Statistical Analysis Plan	
Study Code	D3256C00001
Edition Number	2.0
Date	16May2022

A Phase 2 Multinational, Randomized, Double-blind, Parallel-group, 16-week Placebo-controlled Study with a 36–week Extension to Investigate the Use of Benralizumab for Patients with Moderate to Severe Atopic Dermatitis Despite Treatment with Topical Medications (The HILLIER Study)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
CCI [redacted]	
AD	Atopic dermatitis
ADA	Anti-drug antibodies
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
BMI	Body mass index
CDLQI	Children's Dermatology Life Quality Index
CI	Confidence intervals
CRF	Case report form
CSP	Clinical study protocol
CSR	Clinical study report
DAE	AE leading to discontinuation
DBL	Database lock
DLQI	Dermatology Life Quality Index
DRMI	Dropout reason-based multiple imputation
EASI	Eczema Area and Severity Index
eCRF	Electronic case report form
EOT	End of treatment
ePRO	Electronic patient-reported outcome
EQ-5D-5L	European Quality of Life-5 Dimensions
FAS	Full analysis set
GGT	Gamma-glutamyl transpeptidase
HADS	Hospital Anxiety and Depression Scale
LLOQ	Lower limit of quantification
ICF	Informed consent form
IGA	Investigator Global Assessment
IP	Investigational product
IPD	IP discontinuation
ITT	Intent-to treat

Abbreviation or special term	Explanation
IVRS	Interactive voice response system
IWRS	Interactive web response system
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMRM	Mixed model repeated measurements
MNAR	Missing not at random
NRS	Numeric Rating Scale
CCI [redacted]	
PGI-S	Patient Global Impression of Severity
PK	Pharmacokinetics
POEM	Patient Oriented Eczema Measure
PRO	Patient-reported outcome
PT	Preferred term
Q1	1st quartile
Q3	3rd quartile
Q4W	Every 4 weeks
Q8W	Every 8 weeks
RNA	Ribonucleic acid
SAE	Serious adverse event
SAP	Statistical analysis plan
SC	Subcutaneous
SCORAD	SCORing Atopic Dermatitis
SD	Standard deviation
SF-36v2	Short Form 36-item Health survey Version 2
SI	System international
CCI [redacted]	
SoA	Schedule of Activities
SOC	System organ class
TEAE	Treatment-emergent adverse event
TBL	Total bilirubin
ULN	Upper limit of the normal
VAS	Visual analogue scale

AMENDMENT HISTORY

Category*: Change refers to	Date	Description of change	In line with the CSP?	Rationale
N/A – first version	20 November 2020			
2.0	16 May 2022		Yes	Updates aligned with protocol amendment
2.0	16 May 2022		Yes	Updates from TFL shell development and from comments received after blinded data reviews of outputs and datasets.

* Pre-specified categories are
Primary or secondary endpoints; Statistical analysis method for the primary or secondary endpoints; Derivation of primary or secondary endpoints; Multiple Testing Procedure; Data presentations; Other

1 STUDY DETAILS

This statistical analysis plan (SAP) outlines the analyses to be generated for the global clinical study report (CSR). Additional analyses required for regional submissions will be prespecified in a separate analysis plan and will be submitted to the appropriate authorities.

Table 1 Objectives and Endpoints

Objectives	Estimand description / Endpoints
Primary	
To compare the clinical efficacy of benralizumab 30 mg with placebo in patients with AD despite treatment with topical medications [a].	• Population: Full analysis set • Endpoint: A binary response giving the proportion of patients with an IGA 0/1 and a decrease in IGA of ≥ 2 points at Week 16 relative to baseline • Intercurrent events: Patients who withdraw from the study will be considered as non-responders from the time of withdrawal up to Week 16. Patients who require rescue therapy from Day 29 up to Week 16 will be considered as non-responders from the time of this rescue use up to Week 16. • Summary measure: Odds ratio and difference in proportions between benralizumab and placebo at Week 16
Secondary	
To compare the effect of benralizumab with placebo on supportive measures of clinical efficacy in patients with AD despite treatment with topical medications [a].	• Key secondary endpoint [c]: proportion of patients with skin clearance (EASI-75) at Week 16 • Key secondary endpoint [c]: proportion of patients with an improvement of ≥ 4 points in peak pruritus weekly score at Week 16 • Key secondary endpoint [c]: proportion of patients with skin clearance (EASI-90) at Week 16 • Proportion of patients with skin clearance (EASI-50) at Week 16 • Proportion of patients with skin clearance (EASI-100) at Week 16 • Change from baseline in EASI score at Week 16 • Change from baseline in peak pruritus score at Week 2 • Change from baseline in POEM score at Week 16 • Change from baseline in SCORAD at Week 16
To compare benralizumab with placebo on patient-reported health-related quality of life measures in patients with AD despite treatment with topical medications [a].	• Change from baseline in DLQI and CDLQI at Week 16
To estimate the PK and immunogenicity of benralizumab 30 mg in in patients with AD despite treatment with topical medications [a].	• Serum benralizumab concentration • ADA

Table 1 Objectives and Endpoints

To compare long-term treatment with benralizumab 30 mg Q8W versus benralizumab 30 mg Q4W up to Week 52 in patients with AD despite treatment with topical medications [a].	• Change from baseline in EASI total score at Week 52. • Proportion of patients with an IGA 0/1 and a decrease in IGA of ≥ 2 points at Week 52 relative to baseline • Proportion of patients with EASI-75 at Week 52 • Other supportive efficacy assessments at Week 52 as appropriate.
Safety	
To compare the safety and tolerability of benralizumab with placebo in patients with AD despite treatment with topical medications [a].	Safety and tolerability will be evaluated in terms of AEs, Vital signs, and Clinical laboratory values. Assessments related to AEs cover: • Occurrence/frequency • Relationship to IP as assessed by Investigator • Intensity • Seriousness • Death • AEs leading to discontinuation of IP Vital signs parameters include systolic and diastolic blood pressure, and pulse, as well as respiration rate, body temperature, body weight, and height. Assessments related to vital signs cover: • Observed value • Absolute and percent change from baseline values over time
Tertiary/exploratory	
To explore the effect of benralizumab compared to placebo on healthcare resource utilization due to AD.	Rate of AD-related healthcare resource utilization during the study
To explore the effect of benralizumab compared to placebo on patient-reported health-related quality of life measures in patients with AD despite treatment with topical medications [a].	Change from baseline in scores for: • HADS • PGI-S • SF-36v2 Health Survey (PCS and MCS) • EQ-5D-5L Observed scores for: • CCI [redacted] • Patient-reported experience (free text entry)
CCI [redacted]	

Table 1 Objectives and Endpoints

CCI [redacted]

[a] The locally-approved regimen of topical medication.

[b] Rescue therapy is defined in CSP Section 6.5.3.

[c] The key secondary endpoints will use the same estimand as outlined for the primary endpoint. For all other endpoints, the estimands will be detailed in the SAP.

CCI [redacted], AD, Atopic dermatitis; ADA, anti-drug antibodies; AE, adverse event; CDLQI, The Children's Dermatology Life Quality Index; CSP, Clinical study protocol; DLQI, Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; EQ-5D-5L, European Quality of Life-5 Dimensions; HADS, Hospital Anxiety and Depression Scale; IGA, Investigator Global Assessment; IP, investigational product; CCI [redacted]; PGI-S, Patient Global Impression of Severity; PK, pharmacokinetics; POEM, Patient Oriented Eczema Measure; Q4W, every 4 weeks; Q8W, every 8 weeks; CCI [redacted] SAP, Statistical Analysis Plan; SCORAD, SCORing Atopic Dermatitis; SF-36v2, Short Form 36-item Health Survey Version 2; CCI [redacted]

1.1 Study design

1.1.1 Overall Design

This is a Phase 2 multinational, randomized, double-blind, parallel-group, 16-week placebo-controlled study with a 36–week extension. The study will evaluate the efficacy and safety of benralizumab 30 mg in male and female participants ≥ 12 years of age with moderate to severe atopic dermatitis (AD) who remain symptomatic despite treatment with standard of care treatment with topical medications. At this time adolescent participants will not be enrolled in sites in the United States. Adolescents may be recruited when and where approval is granted. The study is designed to assess the safety and efficacy of benralizumab 30 mg compared with placebo and identify the appropriate maintenance administration frequency (every 4 weeks (Q4W) versus every 8 weeks (Q8W)).

This study consists of the following consecutive periods:

- A 1- to 4-week run-in period, including a 7-day washout period of topical medications prior to randomization;
- A 16–week placebo–controlled, double-blind treatment period;

- A 36-week blinded-to-dosing regimen extension period for maintenance treatment.

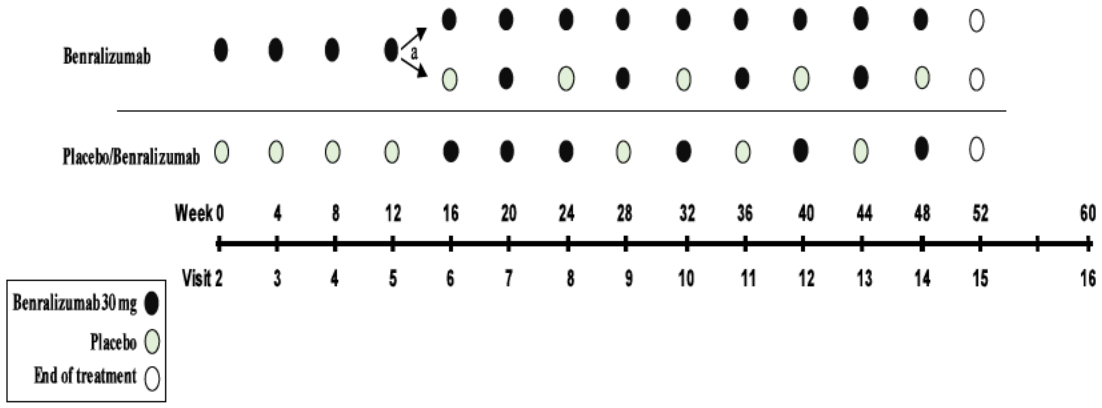
Following informed consent, all eligible participants will enter a run-in period of 1 to 4 weeks during which inclusion/exclusion criteria will be assessed, medical history taken, and complete physical exam will be conducted (Visit 1, CSP Table 3). Potentially eligible participants will enter a 7-day washout period during which all topical medications for AD except the stable emollient moisturizer regimen must be discontinued. Participants will be provided with a handheld device to respond to patient-reported outcome (PRO) questionnaires during the study.

Following the 1- to 4-week run-in period, including the 7-day washout period, a minimum of approximately 160 participants will be randomized at Visit 2, in a ratio of 1:1:2, respectively, to 1 of the following 3 treatment sequences (see also Figure 1 and Figure 2):

- Benralizumab 30 mg Q4W until Week 16, followed by an extension period with benralizumab 30 mg Q4W administered until Week 52 (n = 40);
- Benralizumab 30 mg Q4W until Week 16, followed by an extension period with benralizumab 30 mg Q8W administered until Week 52 (n = 40);
- Placebo Q4W until Week 16, followed by an extension period with benralizumab 30 mg Q4W administered until Week 28, and then benralizumab 30 mg Q8W administered until Week 52 (n = 80).

Figure 1 presents an overview of investigational product (IP) administration for each randomized treatment group.

Figure 1 Investigational Product Administration Q4W by Randomized Treatment Sequence



[a] In the extension phase, participants will receive benralizumab on a Q8W or Q4W dosing regimen, as predetermined at randomization (Visit 2). Participants randomized to receive benralizumab Q8W in the extension phase will also receive placebo at intervening study visits when they are not receiving benralizumab.

Q4W, every 4 weeks; Q8W, every 8 weeks.

All participants who prematurely discontinue IP (IPD) should return to the study site for the IPD visit 4 weeks (± 7 days) after the last dose of IP for procedures, or as soon as feasible if this interval is missed (eg, if decision on discontinuation was made later), as specified in Table 3 of the CSP.

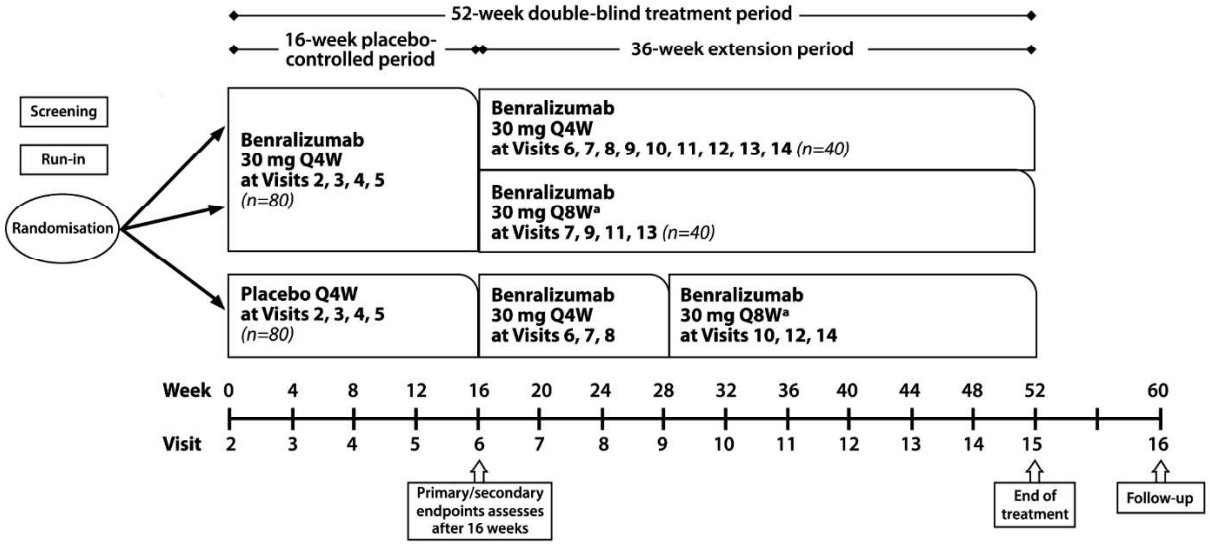
At the IPD visit, the participant will be offered the following options for further follow-up:

- Participants are encouraged to return to all scheduled site visits and perform all procedures but without IP administration, until the end of the study (Visit 16).
- Participants unwilling or unable to attend the scheduled site visits until the end of the double-blind treatment period, will be offered a follow-up option that includes monthly telephone contact instead. During follow-up telephone contact, the Investigator will collect information about concomitant medications, information on AD symptoms, and AE/SAE(s).

