

1. If CM end day is missing and CM month/year are non-missing then impute CM day as the minimum of treatment end date and the last day of the month.

2. If CM end day/month are missing and CM year is non-missing then impute CM day as the minimum of treatment end date and the end of the year (31DECYYYY).

3. If imputed CM end date is less than the CM start date, use the CM start date as the imputed CM end date.

Impute CM start date:

1. If the CM start date year value is missing, the imputed CM start date is set to one day prior to treatment start date.

2. If the CM start date year value is less than the treatment start date year value, the CM started before treatment. Therefore:

    a. If the CM month is missing, the imputed CM start date is set to the mid-year point (01JulYYYY).

    b. Else if the CM month is not missing, the imputed CM start date is set to the mid-month point (15MONYYYY).

3. If the CM start date year value is greater than the treatment start date year value, the CM started after treatment. Therefore:

    a. If the CM month is missing, the imputed CM start date is set to the year start point (01JanYYYY).

    b. Else if the CM month is not missing, the imputed CM start date is set to the month start point (01MONYYYY).

4. If the CM start date year value is equal to the treatment start date year value:

    a. And the CM month is missing or the CM month is equal to the treatment start date month, then the imputed CM start date is set to one day prior treatment start date.

    b. Else if the CM month is less than the treatment start date month, the imputed CM start date is set to the mid-month point (15MONYYYY).

    c. Else if the CM month is greater than the treatment start date month, the imputed CM start date is set to the month start point (01MONYYYY).

If complete (imputed) CM end date is available and the imputed CM start date is greater than the (imputed) CM end date, then imputed CM start date should be set to the (imputed) CM end date.

5.1.3 Prior therapies date imputation

See Section 5.1.2.

5.1.4 Post therapies date imputation

See Section 5.1.2.

5.1.5 First diagnosis date imputation

1. If first diagnosis date is complete then imputed first diagnosis date = first diagnosis date.
2. If only year is present in the first diagnosis date then do the following:

    a. If the year part of the first diagnosis date is equal to the year part of the inform consent date then imputed first diagnosis date = Jan 1 of the year of the first diagnosis date

    b. Otherwise imputed first diagnosis date = Jul 1 of the year of the first diagnosis date
3. If only month and year part are present in the first diagnosis date then do the following:

    a. If the month and year part of the first diagnosis date is equal to the month and year part of the inform consent date then imputed first diagnosis date = 1st of the month of the year of the first diagnosis date

    b. Otherwise imputed first diagnosis date = 15th of the month of the year of the first diagnosis date.

5.2 Statistical models

5.2.1 Analysis of continuous data

Mixed model repeated measures (MMRM)

Continuous variables analyzed via mixed model repeated measures (MMRM) assume that data are missing at random (MAR) assumption.

Treatment group, visit, baseline IGA and geographical region will be fitted as factors and baseline values will be fitted as continuous covariates. Treatment group by visit and visit by baseline value will be included as interaction terms in the model. A term for visit will be included in the repeated statement (in SAS PROC MIXED or PROC GLM) and an unstructured covariance matrix will be used thus allowing adjustment for correlations between time points within patients. The denominator degrees of freedom will be calculated using the Kenward-Roger method.

The model described above allows the relationship to baseline to vary for each visit, but assumes that the effects of baseline IGA and geographical region will remain constant over all visits. From this analysis, the adjusted means for each treatment group and at each timepointt, the difference between the adjusted means (ZPL389 groups vs placebo) at each timepoint, 95% confidence interval around these differences.

SAS code for mixed model:

```
proc mixed data=aaa;
class TRT USUBJID AVISITN IGA_BASELINE GEOGRAPHY;
```

```
model CHG=TRT AVISITN IGA_BASELINE GEOGRAPHY BASE TRT*AVISITN
    BASE*AVISITN
          / s ddfm=kr;
lsmeans TRT*AVISITN / diff cl;
repeated AVISITN / type=un subject=USUBJID;
run;
```

In case the MMRM model does not converge the following sequential steps will be used:

1. change ddfm=kr to ddfm=bw. If still no convergence, perform step 2.
2. change type=un to type=cs. If still no convergence, perform step 3.
3. remove covariates in the following order until convergence: GEOGRAPHY, BASE*AVISITN.

5.2.2 Analysis of binary data

Risk difference and 100*(1-α)% confidence interval

For an investigational drug group with n_1 subjects at risk, independent from the control group (e.g., placebo or comparator) with n_0 subjects at risk, of whom x_1 and x_0 experience a certain event, the risk difference is estimated as p_1-p_0 with $p_1 = x_1/n_1$ and $p_0 = x_0/n_0$.

Exact unconditional confidence limits for the risk difference will be obtained with SAS procedure PROC FREQ and option RISKDIFF in the TABLES statement, specifying the RISKDIFF option also in the EXACT statement.

Logistic regression

Where logistic regression is specified, the binary outcome variable will be evaluated using a logistic regression model with treatment regimen, baseline IGA, geographical region (if possible).