All Week 16 analyses at primary database lock (DBL) will evaluate the two initial dose treatment arms: benralizumab 30 mg (total n = 80), and placebo (total n = 80). Analyses of data for Week 52 end of study reporting will include data from the entire study, presenting the three randomized treatment groups: benralizumab followed by Q4W, benralizumab followed by Q8W and placebo followed by benralizumab Q4W-Q8W.

For an overview of the study design, see Figure 2. For further details see CSP Section 4.1.

Figure 2 Study Design



[a] To maintain blinding, participants will receive investigational product every Q4W during the extension period. Placebo will be administered Q8W, occurring 4 weeks after each benralizumab administration.

n, number; Q4W, every 4 weeks; Q8W, every 8 weeks.

Participants and Investigators will remain blinded to the treatment and regimen until the final DBL.

1.1.2 Investigational Product Dosing Regimen

Participants who meet entry criteria will be randomized 1:1:2 by the interactive voice/web response system (IxRS) to 1 of the 3 treatment sequences described in Section 1.1.

Blinded IP (benralizumab or placebo) will be administered Q4W at Weeks 0, 4, 8, and 12 (Figure 1).

From Week 16 onwards, participants initially randomized to benralizumab 30 mg will receive benralizumab 30 mg Q4W or Q8W interval, according to their original randomization at Visit 2 (Week 0) (Figure 1).

At Weeks 16, 20, and 24, participants initially randomized to placebo will receive benralizumab 30 mg Q4W. At Week 28, these participants will transition to receive benralizumab 30 mg Q8W interval (Figure 1).

All participants on Q8W regimen from Week 16 onwards will continue to receive treatment containing placebo at their Q4W visit intervals (Figure 1), thus maintaining the blinded treatment regimen.

The final dose of benralizumab or placebo will be administered at Week 48.

Please refer to the clinical study protocol (CSP) for more details.

1.2 Number of participants

Approximately 270 participants are expected to be enrolled/screened in order to achieve at least 160, and a maximum of 200, eligible study participants randomly assigned to study intervention to ensure that a broad distribution of participants is recruited across the range of ages and blood eosinophil levels to allow potential identification of responding subpopulations and appropriate cut-offs for future studies, if necessary.

Randomization will be stratified by blood eosinophils (< 300 cells/μL, ≥ 300 cells/μL) collected at screening and age at screening (≥ 12 to < 18 years and ≥ 18 years). Participants will be randomized in minimum cohort sizes across the baseline blood eosinophil and age subgroups; the intended distribution of participants in total across the treatment groups for each combination of these factors is summarized in Table 2. If the minimum recruitment target in any cohort is achieved, that cohort may remain open in order to fulfil minimum recruitment targets for the remaining cohorts. Thus, minimum recruitment targets may be exceeded in cohorts that remain open while fulfilling recruitment targets for other cohorts. Recruitment will stop when a maximum of 200 participants in total is achieved.

Table 2 Number of Participants Randomized by Stratification Factors

	Baseline blood eosinophils	
	<300 cells/µL	≥300 cells/µL
Adolescents (≥ 12 to < 18 years), n	20	30
Adults (≥ 18 years), n	40	70

For the primary analysis in the overall population, a minimum of 80 participants per treatment group will provide a high level of power (> 95%) to detect a 30% difference between benralizumab 30 mg Q4W and placebo for the primary endpoint. This calculation is based on a 2-sided test and a 5% significance level and assumes a response rate of 40% for benralizumab 30 mg Q4W and 10% for placebo.

For subgroup analyses, a minimum of 35 participants per treatment group will provide at least 80% power to detect a 30% difference between benralizumab 30 mg Q4W and placebo (using the same assumptions as for the primary analysis). Thus, there is a high probability of detecting differences between treatments at the 2-sided 5% level in the baseline blood eosinophils ≥ 300 cells/µL subgroup (with 50 participants per treatment) and in the adult subgroup (with 55 participants per treatment). In addition, there is a high chance of demonstrating statistical significance at the 2-sided 10% level in the adolescent subgroup (with 25 participants per treatment). Note that in the baseline blood eosinophils < 300 cells/µL subgroup, 30 participants per treatment group will provide an adequate amount of data to explore treatment differences, and will allow different cut-offs between 'low' and 'high' eosinophils at baseline to be assessed.

2 ANALYSIS SETS AND PERIODS

2.1 Definition of study analysis periods

The first set of analyses will be run after the primary DBL, which will occur after all randomized participants have completed the initial 16-week treatment period or discontinued prior to Week 16.

For Week 16 analyses, all data over the entire 16-week placebo-controlled treatment period (or last dose in placebo-controlled period + 30 days for those who do not enter the extension period) will be included in the analyses. Descriptive by-visit summaries and data listings will be presented for all available data for the entire study. Additionally, a limited number of key summaries may be presented for the entire 52-week treatment period including all available data.

Analyses for the first 16 weeks will compare the two initial treatment groups, benralizumab 30 mg Q4W Total and placebo Q4W. Summaries will also display the randomized treatment sequences (benralizumab 30 mg Q4W/Q4W, benralizumab 30 mg Q4W/Q8W).

The final set of analyses will be produced at the final DBL, which will occur when all participants have completed the 52-week treatment period and/or the IPD/EOT/Week 60 follow-up visit.

For Week 52 reporting, all data over the entire 52-week treatment period (and Week 60 follow-up visit and/or IPD/EOT visit) will be presented in the analyses, summaries, and data listings. Additionally, some summaries will be presented only on data from the extension period.

The start of the extension period is defined as starting after the date/time of first dose of IP in the extension period (Week 16 IP administration), ie assessments at Week 16 visit will be included into the Week 16 reporting.

Analyses for the Week 52 reporting will be presented by the 3 randomized treatment sequences (benralizumab 30 mg Q4W/Q4W, benralizumab 30 mg Q4W/Q8W and placebo/benralizumab). Summaries from baseline to Week 52 will also display the initial benralizumab 30 mg Q4W Total group.

2.2 Definition of analysis sets

Six analysis sets are defined below: all participant analysis set, full analysis set (FAS), safety analysis set, pharmacokinetic (PK) analysis set, extension period analysis set, and placebo-to-benralizumab extension analysis set. Participants must have provided their informed consent. If no signed informed consent is collected (important protocol deviation), then the participant will be excluded from all analysis sets defined below.

Note: "Enrolled" is defined as a participant's agreement to participate in a clinical study following completion of the informed consent process. “Randomized” participants are defined as those who undergo randomization and receive a randomization number. Potential participants who are screened for the purpose of determining eligibility for the study but are not randomized to the study will be considered “screen failures”, unless otherwise specified by the protocol.

2.2.1 All participants analysis set

This analysis set comprises all participants who sign the informed consent form (ICF) for the study, and will be used for the reporting of disposition and screening failures.

2.2.2 Full analysis set

The FAS will comprise of all randomized participants who receive at least one dose of IP, irrespective of their protocol adherence and continued participation in the study.

Participants will be analyzed according to their randomized treatment irrespective of whether or not they were prematurely discontinued, according to the intent-to treat (ITT) principle. Data for participants who withdraw consent to participate in the study will be included up to the date of permanent discontinuation.

All efficacy data will be summarized and analyzed using an ITT approach based on the FAS. For consistency, demographic and baseline characteristics will be presented using the FAS.

2.2.3 Safety analysis set

The safety analysis set will comprise of all participants who receive at least one dose of IP. Participants will be classified according to the treatment they actually received.

Erroneously treated participants (eg, those randomized to treatment A but actually given treatment B) will be included in the group of the treatment they received. A participant who receives at least one dose of active IP will be classified as active and included in the active IP treatment group.

For the extension period, participants will be reported in their randomized treatment groups. Any deviations from the randomized treatment will be listed and considered when interpreting the safety data. These will not be identified until routine unblinding after the primary or final DBL.

All safety summaries and anti-drug antibodies (ADA) analyses will be based on the safety analysis set.

2.2.4 Pharmacokinetic analysis set

The PK analysis set will comprise of all participants who received benralizumab and from whom PK blood samples are assumed not to be affected by factors such as protocol deviations (eg, received wrong dose) and who had at least one quantifiable serum PK observation post first dose. A final list of protocol deviations that would lead to exclusion from this analysis set will be determined prior to DBL but may need to be updated post unblinding.

All PK summaries will be based on the PK analysis set.

2.2.5 Extension period analysis set

The extension period analysis set will comprise of all participants who start or carry on receiving at least one dose of benralizumab after the end of the Week 16 placebo-controlled

period, and thus entering the extension period. For the extension period, participants will be reported in their randomized treatment groups.

2.2.6 Placebo-to-benralizumab extension analysis set

All participants who received benralizumab at the start of the extension period who were previously randomized to placebo treatment will be included in this analysis set.

2.3 Violations and deviations

Participants who do not meet eligibility criteria but are still randomized and received at least one dose of IP will be analyzed according to the analysis sets described in Section 2.2. There is no intention to perform a per-protocol analysis in this study.

2.3.1 Important protocol deviations

Important protocol deviations are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a participant's rights, safety, or well-being.

The final list of important protocol deviations will be documented prior to unblinding the study data (primary DBL and final DBL) and may include but not be limited to the following categories:

- Eligibility criteria not met (participants incorrectly randomized)
  - Deviations from key inclusion criteria
  - Deviations from key exclusion criteria
- Deviations from informed consent procedures
- Discontinuation criteria for IP met but participant not withdrawn from IP
- Deviations from IP management and administration
- Received prohibited/restricted concomitant medication
- Other important protocol deviations
  - Scheduled study assessments not done or incorrectly performed which could impact the primary or key secondary endpoints
  - Unblinding of treatment assignment for reasons unrelated to participant safety
  - Other severe non-compliance (such deviations will be clearly described in the CSR).

Only important protocol deviations will be summarized and listed in the CSR.

Protocol deviations will be detected by means of on-site monitoring (observable protocol deviations) and programmatically derived important protocol deviations. The process for identification and assessment will be detailed in a separate protocol deviation plan.

2.3.2 Impact on analyses due to COVID-19 pandemic

Given the uncertainty surrounding the future impact of the COVID-19 worldwide pandemic on clinical trials, operational procedures are being implemented in this study to maintain the integrity of collected data. Efforts may be made to collect data via alternative means where possible, when on-site visits cannot be performed.

If there is a sufficient number of protocol deviations or study disruptions as a result of COVID-19, then sensitivity analyses will be conducted to evaluate their impact on the interpretation of results. Protocol deviations, including doses or visits missed due to COVID-19 related protocol deviations will be described separately in the CSR.

COVID-19 disruptions are broadly defined as follows:

- Changes to visit schedules, missed visits, changes to study procedures;
- Discontinuation of IP or changes to the IP supply;
- The introduction of alternative monitoring approaches.

Confirmed or suspected cases of COVID-19 will be listed and included as adverse events (AEs) as appropriate. This detail will form part of the protocol deviation plan.

3 PRIMARY AND SECONDARY VARIABLES

3.1 General definitions

3.1.1 Baseline definition

Generally, for efficacy data, the last recorded value on or prior to the date of randomization will serve as the baseline measurement. If time is collected, the assessment performed the same day but at a time prior to randomization will be included in the baseline definition. This includes both, the Week 16 primary and the Week 52 final analyses.

Generally, for safety data, the last recorded value on or prior to the date of the first dose of IP will serve as the baseline measurement. If time is collected, the assessment performed the same day but at a time prior to the first dose of IP will be included in the baseline definition. This includes both, the Week 16 primary and the Week 52 final analyses.

For the Peak Pruritus Numeric Rating Scale (NRS), the average of the daily scores for the 7 days prior to the day of randomization will be the baseline. If there is no value prior to the date of randomization, then the baseline value will not be imputed and will be set to missing. Further details on scoring the Peak Pruritus NRS will be provided in Section 3.3.2.

For PROs with recall periods of 1 week or more (POEM, DLQI/CDLQI, HADS, PGI-S, CCI [redacted] EQ-5D-5L, SF-36v2), the time component of the definition of baseline (ie time of assessment relative to randomization) will not be considered; assessments on the same

day as randomization will be used to calculated baseline. If there is no assessment prior to randomization (or on the same day), then the baseline value will not be imputed and will be set to missing. Further details on scoring for PROs will be provided in Section 3.3.

When time of assessment is not recorded or missing, it is assumed that assessments recorded on the date of first dose of IP were performed prior to dosing, except in cases of protocol specified post-dose assessments. If there is no value prior to randomization (or the first dose of IP, depending on the endpoint), then the baseline value will not be imputed and will be set to missing. No data known to be collected post first dose will be used in determining the baseline value, unless otherwise specified.

For laboratory results and vital signs, baseline will be defined as the last non-missing assessment prior to first dose of IP. If no time is recorded for an assessment, and the assessment takes place at Visit 2, this will be assumed to be a pre-dose assessment. For physical examinations (complete or brief) baseline data will be collected at Visit 2 before first dose of IP.

Additional analyses for the participants who switch from placebo to benralizumab at Week 16 may be performed where the baseline value is set to the last recorded value prior to starting benralizumab (ie likely the Week 16 measurement) to obtain an assessment of the changes occurring while actually receiving benralizumab.

3.1.2 Change from baseline

Absolute change from baseline outcome variables are computed as:

(post-baseline value – baseline value).

Percent change from baseline outcome variables are computed as:

*((post-baseline value – baseline value) / (baseline value)) *100*

If either the post-baseline value or the baseline value is missing, then both change from baseline and percent change from baseline value will also be set to missing. If the baseline value is zero, the percent change will be set to missing.

Calculations of change from Week 16 will follow the same approach.

3.1.3 Visit windows

For the Peak Pruritus NRS, weekly summary scores will be calculated using the 7 daily scores for each week, starting with randomization day (e.g. Days 1 to 7 for Week 1, Days 8 to 14 for Week 2), regardless of visit attendance days.

For Peak Pruritus NRS Week 16 score, the 7 days prior to the scheduled Week 16 timepoint (ie Days 106 to 112) will be used for calculation. However, only days prior to the first dose of IP in the extension period will be used for the Week 16 score derivation. If the Week 16 visit, and therefore the start of extension period IP dosing, occurs ahead of schedule (ie at Day 112 or earlier), then the 7 days prior to this day of first dose of extension period IP will be used for the Week 16 score, and the Week 15 score will be set to missing. If the Week 16 score cannot be calculated on the 7 daily scores, then an additional 1-day shift to assess the 8 daily scores before extension IP dose day will be assessed, and the Week 15 score will be set to missing. For Week 17 score, the scheduled 7 days before day 120 (ie Days 113-119) will be used. If the Week 16 visit occurs at Day 113 or later, Days 106 to 112 will be used in the Week 16 score as planned. Further details on scoring for daily Peak Pruritus NRS will be provided in Section 3.3.2.

For efficacy endpoints that present visit-based data, visit windows will be applied. The adjusted analysis-defined windows will be based on the collection schedule listed in the protocol and variables will be windowed to the closest scheduled visit for that variable.

Visit windows following baseline will be constructed in such a way that the upper limit of the interval falls halfway between two visits (the lower limit of the first post-baseline visit will be study day 2). If an even number of days exists between two consecutive visits, then the upper limit will be taken as the midpoint value minus 1 day. Visit windows are constructed so that every observation collected can be allocated to a particular visit. No visit windows will be defined for screening visits.

For assignment of data to adjusted analysis-defined visit windows, study day will be defined as follows:

(Date of assessment – date of randomization) +1

Following this definition, for example if the day of randomization is study day 1 and date of assessment is the planned date of Visit 3 (Week 4), the study day will be 29 (28+1).

Any assessments performed after entering the extension period (ie the day after first dose of IP in the extension period at Week 16 visit) will be assigned to visit windows in the extension period and will not be assigned to visit windows in the 16-week placebo-controlled treatment period. Assessments at Week 16 visit will be assigned to Week 16.

If multiple assessments are recorded within a single adjusted visit window, please refer to the rules below:

- If there are 2 or more observations within the same visit window, then the non-missing observation closest to the scheduled study day will be used in the analysis.

- If 2 observations are equidistant from the scheduled study day, then the non-missing observation with the earlier collection date will be used in the analysis.
- If 2 observations are collected on the same day then the non-missing observation with the earlier collection time will be included in the analysis.

If a visit window does not contain any observations, then the data will remain missing.

For endpoints which are not collected at every visit (including PROs with recall periods of 1 week or more), the above rules will be applied to derive adjusted analysis-defined visit windows based on the protocol defined visit schedule for that endpoint.

Similarly, visit windows will be applied for laboratory parameters. Vital signs will be based on the case report form (CRF) visit designation.

When IP dosing needs to be postponed, it is recommended that all scheduled treatment visit procedures (except for IP administration) are still performed within the visit window. For further details see CSP Section 6.1.2.

3.1.4 Prior/concomitant medications

Medications will be categorized according to the WHO Drug Reference List dictionary which employs the Anatomical Therapeutic Chemical (ATC) classification system.

A medication will be regarded as prior if it was stopped on or before the date of randomization (medication stop date ≤ date of randomization).

A medication will be regarded as 'concomitant' if the start date is on or after the date of randomization or if it started prior to the date of randomization and was ongoing after the date of randomization. If a medication was started and stopped on the date of randomization, it will be considered as concomitant.

Concomitant medications for the 16-week placebo-controlled analysis reporting will include all concomitant medications with start date prior to entering the extension period.

Concomitant medications for the Week 52 end of study reporting over the entire study duration will include all concomitant medications. Concomitant medications reported for the extension period alone are those with start date after entering the extension period, and those ongoing from the 16-week placebo-controlled period.

The handling of partial/missing dates for assigning prior/concomitant medications is detailed in Appendix 8.3. Duration of prior/concomitant medications will not be calculated using imputed dates and will instead be set to missing. No other medication data will be imputed.

Throughout the study, participants will be required to maintain stable doses of their topical moisturizer for AD (CSP Section 6.5.1). If medically necessary for intolerable AD symptoms

that occur during the study, rescue therapy may be prescribed to participants at the discretion of the Investigator. For more detail see CSP Section 6.5.3.

Medication data will be presented for both, the 16-week placebo-controlled treatment period and the Week 52 reporting.

3.1.5 Intercurrent Events

Intercurrent events are withdrawal from the study at any time or requiring rescue therapy from Day 29 onwards. However, participants who started on placebo and switched to benralizumab in the extension phase are additionally permitted rescue therapy for 28 days from starting benralizumab, and thus additionally any rescue therapy use from Day 141 (ie 29 days after first dose of benralizumab) onwards will be considered as intercurrent event for these participants.

Withdrawal from the study will be considered from the day after withdrawal onwards, thus data on the day of withdrawal will be used as recorded. Rescue therapy from Day 29 (of first dose, and additionally for placebo-benralizumab participants from first benralizumab dose) onwards will be considered for intercurrent event imputation from the day of taking rescue medication onwards.

For responder or binary analyses, participants with intercurrent events will be considered as non-responders from the time these events occur. Participants with missing data for binary endpoints will also be considered non-responders at affected post baseline visits. Intercurrent events will be considered for analyses and descriptive summaries.

For change from baseline analyses or number and percentage of score category summaries, participants with intercurrent events will be considered as having missing scores from the time these events occur. Intercurrent events will be considered for analyses and descriptive summaries.

For summaries of number and percentage of individual questionnaire scores, all data will be presented as recorded, without imputation for intercurrent events.

The categorisation and imputations of responder status and reasons for non-response are provided in following table.

Table 3 Response Category Imputations

Response Category	Imputation Rule
Responder	Keep data as recorded, unless after taking rescue medication
Non responder	
Recorded as non-responder	Keep data as recorded, even if after taking rescue medication

Table 3 Response Category Imputations

Response Category	Imputation Rule
Missing	For missing questionnaire/visit, where participant remains in study and could be Responder or Recorded Non-responder or Rescue medication and remain in study
Rescue medication and remain in study	Impute if recorded as Responder and participant ongoing in study. Impute from day of rescue medication onwards, but only for assessments where participant is a Recorded Responder
Rescue medication and withdrawn from study	For participants who have taken rescue medication before withdrawal. Impute for post-withdrawal assessments
No rescue medication and withdrawn from study	For participants who have Not taken rescue medication before withdrawal. Impute for post-withdrawal assessments

Participants can move between different visits between categories of Responder, Recorded Non-responder and Missing. Participants can move between different visits between categories of Recorded Non-responder, Missing and Rescue medication and remain in study. After a participant has withdrawn, they will stay in the imputed withdrawal category.

3.2 Primary outcome variable

The proportion of participants with an Investigator Global Assessment (IGA) 0/1 and a decrease in IGA of ≥ 2 points at Week 16 relative to baseline will be used as the primary efficacy variable.

The IGA is an instrument used in clinical studies to rate the severity of AD globally, based on a 5-point scale ranging from 0 (clear) to 4 (severe). The IGA uses clinical characteristics of erythema, infiltration, papulation, oozing, and crusting as guidelines for the overall severity assessment. The validated IGA scale for AD will be completed by the Investigator on an electronic device supplied to the site according to the Schedule of Activities (SoA) (CSP Table 3) to capture these key measures of AD disease activity and describe the overall appearance of AD lesions (Table 4).

Table 4 Investigator Global Assessment Score

Score	Morphological description
0 – Clear	No inflammatory signs of atopic dermatitis (no erythema, no induration/papulation, no lichenification, no oozing/crusting). Post-inflammatory hyperpigmentation and/or hypopigmentation may be present.
1 -- Almost clear	Barely perceptible erythema, barely perceptible induration/papulation, and/or minimal lichenification. No oozing or crusting.
2 -- Mild	Slight but definite erythema (pink), slight but definite induration/papulation, and/or slight but definite lichenification. No oozing or crusting.
3 -- Moderate	Clearly perceptible erythema (dull red), clearly perceptible induration/papulation, and/or clearly perceptible lichenification. Oozing and crusting may be present.

Table 4 Investigator Global Assessment Score

Score	Morphological description
4 -- Severe	Marked erythema (deep or bright red), marked induration/papulation, and/or marked lichenification. Disease is widespread in extent. Oozing or crusting may be present.

A responder at Week 16 is defined as having IGA 0/1 and a decrease in IGA of ≥ 2 points at Week 16 relative to baseline.

In the statistical analyses, a binary variable will take on value 1 if a participant is defined as a responder, and value 0 for non-responder, and will be used as the response variable for the efficacy analysis.

Intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5). Intercurrent events will be considered for analyses and descriptive summaries.

3.3 Secondary outcome variables

3.3.1 Eczema Area and Severity Index (EASI)

The first and third key secondary efficacy variables, derived from Eczema Area and Severity Index (EASI), are the proportion of participants with skin clearance EASI-75 and EASI-90, respectively, at Week 16.

The EASI is a continuous outcome that will be assessed by the Investigator or designee at the visits specified in the SoA (CSP Table 3) on an electronic device supplied to the site. The EASI is a validated tool used in clinical practice and clinical trials to assess the severity and extent of AD (Hanifin et al 2001).

Severity of 4 AD disease characteristics (erythema, induration/papulation, excoriation (scratching), and lichenification) will each be assessed on a scale of 0 (absent) to 3 (severe) in each of 4 body regions (head/neck, trunk, upper limbs, and lower limbs). In addition, the area of AD involvement will be assessed as a percentage by body area in the 4 anatomic areas (head/neck, trunk, upper limbs, and lower limbs) and scored as 0 (no eruption), 1 (1% to 9%), 2 (10% to 29%), 3 (30% to 49%), 4 (50% to 69%), 5 (70% to 89%), or 6 (90% to 100%). The total score for each region is the severity score multiplied by the area score and a multiplier. The multipliers are as follows: head/neck=0.1, trunk=0.3, upper extremities=0.2, lower extremities=0.4. The total body total score is the sum of the region total scores. For each body region the severity scores and area of involvement scores are combined by the ePRO device into a domain score for each body region. The total EASI score will be calculated by the ePRO device by adding the four body region domain scores and ranges from 0 to 72. The ePRO does not allow for questions to be unanswered, so no consideration is required for missing items.

The percentage of participants achieving a 50% reduction from baseline in EASI total score (EASI-50), 75% reduction from baseline in EASI total score (EASI-75), 90% reduction from baseline in EASI total score (EASI-90), and 100% reduction from baseline in EASI total score (EASI-100) will be calculated from the total EASI scores.

The first and third key secondary endpoints are binary responses that classify a participant's skin clearance at Week 16 using the EASI:

- EASI-75 response rate: Participants will be classified as responders if they achieve at least a 75% reduction from baseline in their EASI total score at Week 16. Otherwise, participants will be classified as non-responders.
- EASI-90 response rate: Participants will be classified as responders if they achieve at least a 90% reduction from baseline in their EASI total score at Week 16. Otherwise, participants will be classified as non-responders.

In the statistical analyses, a binary variable will take on value 1 if a participant is defined as a responder, and value 0 for non-responder, and will be used as the response variable for the efficacy analysis.

Intercurrent events for responder analyses will be considered as described above in Section Intercurrent Events (Section 3.1.5). Participants with a missing baseline EASI total score will have missing change from baseline status and thus missing responder status at post baseline visits. Intercurrent events will be considered for analyses and descriptive summaries.

The proportion of participants with skin clearance EASI-50 and EASI-100, respectively, at Week 16 are secondary endpoints and derived as described for the key secondary endpoints above.

Changes from baseline in EASI total score at Week 16 will be analyzed as a secondary endpoint.

Changes from baseline in EASI total score at Week 52 will be analyzed. The proportion of participants with skin clearance EASI-75 at Week 52 will be analyzed.

For change from baseline analyses, intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5). Intercurrent events will be considered for analyses and descriptive summaries.

3.3.2 Peak Pruritus Numeric Rating Scale

The proportion of participants with an improvement of ≥ 4 points in peak pruritus weekly score at Week 16 will be used as the second key secondary efficacy variable.

The Peak Pruritus NRS is a one-item daily assessment of the worst itch the participant experienced over the past 24 hours. This is recorded every evening from Visit 1 onwards on the electronic device supplied to the participants. The score ranges from 0 to 10, with 0 being "no itch" and 10 being "worst itch imaginable." and so a reduction in score is considered an improvement.

If a participant has at least 4 non-missing daily scores in a 7-day week, the peak pruritus weekly score is defined as the sum of the available daily scores divided by the number of days with non-missing daily scores. If there are less than 4 non-missing daily scores, then the weekly score is missing for that week.

The second key secondary endpoint is a binary response that will classify a participant as a responder if there is an improvement of 4 or more points in peak pruritus weekly score at Week 16 relative to baseline. Otherwise, participants will be classified as a non-responder.

In the statistical analyses, a binary variable will take on value 1 if a participant is defined as a responder, and value 0 for non-responder, and will be used as the response variable for the efficacy analysis.

The change from baseline in peak pruritus score at Week 2 will be analyzed as a secondary endpoint.

Intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5). Participants with a missing baseline score will have missing change from baseline status. Intercurrent events will be considered for analyses and descriptive summaries.

3.3.3 Patient-Oriented Eczema Measure (POEM)

The change from baseline in Patient-Oriented Eczema Measure (POEM) score at Week 16 will be analyzed as a secondary endpoint.

The POEM is a continuous outcome that will be completed at the on-site visits specified in the SoA (CSP Table 3) on an electronic device supplied to the site.

The POEM is a 7-item assessment of AD symptoms (itchiness, sleep disturbance, bleeding, weeping, cracking, flaking, and dryness) that characterizes severity as a function of symptom frequency. The participant responds to each of the 7 questions by selecting the approximate number of days "over the last week" when they experienced the symptom, scored as 0 (no days), 1 (1 to 2 days), 2 (3 to 4 days), 3 (5 to 6 days), or 4 (every day).

The seven items are summed, resulting in a score ranging from 0 to 28 (Table 5). Higher scores represent more severe AD symptoms.