All p-values reported on linear hypotheses about regression coefficients will be based on the Wald tests from Type III analyses. In the SAS procedure PROC GENMOD, a Type III analysis will be performed by adding the model options: TYPE3, DIST=BIN, and LINK=LOGIT. Logistic regression will be applied to response variables at each visit, and the following steps will be performed:

1. Run the PROC GENMOD procedure;
2. If convergence not reached, remove the covariates from the model one by one until convergence is reached; start with removing region, followed by baseline IGA;
3. If convergence not reached, perform Fisher's exact test.

It should be noted that this model might not converge if response rates are too low.

SAS code for logistic regression:

```
proc logistic data=aaa;
class TRT STRATA / param=glm;
Model AVAL (event='1') = TRT WEIGHT STRATA;
```

LSMEANS TRT / diff cl exp ilink;

run;

5.2.3 Multiple imputations for response variables

Where multiple imputation is specified, the missing data will be imputed based on the individual treatment arm information.

Multiple imputation (MI) 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.

Missing values will be imputed simultaneously based on an underlying joint normal distribution and using a Markov Chain Monte Carlo (MCMC) method.

The imputations will be done separately for each treatment group including baseline IGA, geographical region, baseline EASI score as additional covariates.

The number of imputations will be set to 100, the seed for the random function will be set to 3892203 for this study. To generate the multiple imputed data sets, the SAS procedure MI can be used as follows:

The input data set should have one record per subject with baseline EASI score and IGA score as well as the variables to be imputed (e.g. changes from baseline in EASI, or post-baseline IGA score).

```
ODS OUTPUT MissPattern=msgpat VarianceInfo=varinfo ParameterEstimates=param;
PROC MI DATA=<easi_iga> OUT=<impdata> SEED=3892203  NIMPUTE=100;
 VAR  <baseline EASI> <baseline IGA>
     <change from baseline EASI week 1> - <change from baseline EASI week primary endpoint>
     <IGA week1> - <IGA week primary endpoint>;
  BY <treatment group>;
RUN;
```

Programming notes:

- The SAS procedure MIANALYZE expects a variable called "_IMPUTATION_" which is generated by the MI procedure. It might be needed to set the SAS option "VALIDVARNAME=UPCASE" temporarily in the program before the MI call, this option should be reset after the MIANALYZE call to VALIDVARNAME=V6.
- In case there are no missings in one treatment group, the MI procedure does not impute any values. In this case the corresponding data need to be imputed manually outside PROC MI and added to the dataset <impdata>.

The imputed data are saved in data set <impdata>. The outcomes of interest, i.e. the EASI 50/75 response and IGA 0 or 1 response will be calculated, e.g. as follows:

The treatment differences for each imputed data set will then be evaluated by Logistic regression and ODDS ratio as described in Section 5.4.2.2. This analysis will be done by _IMPUTATION_ for the comparison to the placebo treatment group. The model should be estimating response probability = 1 by using DESECENDING option. Using the ESTIMATE

option in the GENMOD procedure and the ODS OUTPUT data set “Estimates” provides the estimate for the odds ratio and confidence intervals.

```
PROC GENMOD <option>;
CLASS <stratum> <treatment>;
MODEL <response> = <explanatory variables> / link=logit dist=bin type3;
BY <by-variables);
ESTIMATE “OR. AIN 300 mg VS. Placebo” <treatment> 1 -1/exp;
ODS OUTPUT Estimates=Estimates;
RUN;
```

The MIANALYZE procedure expects the parameter estimate in the variables ESTIMATE, and the corresponding standard error in the variable STDERR. Measurements can be obtained from “Estimates” dataset by selecting the row with ODDS ratio estimates.

```
Data <modified dataset>;
  set Estimates;
  if substr(label,1,3)= “Exp”;
  ESTIMATE=LBetaEstimate;
  STDERR=StdErr;
  effect= “OR”;
  if missing(ESTIMATE) or missing(STDERR) then delete;
RUN;
```

The estimates and standard errors based on the 100 imputed data are then combined by applying Rubin’s rules for multiple imputed data sets, see Little and Rubin (2002).

Programming notes:

- The variables ESTIMATE and STDERR in the input data set for the MIANALYZE procedure may not be missing. Records with missing values need to be deleted and the variable _IMPUTATION_ needs to be renumbered and regenerated since for each by-group the procedure expects consecutive numbers starting at 1.
- The ESTIMATE and STDERR in terms of odds ratios from logistic regressions will be transformed to follow a normal distribution before MIANALYZE procedure. They will be transformed back to Odds Ratio to get the corrected ESTIMATE and corresponding CIs.

The SAS procedure MIANALYZE will be applied as follows:

Step 1:

```
DATA <modified dataset_t>;
     SET <modified dataset>;
     estimate=log(ESTIMATE);
     stderr=(log(LBETAUPPERCL)-log(LBETALOWERCL))/(2*1.96);
RUN ;

ODS LISTING CLOSE;
ODS OUTPUT ParameterEstimates=<results> VarianceInfo=<varinfo> ModelInfo=<modelinfo>;
PROC MIANALYZE PARMS=<modified dataset>;
  BY <by-variables>;
  MODELEFFECTS OR;
RUN;
```

```
ODS LISTING;

data <results_back>;
     set <results>;
     estimate=exp(ESTIMATE);
     LCLMEAN=estimate*exp(-1.96*stderr);
     UCLMEAN=estimate*exp(+1.96*stderr);
RUN ;
```