Table 5 Patient-Oriented Eczema Measure Score Bands

Score range	Clinical interpretation
0 to 2	Clear or almost clear
3 to 7	Mild atopic dermatitis
8 to 16	Moderate atopic dermatitis
17 to 24	Severe atopic dermatitis
25 to 28	Very severe atopic dermatitis

The ePRO does not allow for questions to be unanswered, so no consideration is required for missing items.

Intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5). Intercurrent events will be considered for analyses and descriptive summaries.

Changes from baseline in POEM score at Week 52 will be analyzed.

3.3.4 Scoring Atopic Dermatitis (SCORAD)

The change from baseline in SCORing Atopic Dermatitis (SCORAD) score at Week 16 will be analyzed as a secondary endpoint.

SCORAD is a continuous outcome that will be completed at the on-site visits specified in the SoA (CSP Table 3) on the electronic device supplied to the site.

SCORAD is a clinical tool for assessing the severity of AD that evaluates the extent and intensity of AD lesions, in addition to subjective symptoms (Kunz et al 1997). There are 3 components to the assessment: A = extent or affected body surface area, B = severity, and C = subjective symptoms. For component A; the extent of AD is assessed as a percentage of each defined body area (head/neck, anterior trunk, back, upper limbs, lower limbs and genitals) (see CSP Section 8.1.4) and is reported as the sum of all areas, with a maximum score of 100%. For component B, the severity of 6 specific symptoms of AD (redness, swelling, oozing/crusting, excoriation, skin thickening/lichenification, dryness) is assessed using the following scale: 0 (none), 1 (mild), 2 (moderate), or 3 (severe), giving a maximum of 18 total points. For component C, subjective assessments of itch and sleeplessness are recorded for each symptom by the participant or relative on a visual analogue scale (VAS), where 0 is no itch or sleeplessness and 10 is the worst imaginable itch or sleeplessness) respectively. The VAS scores are summed to give a maximum possible score of 20. The SCORAD total score is calculated by the ePRO device using an algorithm (A/5+7(B/2)+C). The maximum total score is 103, with higher values indicating more severe disease. The ePRO does not allow for questions to be unanswered, so no consideration is required for missing items.

Intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5). Intercurrent events will be considered for analyses and descriptive summaries.

Changes from baseline in SCORAD total score at Week 52 will be analyzed.

3.3.5 Dermatology Life Quality Index (DLQI)

The change from baseline in Dermatology Life Quality Index (DLQI) total score at Week 16 will be analyzed as a secondary endpoint.

The DLQI is a continuous outcome that will be completed at the on-site visits specified in the SoA (CSP Table 3) on the electronic device supplied to the site. The DLQI will only be completed by participants who are age 17 or older at Visit 1.

The DLQI is a 10-item assessment of dermatology-specific health-related quality of life in participants age 17 and older (Finlay and Khan 1994). Participants are asked to rate their symptoms and the impact of their symptoms "over the last week" on several domains: symptoms and feelings, daily activities, leisure, work and school, personal relationships, and treatment. Both the total score and domain scores will be calculated.

Each item (except question 7) is rated on a four-point Likert scale: 3 = very much, 2 = a lot, 1 = a little. 0 = not at all, 0 = not relevant, and 0 = question unanswered. Scoring question 7, the first part asks: 'Over the last week, has your skin prevented you from working or studying?' Scoring is for response of Not relevant = 0, and Yes = 3 (as 'prevention' is the biggest possible impact it is scored the maximum of 3). If response is No, a further question is asked: 'How much has your skin been a problem at work or studying', and scored as: A lot = 2, A little = 1, Not at all = 0.

The subscale domain scores are the sum of the individual question scores. If one question is left unanswered, the corresponding subscale should not be scored. The six domain scores are as follows:

Table 6 Dermatology Life Quality Index Domain Scores

Score range	Domains
0 to 6	Symptoms and feelings: Questions 1 (itchy, sore, painful or stinging) and 2 (embarrassed, self-conscious)
0 to 6	Daily activities: Questions 3 (interfered shopping or at home) and 4 (influenced the clothes)
0 to 6	Leisure: Questions 5 (affected social/leisure activities) and 6 (difficult to do any sport)
0 to 3	Work and school: Question 7 (prevented working or studying)
0 to 6	Personal relationships: Questions 8 (problems with partner, friends, relatives) and 9 (caused sexual difficulties)
0 to 3	Treatment: Question 10 (problem using treatment)

The domain subscales for each domain will be presented as a percentage of the maximum score (either 6 or 3).

The DLQI total score is calculated by summing the score of each question, resulting in a maximum of 30 and a minimum of 0. If one question is unanswered, this is allocated a score of 0 and the DLQI score summed in the usual way, out of 30. If two or more questions are unanswered, the questionnaire is not scored. The higher the score, the more quality of life is impaired.

The effects of life scores are calculated as follows from the total score:

- 0-1 = no effect at all on participant's life,
- 2-5 = small effect on participant's life,
- 6-10 = moderate effect on participant's life,
- 11-20 = very large effect on participant's life,
- 21-30 = extremely large effect on participant's life.

Intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5). Intercurrent events will be considered for analyses and descriptive summaries.

Changes from baseline DLQI total score at Week 52 will be analyzed.

3.3.6 Children's Dermatology Life Quality Index (CDLQI)

The change from baseline in Children's Dermatology Life Quality Index (CDLQI) total score at Week 16 will be analyzed as secondary endpoint.

The CDLQI is a continuous outcome that will be completed at the visits specified in the SoA (CSP Table 3) on the electronic device supplied to the site. The CDLQI will only be completed by participants who are age 16 or younger at Visit 1 in regions where adolescent recruitment is permitted.

The CDLQI is a 10-item assessment of dermatology-specific health-related quality of life in children and adolescents ages 5 to16 (Lewis-Jones and Finlay 1995). Participants are asked to rate their symptoms and the impact of their symptoms "over the last week" on several domains: symptoms and feelings, leisure, school and holidays, personal relationships, sleep, and treatment.

Each item is rated on a four-point Likert scale: 3 = very much, 2 = quite a lot, 1 = only a little. 0 = not at all, 0 = question unanswered, and for Question 7 additionally 3 = prevented school. Both the total score and domain scores will be calculated.

The subscale domain scores are the sum of the individual question scores. If one question is left unanswered, the corresponding subscale should not be scored. The six domain scores are calculated as follows:

Table 7 Children's Dermatology Life Quality Index Domain Scores

Score range	Domains
0 to 6	Symptoms and feelings: Questions 1 (scratchy, sore, painful) and 2 (embarrassed, self-conscious, upset, sad)
0 to 9	Leisure: Questions 4 (worn different or special clothes/shoes), 5 (affected going out, playing, doing hobbies) and 6 (avoided swimming or other sports)
0 to 3	School or holidays: Question 7 (affected school work/holiday)
0 to 6	Personal relationships: Questions 3 (affected your friendships) and 8 (calling you names, teasing, bullying, asking questions or avoiding you)
0 to 3	Sleep: Question 9 (sleep been affected)
0 to 3	Treatment: Question 10 (problem with treatment)

The CDLQI total score ranges from 0-30. The total score is the sum of the individual question scores. 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.

Meaning of total score:

- 0-1 = no effect on child's life,
- 2-6 = small effect child's life,
- 7-12 = moderate effect child's life,
- 13-18 = very large effect child's life,
- 19-30 = extremely large effect child's life.

Intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5). Intercurrent events will be considered for analyses and descriptive summaries.

Changes from baseline CDLQI total score at Week 52 will be analyzed.

3.3.7 Serum benralizumab concentration

All PK samples will be collected before administration of study medication in accordance with the visit schedule provided in Table 3 of the CSP. For the PK analysis, it is important that the date, time, and location of each subcutaneous (SC) injection is recorded for each participant.

Concentration results below the lower limit of quantification will be set to lower limit of quantification (LLOQ)/2 for analysis and will be listed as <LLOQ.

Drug concentration information that would unblind the study will not be reported to investigative sites or blinded personnel until the study has been unblinded.

All available data will be included, regardless of intercurrent events like taking rescue medication or withdrawal from study.

3.3.8 Immunogenicity

Samples for Anti-drug antibodies (ADA) will be collected before administration of study medication according to the SoA (CSP Table 3). ADA variables, such as ADA responses, will be generated and analyzed as per the details in Appendix 8.2.

ADA information that would unblind the study will not be reported to investigative sites or blinded personnel until the study has been unblinded.

All available data will be included, regardless of intercurrent events like taking rescue medication or withdrawal from study.

3.3.9 Investigator Global Assessment (IGA) at Week 52

The IGA is defined for the primary endpoint variable above (Section 3.2).

The proportion of participants with an IGA 0/1 and a decrease in IGA of ≥ 2 points at Week 52 relative to baseline will be analyzed.

Intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5). Participants with a missing baseline IGA score will have missing responder status at post baseline visits. Intercurrent events will be considered for analyses and descriptive summaries.

3.4 Safety outcome variables

Safety and tolerability will be evaluated in terms of: reported AEs (including serious adverse events (SAEs)), vital signs, physical examination, electrocardiogram and clinical laboratory assessments.

All safety measurements will use all available data for analyses, including data from unscheduled visits and repeated measurements.

Change from baseline to each post-treatment time point where scheduled assessments were made will be calculated for relevant measurements. Safety data will be summarized by means of descriptive statistics and qualitative summaries.

The handling of partial/missing dates for classifying AEs is detailed in Appendix 8.3. Duration of AEs will not be calculated using imputed dates and will instead be set to missing. No other safety data will be imputed.

3.4.1 Adverse Events

AEs (including SAEs) experienced by the participants will be collected throughout the entire study and will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA) that will have been released for execution at AstraZeneca/designee per the Data Management Plan.

AEs for the 16-week placebo-controlled analysis reporting will include all AEs with start date prior to entering the extension period.

AEs for the Week 52 end of study reporting over the entire study duration will include all AEs as defined below. AEs reported for the extension period alone are those with start date on or after the date of first dose of treatment during the extension period.

As per CSP Section 9.4.3, the following events are considered treatment emergent:

- AEs with an onset date on or after the first dose of IP;
- Worsening of pre-existing events on or after first dose of IP.

For the analysis, an AE will be defined as Treatment Emergent AE (TEAE) if it has an onset date on or after the first dose of IP. Applying this definition, also the worsening of pre-existing events on or after the first dose of IP will be included in the analysis as Investigators are instructed to record these as new events.

AE data will be categorized according to their onset date into the following study periods:

- AEs in the pre-treatment period are defined as those with an onset prior to the day of first dose of IP.
- AEs in the on-treatment period for 16-week placebo-controlled period are defined as those with onset date between day of first dose of IP and the day prior to first dose in extension period; or the last dose in placebo-controlled period + 30 days for those who do not enter the extension period.
- AEs in the on-treatment period for extension period only are defined as those with onset date between day of first dose of IP in the extension and the last dose + 30 days.
- AEs in the on-study period for Week 52 end of study reporting are defined as those with onset on or after the day of first dose of IP.

If an AE has a missing onset date then unless the stop date of the AE indicates otherwise, this will be considered an on-treatment AE. Similarly, if an AE has a partial onset date, then unless

the partial onset date or the stop date indicates otherwise, this will be considered an on-treatment AE.

AEs that have missing causality (after data querying) will be assumed to be related to study drug.

3.4.2 Clinical laboratory variables

Blood and urine samples for determination of clinical chemistry, hematology and urinalysis parameters will be taken at the times detailed in the CSP and will be assessed in a central laboratory. The parameters outlined in Section 8.2.1, Table 10 of the CSP will be collected.

In summaries, figures, and listings, lab results and normal ranges will be presented in System International (SI) units. Eosinophil data will be presented in both SI and conventional units (cells/µL) in summaries.

Changes in hematology and clinical chemistry variables between baseline and each post-baseline assessment will be calculated. For values recorded with a leading greater than or less than (‘>’, ‘<’) symbol, the reported numeric value will be used for analysis and the value with the symbol will be included in the listings, unless otherwise specified. For example, a value of <0.01 will be analyzed as 0.01 and listed as <0.01.

Absolute values will be compared to the relevant reference range and classified as low (below range), normal (within range or on limits) or high (above range). The central reference ranges will be used for laboratory variables. All absolute values falling outside the reference ranges will be flagged.

Maximum or minimum value post-baseline will be calculated over the on-treatment period. This will be defined at both reporting occurrences, the 16-week placebo-controlled analysis reporting and the Week 52 end of study reporting.

For the liver function tests: aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), gamma-glutamyl transpeptidase (GGT) and total bilirubin (TBL), the multiple of the central laboratory upper limit of the normal (ULN) range will be calculated for each data point.

$$Multiple = Value / ULN$$

That is, if the ALT value was 72 IU/L (ULN=36) then the multiple would be 2.

Urinalysis data with qualitative results will be categorized as negative (0), positive (+), or strongly positive (++, +++, or >+++) at each timepoint.

3.4.3 Vital signs, physical examination and electrocardiogram

Pre-dose vital signs (pulse, systolic blood pressure, diastolic blood pressure, respiration rate, body temperature, body weight, and height) will be obtained in accordance with the schedule provided in Table 3 of the CSP. Body temperature will be measured in Celsius in accordance with local standards; weight will be recorded in kilograms.

Body mass index (BMI) will be calculated from the height and weight as follows:

$$BMI\ (kg/m^2) = Weight\ (kg) / (Height\ (m))^2$$

Changes in vital signs variables between baseline and each subsequent scheduled assessment will be calculated.

Absolute values will be compared to the reference ranges in Table 8 and classified as low (below range), normal (within range or on limits) or high (above range). All values (absolute) falling outside the reference ranges will be flagged.

Table 8 Vital signs reference ranges

Parameter	Standard Unit	Lower Limit	Upper Limit
Diastolic Blood Pressure	mmHg	60	120
Systolic Blood Pressure	mmHg	100	160
Pulse Rate	Beats/min	40	120
Respiratory Rate	Breaths/Min	8	28
Body Temperature	Celsius	36.5	38
Weight	kg	35	200

Physical examinations (complete or brief), including height and weight, will be conducted in accordance with the schedule provided in Table 3 of the CSP. Any new findings or aggravated existing abnormalities, judged as clinically significant by the Investigator, will be reported as an AE as described in CSP Section 8.3.5.

The complete physical examination will include an assessment of the following: general appearance, skin, head and neck (including eyes, ears, nose, mouth, and throat), lymph nodes, abdomen, musculoskeletal (including spine and extremities), cardiovascular, respiratory, and neurological systems. The brief physical examination will include an assessment of the general appearance, abdomen, cardiovascular, and respiratory system. For both types of physical examination, only information on whether the assessment was performed or not is to be recorded.

Single 12-lead electrocardiogram (ECG) will be obtained locally during screening (Table 3 of CSP) using an ECG machine that automatically calculates the heart rate and measures PR,

QRS, and QT intervals. The ECG results will be interpreted locally for the overall interpretation (normal/abnormal) and determination of clinical significance of any potential ECG findings (yes/no) and abnormal reason, which will be recorded in the eCRF.

3.5 Tertiary/exploratory outcome variables

3.5.1 Healthcare resource utilisation

At randomization, retrospective AD-related healthcare resource utilization information will be collected with a one-year recall period. At all subsequent visits (see CSP Table 3), AD-related healthcare resource utilization information will be collected with a recall period of 'since last scheduled visit'.

The data collected may be used to conduct exploratory economic analyses and include:

- Number and duration of hospitalization (length of stay, overall and by wards [eg, general ward, intensive care unit]);
- Number and type of diagnostic and therapeutic tests and procedures;
- Number of outpatient medical encounters and interventions (including GP, specialists, or emergency room visits, and medications).

Note: Cases of hospitalization occurring after signing of the ICF must be reported as an SAE (CSP Section 8.3).

All available data will be included, regardless of taking rescue medication or withdrawal from study.

Healthcare resource utilisation will be summarized descriptively. Any additional analyses will be performed separately and not form part of this SAP.

3.5.2 Hospital Anxiety and Depression Scale (HADS)

The Hospital Anxiety and Depression Scale (HADS) will be completed at the on-site visits specified in the SoA (CSP Table 3) on the electronic device supplied to the site.

The HADS is a mental health assessment comprising 7 items on anxiety and 7 items on depression. All questions are rated on a 0 to 3 scale, with higher scores representing worse mental health; scores each of for the 2 subscales thus range from 0 to 21. Scores between 0 and 7 are considered normal, scores between 8 and 10 are possible cases of anxiety/depression, scores 11 or greater are considered likely cases of anxiety/depression. The ePRO does not allow for questions to be unanswered, so no consideration is required for missing items.

Intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5). Intercurrent events will be considered for analyses and descriptive summaries.

The change from baseline in HADS anxiety and depression scores will be analyzed at Week 16 and at Week 52.

3.5.3 Patient Global Impression of Severity (PGI-S)

The Patient Global Impression of Severity (PGI-S) will be completed at the on-site visits specified in the SoA (CSP Table 3) on the electronic device supplied to the site.

The PGI-S is a single item assessment of the participant's perception of overall symptom severity "over the past 7 days". The assessment uses a 6-point categorical response scale (0 = no symptoms, 1 = very mild, 2 = mild, 3 = moderate, 4 = severe, 5 = very severe). The ePRO does not allow for questions to be unanswered, so no consideration is required for missing items.

All available data will be included, regardless of intercurrent events like taking rescue medication or withdrawal from study.

The PGI-S will be summarized descriptively.

3.5.4 SF-36 (Version 2), Acute Recall

The Short Form 36-item Health Survey Version 2 (acute recall) (SF-36v2) will be assessed at on-site Visits 1, 2, 6, 7, 12, 16, and IPD visit, and at home every 28±3 days after Visit 2 until Visit 6 is confirmed, on the electronic device supplied to the participants.

SF-36v2 is a 36-item questionnaire on functional health and well-being. The recall period is 1 week. Responses to 35 of the 36 items are used to compute 8 domain scores and 2 component summary measures. The remaining item (the "Health Transition" item) asks participants to rate how their current state of health compares to their state of health 1 week ago. The Heath Transition item is not used to calculate the domain scores.

The 8 domains are: Physical Functioning (PF), Role Limitations due to Physical Health (RP), Bodily Pain (BP), General Health Perceptions (GH), Vitality (VT), Social Functioning (SF), Role Limitations due to Emotional Problems (RE), and Mental Health (MH). The component summary measures (physical component summary [PCS] and mental health component summary [MCS]) are computed from domain scores to give a broader metric of physical and mental health-related quality of life. The transformed score range for each of the 8 domains and for PCS and MCS is 0-100; higher scores indicate better health state. The ePRO device does not allow for questions to be unanswered, so no consideration is required for missing items.

The change from baseline in SF-36v2 (PCS and MCS) scores will be calculated.

The threshold values for the SF-36v2 PCS, MCS, and domain scores listed in Table 9 are suitable for interpreting change at the participant level and are referred to as the responder thresholds or responder definitions (Maruish 2011).

Table 9 Threshold Values for the SF-36v2 Scale and Summary Measures

Threshold	SF-36v2 score									
	PCS	MCS	PF	RP	BP	GH	VT	SF	RE	MH
Group difference	2	3	3	3	3	2	2	3	4	3
Individual change	3.4	4.6	4.3	3.4	6.2	7.2	6.2	6.9	4.5	6.2

BP, Bodily Pain; GH, General Health Perceptions; MCS, mental health component summary; MH, Mental Health; PCS, physical component summary; PF, Physical Functioning; RE, Emotional Problems; RP, Role Limitations due to Physical Health; Short Form 36-item Health Survey Version 2; VT, Vitality.

Categorical analyses of the change in each of the SF-36v2 scales and summary measures will be compared to its respective individual change threshold (see Table 9) to indicate whether each participant experienced a change deemed clinically significant and important. A participant will be classified as a responder if the change from baseline ≥ individual change threshold, or a non-responder if change from baseline < individual change threshold. If data are missing, then the participant will be classified as a non-responder.

Intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5). Intercurrent events will be considered for analyses and descriptive summaries.

The change from baseline in SF-36v2 PCS and MCS will be analyzed at Week 16, and at Week 52.

CCI

CCI

[redacted]

3.5.6 Patient-reported Experience

Participants will be asked to respond in writing (free text) to 3 open-ended questions about their experience with AD and their study intervention. The assessment will be conducted during the site visits specified in the SoA (CSP Table 3), using a provisioned tablet to access the secure web form.

Data from the free text collection will be used for exploratory descriptive analysis using machine learning technologies. Due to the qualitative nature of the data and the analysis, the results will be presented in a separate report (not in the CSR) and the data will not be entered into the study database.

The Patient-reported Experience will be analyzed separately and not form part of this SAP.

3.5.7 Body Surface Area Involvement of Atopic Dermatitis

Body surface area affected by AD will be assessed for each section of the body using the Palmar method with rule of nine. The possible highest score for each region is: head and neck [9%], anterior trunk [18%], back [18%], upper limbs [18%], lower limbs [36%], and genitals [1%], and will be reported as a percentage of all major body sections summed together. Participants will undergo this assessment at time points according to the SOA (CSP Table 3). The body surface area affected scoring will also be calculated on the electronic device that is supplied to the site. The ePRO does not allow for questions to be unanswered, so no consideration is required for missing data.

Intercurrent events will be considered as described above in Section Intercurrent Events (Section 3.1.5) Intercurrent events will be considered for analyses and descriptive summaries.

3.5.8 European Quality of Life-5 Dimensions (EQ-5D-5L)

The European Quality of Life-5 Dimensions (EQ-5D-5L) questionnaire will be assessed at on-site Visits 1, 2, 6. 7, 12, 16, and IPD visit, and at home every 28±3 days after Visit 2 until Visit 6 is confirmed, on the electronic device supplied to the participants.

The EQ-5D-5L assesses 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 5 response options (no problems, slight problems, moderate problems, severe problems, and extreme problems) that reflect increasing levels of difficulty.

The participant will be asked to indicate his/her current health state by selecting the most appropriate level in each of the 5 dimensions. The questionnaire also includes a VAS, where the participant will be asked to rate current health status on a scale of 0 to 100, with 0 being the worst imaginable health state.

The European Quality of Life-5 Dimensions (EQ-5D-5L) will be analyzed separately and not form part of this SAP.

CCI

[redacted]

CCI

[redacted]

3.5.11 Biomarkers

CCI

[redacted]

3.5.12 Clinical skin photography

Clinical skin photography is a method to capture and document images of the skin. To evaluate alternative methods of clinical assessments, skin photography will be used in this study. Full body photographs will be taken as outlined in the photography manual and according to the SoA (CSP Table 3). At visits when clinical skin photography is to be performed and assessed before the Investigator sees the participant for the study visit, the photographs should be taken by a trained site personnel and not by the Investigator performing efficacy assessments on the participant. The site Investigator will then review the photographs and perform the clinical efficacy assessments (IGA, EASI, SCORAD, and BSA) based on review of the photographs. The Investigator photograph assessment will be recorded in the electronic case report form (eCRF). The same site Investigator should then see the participant and perform the usual in-person visit assessments including the in-person assessment of efficacy parameters, and document the data. The first assessments (based on skin photography) should not be altered even if differences are observed after assessing the participant in person. The final in-person Investigator assessment documented should reflect

the Investigator's overall clinical judgment. Photographic equipment will be provided to the sites.

The clinical efficacy assessments (IGA, EASI, SCORAD, and BSA) are described for primary and secondary endpoints in Sections 3.2 and 3.3 above. For these exploratory analyses of agreement between the in person and photograph assessment, the total scores or matching sub-scores will be used.

All available data will be included, regardless of intercurrent events like taking rescue medication or withdrawal from study.

The photographic assessment results will be examined against the in-person assessment results at Week 16 analyses.

4 ANALYSIS METHODS

The primary DBL is targeted to occur when all participants have completed the 16-week placebo-controlled treatment period. The final DBL will occur when all participants have completed the 52-week treatment period and the Week 60 follow-up visit and/or the IP discontinuation (IPD) /EOT visit.

All data will be included in the analyses, regardless of whether IP was prematurely discontinued or delayed, and irrespective of protocol adherence, unless the participant withdrew consent to further study participation.

All personnel involved in the analyses of the study will remain blinded until the primary DBL and protocol deviations are identified.

The CSR will be based on the final DBL.

4.1 General principles

The primary efficacy analyses will be based on the double-blind, 16-week placebo-controlled treatment period. Final analyses will be based on Week 52 end of study reporting. Efficacy endpoints will be analyzed using the FAS, the analysis of safety endpoints will be based on the safety analysis set. Analysis sets are defined in Section 2.2.

For the primary Week 16 analyses of binary endpoints, assessments after the use of rescue medication from Day 29 onwards will be considered as non-responders from the point of the rescue medication onwards. Any missing visits (including data after withdrawal of study at any time) will also be considered as non-responders, as described in Table 11 below.

For the primary Week 16 analyses of continuous repeated measures endpoints, any data after the use of rescue medication from Day 29 onwards or after withdrawal from study at any time will be treated as missing and the MMRM analyses fitted on the remaining available data.

For Week 52 analyses, for participants on benralizumab from randomization, rescue therapy use from Day 29 onwards will result in the participants being considered non-responders from the point of rescue use onwards. However, for participants who started on placebo and switched to benralizumab in the extension phase, rescue therapy users will only be considered as non-responders for the Week 52 analyses if it was taken from Day 141 (ie 29 days after first dose of benralizumab) onwards. Participants after withdrawal of study at any time will be considered as non-responders. Similarly, for change from baseline analyses, those intercurrent events will be considered as missing data.

Summary data will be presented in tabular format by treatment group. Categorical data will be summarized using frequencies and percentages, where the denominator for calculation is the underlying analysis set population unless otherwise stated. For efficacy reporting, the proportions of responders presented over time will include a number in analysis row to show the number of participants who have had the opportunity to reach each timepoint by the primary analysis based on their randomisation date, and this number in analysis will be used as the denominator for each visit.

Continuous data will be summarized with descriptive statistics of number of available observations, mean, standard deviation (SD), median, minimum and maximum, and quartiles where more appropriate. Mean and median will be reported to one decimal place more than the data recorded, the SD to one decimal place more than the mean, other items to the precision recorded (maximum rounded to 4 decimal places). Data listings will be sorted by treatment and participant number.

All point estimates will be presented together with 95% confidence intervals (CIs) as measures of precision and nominal 2-sided p-values, rounded to 4 decimal places.

For Week 16 analyses, all data over the entire 16-week placebo-controlled treatment period (or last dose in placebo-controlled period + 30 days for those who do not enter the extension period) will be presented in summaries and data listings (see Section 2.1).

For Week 52 end of study reporting, all data over the entire 52-week treatment period (and Week 60 follow-up visit and/or IPD/EOT visit) will be presented in summaries and data listings (see Section 2.1).

Analyses for the first 16 weeks will compare the two initial treatment groups, benralizumab 30 mg Q4W Total and placebo Q4W. Summaries will also display the randomized treatment sequences (benralizumab 30 mg Q4W/Q4W, benralizumab 30 mg Q4W/Q8W).

Analyses for the Week 52 end of study reporting will be presented by the 3 randomized treatment sequences (benralizumab 30 mg Q4W, benralizumab 30 mg Q8W and placebo/benralizumab), as noted in Section 1.1.1. Summaries will also display the initial benralizumab 30 mg Q4W Total group. For efficacy analyses, the Q4W versus Q8W dosing regimen will be statistically compared for participants originally randomized to benralizumab 30 mg Q4W initial dose, with all data (including placebo/benralizumab group) included in the model.

Additionally, a small number of key efficacy and safety tables will also be presented for data in the extension period alone to explore where a direct comparison of Q4W versus Q8W dosing can be made for participants previously receiving benralizumab, and to summarize time on benralizumab alone for participants previously receiving placebo. Adverse events will be presented for each period separately and only overall for key tables.

The data analyses will be conducted using the SAS® System (SAS Institute Inc., Cary, NC). All SAS® programs used to generate analytical results will be developed and validated according to AstraZeneca SAS® programming standards and validation procedures.

Table 11 Primary and key secondary efficacy and main safety estimands

Statistical Category	Estimand[a]				Section
	Treatment Condition[a]	Endpoint (Population)	Intercurrent Event Strategy[a]	Population Level Summary[a] (Analysis)	
Primary Objective: To compare the clinical efficacy of benralizumab 30 mg with placebo in patients with AD despite treatment with topical medications					
Primary	Treatment with benralizumab versus placebo, regardless of compliance, where rescue from Day 29 indicates treatment failure	• A binary response giving the proportion of patients with an IGA 0/1 and a decrease in IGA of ≥ 2 points at Week 16 relative to baseline (FAS)	• All data up to Week 16 will be included • Treatment discontinuation – treatment policy • Rescue medication from Day 29 onwards – non-responders • Withdrawal from study – non-responders	• Odds ratio, with corresponding 95% CIs (logistic regression)	4.2.6.1
Key secondary objective: To compare the effect of benralizumab with placebo on supportive measures of clinical efficacy in patients with AD despite treatment with topical medications					
Key secondary	Treatment with benralizumab versus placebo, regardless of compliance, where rescue from Day 29 indicates treatment failure	• Proportion of patients with skin clearance (EASI-75) at Week 16 (FAS) • Proportion of patients with an improvement of ≥ 4 or more points in peak pruritus weekly score at Week 16 (FAS) • Proportion of patients with skin clearance (EASI-90) at Week 16 (FAS)	• All data up to Week 16 will be included • Treatment discontinuation – treatment policy • Rescue medication from Day 29 onwards – non-responders • Withdrawal from study – non-responders	• Odds ratio, with corresponding 95% CIs (logistic regression)	4.2.7.1, 4.2.7.2
Safety Objective: To compare the safety and tolerability of benralizumab with placebo in patients with AD despite treatment with topical medications					
Safety	Treatment with benralizumab versus placebo, regardless of rescue medication	Presence of TEAEs through EOT (safety) Presence of serious TEAEs through EOT (safety)	Remained adherent to intervention (on-treatment)	• Categorical descriptive	4.2.8

[a] All estimand attributes explicitly identified for primary and key secondary estimands only.

4.1.1 Testing strategy to account for multiplicity considerations

A hierarchical testing procedure will be used to control the overall Type I error rate at 5% across the primary and key secondary endpoints

Hypothesis testing for the primary and key secondary endpoints (Sections 4.2.6.1, 4.2.7.1, 4.2.7.2) will be sequentially in the following order and based on a 2-sided test and carried out using a 5% significance level:

1 Proportion of participants with IGA 0/1 (ie, clear or almost clear) and a decrease in IGA of 2 or more points at Week 16, relative to baseline;
2 Proportion of participants with EASI-75 at Week 16;
3 Proportion of participants with an improvement of 4 or more points in peak pruritus weekly score at Week 16, relative to baseline;
4 Proportion of participants with EASI-90 at Week 16.

If the treatment comparison for (1) is statistically significant, testing will proceed to (2), and so on through the testing hierarchy. If, at any point in this hierarchy, a null hypothesis cannot be rejected at the 5% significance level in favor of benralizumab, further testing will stop, and no subsequent null hypotheses will be rejected.

4.2 Analysis methods

4.2.1 Participant disposition

Participant disposition will be summarized using the all participants analysis set. The total number of participants will be summarized for the following groups: those who enrolled, those who entered run-in, and those who were not randomized (and reason). The number and percentage of participants within each treatment group will be presented by the following categories: (a) for 16-week placebo-controlled treatment period: randomized, received treatment with IP, did not receive treatment with IP (and reason), completed treatment period with IP, discontinued treatment with IP (and reason), discontinued treatment with IP but completed study follow-up, completed treatment, and withdrawn from study (and reason), (b) for extension period: entering, not entering (and reason) , completed treatment period with IP, discontinued treatment with IP (and reason), discontinued treatment with IP but completed study follow-up, completed treatment, and withdrawn from study (and reason).

Screen failure information will be listed for the all participants analysis set.

For participants in the FAS, the number of participants remaining on treatment, participants discontinued IP but still in study follow-up, and participants who withdraw from the study will be summarized by treatment group and scheduled visit.

The number of participants randomized by country and center will also be summarized by treatment group in the FAS.

Disposition summaries will be presented for both, the 16-week placebo-controlled treatment period and the Week 52 end of study reporting.

4.2.2 Demography data and participant characteristics

Demography and baseline characteristics will be summarized by treatment group and overall in the FAS, using frequency and percentages (for categorical variables) and descriptive statistics of n, mean, standard deviation, median, minimum, and maximum, and where appropriate 1st quartile (Q1) and 3rd quartile (Q3) (for continuous variables). If there are major differences in the number of participants between the FAS and safety analysis set, these summaries will be repeated for the safety analysis set.

Age will be derived from the date of informed consent and date of birth, rounded down to the nearest integer. For participants in countries where date of birth is not recorded, the age as recorded in the eCRF will be used.

Various baseline characteristics will also be summarized, including participant characteristics (screening age (as continuous and categorical variable (≥ 12 to < 18 years/≥ 18 years)), weight, height, BMI, ethnicity, race, screening blood eosinophils (< 300 cells/μL/≥ 300 cells/μL)). Medical and surgical histories will be summarized separately for past and current conditions by MedDRA Preferred Term (PT) within MedDRA System Organ Class (SOC). History of AD and age (years) at first appearance and diagnosis of AD and time since first appearance and diagnosis (years) will be summarized separately.

The handling of partial/missing dates for age (years) at/time since first appearance/diagnosis of AD is as follows, to allow for durations to be presented: If the day of date of appearance/diagnosis is missing it will be set to the first day of the month that the appearance/diagnosis occurred. If the date of appearance/diagnosis is missing both the day and month, and the year of appearance/diagnosis is in the same year as the year of birth, then set the date to be equal to date of birth (if both the day and month of birth are missing then set to January 1 of the year of birth; if the day of birth is missing then set to the first day of the month of birth); otherwise the date will be set to January 1 of the year of diagnosis.

4.2.3 Prior and concomitant medications

The frequency and percentage of participants reporting usage of prior medications, those reporting use of allowed concomitant medications, and those reporting usage of disallowed concomitant medications (refer to Section 6.5 of the CSP) will be summarized by treatment group and overall. Background medications will be summarized by treatment group and overall.

The summary tables will present data by generic term within ATC code.

The frequency and percentage of participants reporting usage of rescue medication will be summarized by treatment group and overall for scheduled visits.

4.2.4 Study treatments

Exposure to IP administration will be calculated in days as:

(last dose date of IP - first dose date of IP + 1)

where for 16-week placebo-controlled treatment period last dose date of IP will be the last dose date prior to entering the extension period, for the Week 52 end of study reporting the first and last dose date for the entire study will be used, and for the extension period alone the first dose date will be the first dose in the extension period.

Exposure will be summarized descriptively by treatment group for the safety analysis set.

The number and percentage of participants with duration of IP exposure in each of the following categories will also be summarized:

- <16 weeks
- ≥16 weeks and <24 weeks
- ≥24 weeks and <36 weeks
- ≥36 weeks and <52 weeks

IP exposure will be presented for the 16-week placebo-controlled treatment period, the 36-week extension period, and the overall Week 52 end of study reporting.

4.2.5 Compliance

Study treatment compliance with IP administration will be summarized descriptively by treatment group for the FAS and calculated as:

(Total doses administered / total doses expected) x 100.

The total number of doses expected includes all visits with protocol scheduled IP administration on or before a participant's IP discontinuation or treatment completion date.

Compliance will be presented for the 16-week placebo-controlled treatment period, the 36-week extension period, and the overall Week 52 end of study reporting.

4.2.6 Primary outcome variable

4.2.6.1 Primary analysis

The primary endpoint is the proportion of participants with IGA 0/1 (ie, clear or almost clear) and a decrease in IGA of 2 or more points at Week 16, relative to baseline. The null hypothesis is: the proportion of participants on benralizumab 30 mg Q4W is equal to the proportion of participants on placebo. Whereas, the alternative hypothesis is: the proportion of participants on benralizumab 30 mg Q4W is not equal to the proportion of participants on placebo. That is,

H_0: Difference in proportions (benralizumab 30 mg Q4W – placebo) = 0

H_1: Difference in proportions (benralizumab 30 mg Q4W – placebo) ≠ 0

Hypothesis testing for the primary analyses will be performed according to the multiple testing procedure described in Section 4.1.1.

For the primary and key secondary endpoints, the primary estimand will be based on the FAS (Section 2.2.2). Intercurrent events will consist of participants who withdraw from the study or require rescue therapy (defined in CSP Section 6.5.3). The primary estimand will regard these participants as non-responders from the time such events occur up to Week 16. A participant with missing data at a specific time point will also be considered as a non-responder at that time point.

The primary endpoint is a binary response that classifies a participant's skin clearance at Week 16 using the IGA score (see Section 3.2). Participants will be classified as responders if IGA 0/1 (ie, clear or almost clear) and they have a decrease in IGA of 2 or more points relative to baseline. Otherwise, participants will be classified as non-responders.

For the primary analysis, a logistic regression model will be fitted to the primary endpoint using a logit link function. The model will include treatment group and baseline covariates for age at screening (≥ 12 to < 18 years/≥ 18 years), blood eosinophils at screening (< 300 cells/μL; ≥ 300 cells/μL), and baseline IGA score. The strata variables will be used as recorded in the database, however for missing eosinophil data from the central laboratory the IVRS results will be used.

The model will be used to estimate the proportion of responders for benralizumab 30 mg Q4W and placebo, the difference between these proportions (benralizumab 30 mg Q4W – placebo) and an odds ratio, with corresponding 95% CIs. A p-value, corresponding to a 2-sided test, will be presented to compare the benralizumab 30 mg Q4W and placebo treatment groups. Marginal standardization methods (Bartlett 2018) will be used to calculate the proportions of responders, the difference in proportions and the confidence intervals from the model. Note

that if the logistic regression model does not converge to a solution due to low response rates in certain strata, these data will be analyzed using a Cochran–Mantel–Haenszel test.

The observed proportion of responders will be summarized using descriptive statistics by treatment group and each scheduled visit.

Sensitivity analyses to alternative missing data assumptions and alternative estimand approaches may be performed as described in Section 4.2.6.3 below.

4.2.6.2 Subgroup analyses for the primary outcome variable

To explore the uniformity of the detected overall treatment effect on the primary efficacy variable, subgroup analyses and statistical modelling including testing for interaction between treatment and covariates will be performed for the following factors:

- Region (Europe, North America, Rest of World)
- Gender (male, female)
- Age group categories (< 18 years, ≥ 18 – 35 years, ≥ 35 years) at screening from eCRF
- Age group categories (< 21 years, ≥ 21 – 35 years, ≥ 35 years) at screening from eCRF
- Blood eosinophils (< 300 cells/µL, ≥ 300 cells/µL) at screening from central laboratory results
- BMI (≤ 30 kg/m^2, > 30 kg/m^2)
- Race (White, Asian, Other)
- Duration of AD at enrolment (< median, ≥ median; calculated on FAS)
- Age at onset of AD (< 18 years, ≥ 18 years; if sufficient data: < 12 years, ≥12 – 18 years, ≥ 18 years)
- Baseline IGA score (3, ≥ 4)
- Baseline EASI (< median, ≥ median)
- Previous biologic use (yes, no; if sufficient data: responded to, failed, naïve) at screening from eCRF
- Previous immunosuppressant (yes, no; if sufficient data: responded to, failed, naïve) at screening from eCRF
- Co-morbid asthma (yes, no) at screening from eCRF
- IgE (< median, ≥ median) from central laboratory results; calculated on FAS
- Staph aureus (lesional) (yes, no)
- Staph aureus(non-lesional) (yes, no)
- Aeroallergens (yes, no) at screening from eCRF
- Food allergies (yes, no) at screening from eCRF

For each of the subgroup factors in turn, a separate logistic regression model will be fitted using the same model terms as used for the primary analysis (described in Section 4.2.6.1),

with additional terms for the subgroup main effect (unless the subgroup main effect is already included) and the treatment by subgroup interaction. If any model does not converge, respective sub-groups may be collapsed appropriately.

Similar outputs will be presented for each subgroup as for the primary analysis. The p-value for the interaction term by each treatment group will be presented in the analysis summary tables and forest plots will present the odds ratios with confidence intervals for each level of the subgroups.

It is important to note that the study has not been designed or powered to assess efficacy within any of these pre-defined subgroups (except for strata of age and eosinophils), and as such these analyses are considered as exploratory.

4.2.6.3 Sensitivity analyses for the primary outcome variable

Sensitivity analyses may explore the effect of including other baseline covariates if imbalances in important factors are observed.

For the primary analyses of binary endpoints, participants who require any rescue therapy during the study will be classified as non-responders from the time of rescue use up until Week 16. Participants with missing data for binary endpoints will also be considered non-responders. Sensitivity analyses that may be performed to explore the impact of rescue therapy use and missing data on the results may include:

- Participants with rescue therapy use (from Day 29 onwards) remain as non-responders, other missing data at Week 16 will be imputed using LOCF approach;
- Treatment policy approach to handle rescue therapy use where all observed data is included in the analysis regardless of rescue therapy, missing data will be considered non responders;
- All observed data will be included in the analysis regardless of rescue therapy, but with no imputation of missing data to be non-responders.

Additional sensitivity analyses for the primary and key secondary endpoints based on different missing data mechanism assumptions, may be used to explore the robustness of any treatment effect including multiple imputation approaches, if the amount of missing data warrants further investigation. Details of multiple imputation techniques for continuous endpoints are specified in Appendix 8.1, and can be adapted for binary endpoints if required.

4.2.7 Secondary outcome variables

4.2.7.1 Proportion of skin clearance EASI

The first and third key secondary endpoints of proportion of participants with skin clearance EASI-75 and EASI-90 at Week 16 for benralizumab 30 mg Q4W and placebo treatment groups will be analyzed using a logistic regression model in a similar way to that described for

the primary endpoint analyses (Section 4.2.6.1) including baseline EASI as a covariate instead of baseline IGA.

The observed proportion of responders will be summarized using descriptive statistics by treatment group and each scheduled visit.

For each of the first and third key secondary endpoints, the null hypothesis is: the proportion of participants on benralizumab 30 mg Q4W is equal to the proportion of participants on placebo. Whereas, the alternative hypothesis is: the proportion of participants on benralizumab 30mg Q4W is not equal to the proportion of participants on placebo. That is,

H_0: Difference in proportions (benralizumab 30 mg Q4W – placebo) = 0

H_1: Difference in proportions (benralizumab 30 mg Q4W – placebo) ≠ 0

Sensitivity analyses for the key secondary endpoints may be considered if appropriate to explore the impact of rescue therapy use or missing data. Details are described above for the primary endpoint (Section 4.2.6.3).

The secondary endpoints of proportion of participants with skin clearance EASI-50 and EASI-100 at Week 16 will be analyzed using the same logistic regression model as the key secondary endpoint.

The proportion of participants with skin clearance EASI-75 at Week 52 will be analyzed with logistic regression model as described above, comparing effects on the Q4W and Q8W maintenance regimens.

4.2.7.2 Proportion with improvement in peak pruritus numeric rating scale

The second key secondary endpoint of proportion of participants with an improvement of ≥ 4 points in peak pruritus weekly score at Week 16 for benralizumab 30 mg Q4W and placebo treatment groups will be analyzed using a logistic regression model in a similar way to that described for the primary endpoint analyses (Section 4.2.6.1).

The observed proportion of responders will be summarized using descriptive statistics by treatment group and each scheduled visit.

For the second key secondary endpoint, the null hypothesis is: the proportion of participants on benralizumab 30 mg Q4W is equal to the proportion of participants on placebo. Whereas, the alternative hypothesis is: the proportion of participants on benralizumab 30mg Q4W is not equal to the proportion of participants on placebo. That is,

H_0: Difference in proportions (benralizumab 30 mg Q4W – placebo) = 0

H_1: Difference in proportions (benralizumab 30 mg Q4W – placebo) ≠ 0

Sensitivity analyses for the key secondary endpoint may be considered if appropriate to explore the impact of rescue therapy use or missing data, respectively. Details are described above for the primary endpoint (Section 4.2.6.3).

The proportion of participants with improvement of ≥ 4 in peak pruritus weekly score at Week 52 will be analyzed as described above, comparing effects on the Q4W and Q8W maintenance regimens.

4.2.7.3 Change from baseline in EASI score

The secondary endpoint of change from baseline in EASI score will be analyzed by a MMRM model at Week 16.

The model will include change from baseline in EASI score as the dependent variable (derived for each participant and every scheduled visit up to Week 16); baseline EASI score as a continuous covariate; treatment, age at screening (≥ 12 to < 18 years/≥ 18 years), blood eosinophils at screening (< 300 cells/µL; ≥ 300 cells/µL), and visit as categorical factors; and treatment-by-visit as an interaction term. The strata variables will be used as recorded on the database. An unstructured variance-covariance matrix will be used to model within-subject errors, and the Kenward-Roger approximation will be used to estimate denominator degrees of freedom and to adjust standard errors, the OBSMARGINS option will be used to derive the LSMEAN estimates assuming mean values for the baseline covariates. If the model fails to converge, alternative variance-covariance models will be tried in the following order: Toeplitz, first-order regressive, compound symmetry.

For each visit, the model will be used to estimate the mean change from baseline for each treatment group (benralizumab 30mg Q4W and placebo) and difference between treatment groups, with corresponding 95% CIs and the p-value, corresponding to a 2-sided test.

The distributional assumptions of the MMRM model will be checked by examining plots of residuals. If the residuals suggest that it is unreasonable to assume that data follow a Normal distribution (eg, data are skewed and the constant variance assumption is not met), alternative models will be explored. For example, data may be log transformed prior to fitting the model, or a non-parametric method may be deemed more appropriate.

The observed means and mean changes from baseline will be summarized using descriptive statistics by treatment group and each scheduled visit up to Week 16.

Additional sensitivity analyses based on different missing data mechanism assumptions may be used to explore the robustness of any treatment effect, including

- Participants with rescue therapy use (from Day 29 onwards) will be imputed using LOCF approach and analyzed with analysis of covariance (ANCOVA);
- Treatment policy approach where all available data is included regardless of rescue use, and analyzed with MMRM;
- Multiple imputation techniques if the amount of missing data due to intercurrent events warrants further investigation.

Details of multiple imputation techniques for continuous endpoints are specified in Appendix 8.1.

Changes from baseline in EASI total score at Week 52 will be analyzed with MMRM model as described above, comparing effects on the Q4W and Q8W maintenance regimens.

4.2.7.4 Change from baseline in peak pruritus numeric rating scale

The change from baseline in peak pruritus score will be analyzed using a MMRM model as described above for the change in EASI (Section 4.2.7.3). The change at Week 2 along with other supportive timepoints will be estimated from this model.

Descriptive summary statistics for scores and changes from baseline will be produced by treatment group and visit.

4.2.7.5 Change from baseline in POEM score

The change from baseline in POEM score at Week 16 and at Week 52 will be analyzed using MMRM model as described for the change in EASI (Section 4.2.7.3).

Descriptive summary statistics for scores and changes from baseline will be produced by treatment group and visit.

4.2.7.6 Change from baseline in SCORAD Index

The change from baseline in SCORAD Index at Week 16 and at Week 52 will be analyzed using MMRM model as described for the change in EASI (Section 4.2.7.3).

Descriptive summary statistics for scores and changes from baseline will be produced by treatment group and visit.

4.2.7.7 Change from baseline in DLQI

The change from baseline in DLQI at Week 16 and at Week 52 will be analyzed using MMRM model as described for the change in EASI (Section 4.2.7.3).

Descriptive summary statistics for scores and changes from baseline will be produced by treatment group and visit.

4.2.7.8 Change from baseline in CDLQI

The change from baseline in CDLQI at Week 16 and at Week 52 will be analyzed using MMRM model as described for the change in EASI (Section 4.2.7.3).

Descriptive summary statistics for scores and changes from baseline will be produced by treatment group and visit.

4.2.7.9 Serum benralizumab concentration

Serum benralizumab concentration will be summarized using descriptive statistics at each visit by treatment group based on the PK analysis set.

Serum benralizumab concentrations will be presented for both, the 16-week placebo-controlled treatment period and the Week 52 end of study reporting.

4.2.7.10 Immunogenicity

Anti-drug antibodies to benralizumab will be summarized using descriptive statistics at each visit by treatment group based on the safety analysis set. The ADA titres-time profiles of benralizumab by treatment group will be generated. The impact of ADA on PK and eosinophil level will be assessed. The potential association of ADA with safety and efficacy will be evaluated.

Further details are provided in Appendix 8.2.

ADA will be presented for both, the 16-week placebo-controlled treatment period and the Week 52 end of study reporting.

4.2.7.11 Proportion with an IGA response at Week 52

The proportion of participants with an IGA 0/1 and a decrease in IGA of ≥ 2 points at Week 52 relative to baseline, will be analyzed using a logistic regression model in a similar way to that described for the primary endpoint analyses (Section 4.2.6.1) comparing effects on the Q4W and Q8W maintenance regimens.

4.2.8 Safety Analyses

Safety analyses will be performed using the safety analysis set. Summaries will be presented by the 3 randomized treatment sequences (benralizumab 30 mg Q4W, benralizumab 30 mg Q8W and placebo/benralizumab) and the initial benralizumab 30 mg Q4W Total dose group, as noted in Sections 1.1.1 and 4.1.

A small number of key safety tables will also be presented for data in the extension period alone to explore the safety profile were a direct comparison of Q4W versus Q8W dosing can be made for participants previously receiving benralizumab, and to summarize time on benralizumab alone for participants previously receiving placebo.

Adverse event tables will be reported separately for the 16-week placebo-controlled treatment period and for the 36-week extension period. Some key adverse event tables will be repeated overall for Week 52 whole study.

Participants will be analyzed according to the treatment they received. In general, the baseline value for statistical analysis is the last non-missing value prior to administration of the first dose of IP. Safety data will be presented using descriptive statistics. No safety data will be imputed.

4.2.8.1 Adverse Events

TEAEs will be summarized separately for the on-treatment, on-study, and post-treatment periods, as defined in Section 3.4.1. The primary period of interest for safety summaries will be the on-treatment period for the 16-week placebo-controlled treatment period, and this is the only period that will be used for the placebo controlled first 16 weeks summaries. Summaries will also be produced for the on-treatment period of the extension period. Limited summaries will be produced for the on-study and post-treatment periods for the Week 52 end of study reporting summaries only. AEs in the pre-treatment period (with start date prior to the first dose of IP) will be listed only.

All summaries will be presented by treatment group and key tables will include exposure-adjusted rates to account for any differences in follow up.

The rate of TEAEs per person-years at risk will be calculated as (number of participants reporting the TEAE) / (total time with participants at risk of TEAE) for on-treatment and on-study periods. The post-treatment AEs will be listed. The total period at risk for each participant will be the duration of the on-treatment, post-treatment and on-study periods as defined in Section 3.4.1. Rates will be expressed in terms of events per 100 participant-years.

An overall summary table will be produced showing the number, percentage, and exposure-adjusted rate of participants with at least one TEAE in any of the following categories; TEAEs, SAEs, AEs with outcome of death, and AEs leading to discontinuation of IP (DAEs). AEs will be presented for each treatment group by system organ class (SOC) and preferred term (PT), including the number and percentage of participants reporting at least one event, number of events and exposure-adjusted rates, where appropriate.

TEAEs, AEs with outcome of death, SAEs and DAEs will be summarized by SOC and PT assigned to the event by MedDRA. For each PT, the number, percentage and exposure-adjusted rate of participants reporting at least one occurrence will be presented (ie, multiple occurrences of an AE for a participant will only be counted once).

A summary of the most common (frequency of >5%) TEAEs will be presented by PT. AEs and SAEs causing discontinuation of the IP and SAEs causing discontinuation from the study will be summarized by SOC and PT.

TEAEs, SAEs and DAEs will be summarized by PT and Investigator's causality assessment (related versus not related) and maximum intensity, including reporting of seriousness, death and events leading to discontinuation of IP, as well as other action taken related to IP. If a participant reports multiple occurrences of the same TEAE within the same study period, the maximum intensity will be taken as the highest recorded maximum intensity (the order being mild, moderate, and severe).

Other significant TEAEs will include but may not be limited to injection site reactions and hypersensitivity events. TEAEs of injection site reactions (high level term of injection site reactions) and hypersensitivity (standardized MedDRA query of hypersensitivity [narrow]) will be summarized by preferred term. Injection site reactions will be summarized by injection site location and number of IP administrations. The summary of TEAEs of hypersensitivity will be presented by PT and will be repeated for events causally related to IP as assessed by the Investigator. Hypersensitivity events will be listed.

Key participant information will be presented for participants with AEs with outcome of death, SAEs, and DAEs. An AE listing for the safety analysis set will cover details for each individual AE.

AEs ongoing at primary DBL will be presented for both primary and final DBL reporting.

AEs will be presented for both, the 16-week placebo-controlled treatment period and the Week 52 end of study reporting.

4.2.8.2 Clinical Laboratory Safety Assessments

All protocol-specified continuous laboratory parameters will be summarized descriptively by absolute value at each visit by treatment group, together with the corresponding changes from baseline. Summary statistics for continuous variables will present n, mean, SD, Minimum, Q1, median, Q3, and Maximum.

All parameters will be summarized in SI units, except for blood eosinophil counts which will be summarized in both SI and conventional units. Results reported by the central laboratory in conventional units will be converted to SI units for reporting.

Central laboratory reference ranges will be used for the identification of abnormalities, and a shift table will be produced for each laboratory parameter to display low, normal, and high values. The shift tables will present baseline and maximum/minimum on-treatment value, as

applicable for each parameter and will include participants with both baseline and post-baseline data.

Shift plots showing each individual participant's laboratory value at baseline and at maximum/minimum post-baseline will be produced for each continuous laboratory variable. If any laboratory variables show any unusual features (high or low values or a general shift in the data points) at other time points, then shift plots of these data may be produced.

Data for participants who have treatment-emergent changes outside central laboratory reference ranges will be presented. This data presentation will include all visits for those participants.

Maximum post-baseline TBL elevations by maximum post-baseline ALT and AST will be presented, expressed as multiples of the ULN. TBL will be presented in multiples of the following ULN: ≤1.5, >1.5-2, >2. AST and ALT will be presented in multiples of the following ULN: ≤1, >1-3, >3-5, >5-10, >10.

Maximum post-baseline TBL will be presented (<2 and ≥2 x ULN) and plotted against maximum post-baseline ALT (<3, ≥3-<5, ≥5-<10, and ≥10 x ULN), expressed as multiples of ULN. This will be repeated to show maximum post-baseline TBL against maximum post-baseline AST.

For urinalysis data, a shift table will be generated to present changes from baseline to maximum/minimum post-baseline value, displaying normal and abnormal values for selected parameters and will include participants with both baseline and post-baseline data.

Any data outside the central laboratory reference ranges will be explicitly noted on the listings.

Laboratory parameters will be presented for both the 16-week placebo-controlled treatment period and the Week 52 end of study reporting.

4.2.8.3 Vital signs, physical examination and electrocardiogram

For vital signs data, descriptive statistics for absolute value, change from baseline and percent change from baseline will be presented for each treatment group by scheduled visit. Summary statistics for continuous variables will present n, mean, SD, Minimum, Q1, median, Q3, and Maximum.

For physical examination data, only information on whether the assessment was performed or not is to be recorded.

For ECG at screening, descriptive summaries of normal or abnormal evaluation and clinical significance will be provided by treatment group.

All vital sign and ECG data will be listed.

Vital signs and electrocardiogram data will be presented for both, the 16-week placebo-controlled treatment period and the Week 52 end of study reporting.

4.2.9 Tertiary/exploratory Endpoints

Analyses for exploratory objectives will be specified below, or additionally in an exploratory analysis plan.

4.2.9.1 Healthcare resource utilisation

Descriptive summary statistics will be produced by treatment group and visit.

4.2.9.2 Hospital Anxiety and Depression Scale (HADS)

The change from baseline in HADS scores for anxiety and depression, respectively, will be analyzed at Week 16 and at Week 52 using MMRM models as described for the change in EASI (Section 4.2.7.3).

Descriptive summary statistics for HADS anxiety and depression scores and change from baseline will be produced by treatment group and visit. Summaries of score categories considered normal, possible cases of anxiety/depression, and likely cases of anxiety/depression will be produced by treatment group and visit.

4.2.9.3 Patient Global Impression of Severity (PGI-S)

The PGI-S scores will be summarized descriptively by treatment group and visit for both, the 16-week placebo-controlled treatment period and the Week 52 end of study reporting.

4.2.9.4 SF-36 (Version 2), Acute Recall

The number and percentage of participants above and below the threshold values for the SF-36v2 Scale domains, PCS and MCS will be presented. The binary endpoints will be analyzed at Week 16 and at Week 52 using logistic regression models in a similar way to that described for the primary endpoint analyses (Section 4.2.6.1).

Descriptive summary statistics for scores and change from baseline will be produced by treatment group and visit.

CCI

CCI [redacted]

4.2.9.6 Patient-reported Experience

The Patient-reported Experience will be analyzed separately and not form part of this SAP.

4.2.9.7 Body Surface Area Involvement of Atopic Dermatitis

The change from baseline body surface area involvement at Week 16 and at Week 52 will be analyzed using MMRM models as described for the change in EASI (Section 4.2.7.3). Both, the results from regular assessment and from electronic device calculations will be analyzed separately.

Descriptive summary statistics for scores and change from baseline will be produced by treatment group and visit.

4.2.9.8 European Quality of Life-5 Dimensions (EQ-5D-5L)

The EQ-5D-5L will be analyzed separately and not form part of this SAP.

CCI [redacted]

4.2.9.11 Biomarkers

To explore the pharmacodynamic effect of benralizumab, summaries and plots of absolute levels and changes from baseline in pharmacodynamic biomarkers including CCI [redacted] CCI [redacted] will be produced as appropriate.

To evaluate the association of baseline biomarkers with clinical efficacy, and assess for any markers potentially predictive of benralizumab treatment effect the following analysis techniques may be considered:

- Forest plots for binary efficacy measures (primarily IGA response, EASI-75 response may also be included if necessary) at Week 16 of odds ratio by biomarker quartiles or clinically relevant thresholds pre-defined before unblinding;

- Scatterplots of baseline biomarker levels versus continuous efficacy measures (change from baseline for EASI, POEM and SCORAD) at Week 16, with locally weighted smoothing (LOESS) lines by treatment as appropriate;
- If appropriate, forest plots presenting subgroup analyses of continuous efficacy measures at Week 16 (eg change from baseline for EASI) by subgroups defined by baseline biomarker levels. Groupings of baseline biomarker levels may include splitting biomarkers at the median, at tertiles, or at quartiles, or other clinically relevant thresholds of interest to be pre-defined prior to unblinding. Results would be presented in forest plots of treatment effect (difference between benralizumab and placebo in change from baseline in EASI at Week 16) in each level of the biomarker subgroups.

Additional Biomarker analyses may be explored separately outside of this SAP.

4.2.9.12 Clinical skin photography

Results from clinical skin photography assessments will be compared to in-person assessments of clinical efficacy taken on the same visit. Descriptive summary statistics for clinical skin photography assessments of scores and change from baseline will be produced by treatment group and visit.

Agreement between skin photography and in-person clinical assessment will be explored over all treatment groups (benralizumab, placebo) combined, as follows:

For binary outcomes (IGA response) cross-tabulations of responder/non-responder by skin photography/in-person assessment will be produced, presenting number and percent in each category. Cohen's kappa statistic (κ) will be calculated for evaluating agreement:

κ = (observed agreement [Po] – expected agreement [Pe]) / (1-expected agreement [Pe]), where Pe=0.5.

The results will be interpreted as 0 = agreement equivalent to chance; 0.10–0.20 = slight agreement; 0.21–0.40 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.80 = substantial agreement; 0.81–0.99 = near-perfect agreement; and 1.00 = perfect agreement.

For the continuous outcomes (EASI, SCORAD, BSA) Bland-Altman plots and limits of agreement (Bland and Altman 1986) will be used. Scatter plots will be produced of the difference between the two measurements (y-axis) against the average of the two

measurements (x-axis), and colour coded by time point. This provides a graphical display of bias with 95% limits of agreement (difference ± 1.96x SD of observed differences).

5 INTERIM ANALYSES

No formal interim analysis is planned for the study.

The first set of analyses will be run after the primary DBL, which will occur after all randomized participants have completed the initial 16-week treatment period or discontinued prior to Week 16. The final set of analyses will be produced at the final DBL, which will occur when all participants have completed the 52-week treatment period and/or the IPD/EOT/Week 60 follow-up visit.

An additional analysis may be performed between the primary and final DBLs to report data accumulating during the extension part of the study if needed to support end of Phase 2 decision making.

Participants and investigators will remain blinded to the dosing regimens until the final DBL.

6 CHANGES OF ANALYSIS FROM PROTOCOL

Section 2.2.5 Extension period analysis set is added to identify all participants who enter the extension period.

Sections 3.5.11 and 4.2.9.11 Biomarker analyses added to present in SAP and CSR.

7 REFERENCES

Bland and Altman 1986
Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement Lancet 1986, i, 307-310.

Bartlett 2018
Bartlett J. Covariate adjustment and estimation of mean response in randomised trials. Pharmaceutical Statistics 2018;17:648-66.

Finlay and Khan 1994
Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. Clin Exp Dermatol 1994 May;19(3):210-6.

Hanifin et al 2001
Hanifin JM, Thurston M, Omoto M, Cherill R, Tofte SJ, Graeber M. The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. Exp Dermatol 2001 Feb;10(1):11-8.

Hopkins et al 2009
Hopkins C, Gillett S, Slack R, Lund VJ, Browne JP. Psychometric validity of the 22-item Sinonasal Outcome Test. Clin Otolaryngol 2009 Oct;34(5):447-54. doi: 10.1111/j.1749-4486.2009.01995.x.

Juniper et al 1999
Juniper EF, O'Byrne PM, Guyatt GH, Ferrie PJ, King DR. Development and validation of a questionnaire to measure asthma control. Eur Respir J 1999 Oct;14(4):902-7.

Juniper et al 2005
Juniper EF, Svensson K, Mörk AC, and Ståhl E. Measurement properties and interpretation of three shortened versions of the asthma control questionnaire. Respir Med. 2005 May;99(5):553-8. Epub 2004 Nov 26.

Juniper et al 2006
Juniper EF, Bousquet J, Abetz L, Bateman ED. Identifying 'well-controlled' and 'not well-controlled' asthma using the Asthma Control Questionnaire. Respir Med 2006 Apr;100(4):616-21. Epub 2005 Oct 13.

Kunz et al 1997
Kunz B, Oranje AP, Labrèze L, Stalder JF, Ring J, Taïeb A. Clinical validation and guidelines for the SCORAD index: consensus report of the European Task Force on Atopic Dermatitis. Dermatology 1997;195(1):10-9.

Lewis-Jones and Finlay 1995
Lewis-Jones MS, Finlay AY. The Children's Dermatology Life Quality Index (CDLQI): initial validation and practical use. Br J Dermatol. 1995 Jun;132(6):942-9.

Maruish 2011
Maruish, M. E. (Ed.). User's manual for the SF-36v2 Health Survey (3rd ed.). Lincoln, RI: QualityMetric Incorporated. 2011.

Piccirillo et al 2002
Piccirillo JF, Merritt MG Jr, Richards ML. Psychometric and clinimetric validity of the 20-Item Sino-Nasal Outcome Test (SNOT-20). Otolaryngol Head Neck Surg 2002 Jan;126(1):41-7.

8 APPENDIX

8.1 Accounting for missing data

Missing data can occur due to intercurrent events like non-adherence to randomized treatment, or early discontinuation, or non-adherence of ePRO usage. The amount of missing data is

minimized in this study as participants are encouraged to remain in the study after premature discontinuation of IP and complete visits according to the protocol.

This section summarizes how we will describe the pattern of and reasons for missing data from the study. It will also describe how we plan to account for missing data, including both the primary and sensitivity analyses to assess the robustness of the treatment effect under different underlying assumptions to account for missing data.

8.1.1 Missing data description

Tabular summaries for the percentage of participants by the reason for discontinuation of randomized treatment as well as for withdrawal from the study will be presented by treatment to describe why participants discontinue from randomized treatment or withdraw from the study. The time to discontinuation of randomized treatment and withdrawal from the study will be presented using Kaplan Meier plots (overall and split by treatment related/not treatment related reason for discontinuation, as defined in Table 12). Dependent on these outputs additional exploratory analyses may be produced as deemed necessary to further understand the pattern of missing data.

Descriptive summaries of number and percentage of participants using rescue medication and change of background medications will be provided by treatment group.

For primary endpoint and key secondary endpoints, the summaries of number and percentage of participants with missing/non-missing daily assessments will be presented by treatment group and visit.

8.1.2 Multiple imputation (MI) approach

Analyses will include all participants in the FAS.

The following 4 steps will be used to build the imputation datasets and perform analyses:

1 100 datasets with a monotone missing pattern will be obtained, induced by Markov Chain Monte Carlo (MCMC) method up to Week 16, on participants with missing data not due to reason of interest (eg lost-to-follow-up).
2 For each of the imputed datasets obtained in step 1, the remaining missing data due to reason of interest (eg AE) up to Week 16, will be imputed using the regression method for the monotone pattern with adjustment for covariates as for the primary model.
3 Each of the 100 imputed datasets will be merged with the one dataset imputed by worst case carried forward (WOCF) approach, and then be analyzed using the main statistical model. These 100 datasets will be saved.
4 Apply Rubin's rule (Rubin DB 1987) to combine analysis results (point estimates and standard errors) from 100 imputations. Descriptive statistics including number of

participants, mean, standard error, and least squares (LS) means will be provided for each timepoint. In addition, difference in LS means and the corresponding 95% CIs will be provided along with the p-values for Week 16.

8.1.3 Primary analysis for continuous endpoints under effective estimand using the Missing at Random (MAR) assumption

The primary analysis for the continuous endpoints allows for differences in outcomes over the study treatment period up to 16 weeks. This analysis includes all data until participants withdraw from the study or up to Week 16, whichever is earlier, regardless of if they discontinue from randomized treatment. Data after rescue therapy taken after Day 29 will be considered as missing. The primary analysis uses a MMRM method, adjusting for covariates (Section 4.2.7.3), and assumes that missing data is missing at random (MAR) and is a direct likelihood approach (DL).

8.1.4 Sensitivity analyses: Primary estimand assuming dropout reason-based multiple imputation approach

To examine the sensitivity of the results of the primary analysis to departures from the underlying assumptions, additional analyses will be performed. As with the primary analysis the sensitivity analyses include all data until participants withdraw (up to Week 16) from the study regardless of if they discontinue from randomized treatment.

For sensitivity analyses, missing data will be imputed using controlled sequential MI methods based on pattern mixture models (EMA/CHMP/EWP 2010). This model will assume that some pre-specified subset of participants who withdraw from the study have correlations with future unobserved visits similar to participants in the placebo arm.

The default assumptions that will be used to impute the missing data are as follows:

(a) MAR: Assumes that the trajectory for participants who dropped out in each arm is similar to those observed in their own treatment arm.

(b) Dropout Reason-based Multiple Imputation (DRMI): Assumes that the trajectory for participants in the benralizumab arms who dropped out for a treatment related reason or severe non-compliance of protocol is similar to that of the placebo participants, whereas the remaining participants who dropped out are imputed assuming MAR.

A summary of reasons for participants withdrawing from the benralizumab treatment arm and the corresponding treatment arm used to calculate the imputation response rate under MAR and DRMI is given in Table 12.

Table 12 Parameters for calculating the imputation under MAR and DRMI

Reason for withdrawal	MAR	DRMI
Adverse Event	Benralizumab	Placebo

Reason for withdrawal	MAR	DRMI
Development of study-specific discontinuation criteria*	Benralizumab	Placebo
Death	Benralizumab	Placebo
Severe non-compliance to protocol	Benralizumab	Placebo
Eligibility criteria not fulfilled	Benralizumab	Benralizumab
Participant lost to follow up	Benralizumab	Benralizumab
Participant decision	Benralizumab	Based on review prior to study unblinding
COVID-19	Benralizumab	Benralizumab
Other	Benralizumab	Based on review prior to study unblinding

Note: Participants in the placebo arm are imputed using the mean of the non-missing values in placebo arm.
* Development of study-specific discontinuation criteria are based on the following: Anaphylactic reaction to the investigational product requiring administration of epinephrine; Development of helminth parasitic infestations requiring hospitalization; If 2 doses of the IP are missed during course of the study; A respiratory-related event requiring mechanical ventilation.

Some reasons for withdrawal are clearer to determine as treatment related (AEs, Death, Development of study-specific discontinuation criteria) or non-treatment related (participants lost to follow up, eligibility criteria not fulfilled). Other reasons are less clear such as participant decision and 'Other'; a review of each participant who withdraws from the study will therefore be carried out prior to unblinding the study. The review will include assessment of the reason for discontinuation of randomized treatment for those participants who discontinued randomized treatment and then withdrew from the study and also free text for when the reason for withdrawal or discontinuation of randomized treatment is recorded as subject decision or other. Based on this review the default assumptions for DRMI as described in b) and Table 12 may be changed. A list of these participants and the assumptions made under DRMI will be documented prior to unblinding of the study.

Together with the primary analysis, the sensitivity analyses are considered to cover the range from realistic to plausible worst-case assumptions about missing data.

The DRMI approach has been selected as a conservative approach based on the fact that placebo participants are receiving standard of care and are not expected to change to a substantially more effective treatment after withdrawing from study or study treatment. For participants receiving benralizumab who withdraw from the study due to treatment related reasons it is assumed that at worst they would be on the standard of care treatment, i.e., the placebo treatment group. For participants receiving benralizumab who withdraw from the study due to non-treatment related reasons it seems reasonable to assume they would be similar to those participants who complete treatment.

8.1.5 Overall summary of analyses to account for missing data

Forest plots will be used to show the primary analysis results along with the missing data sensitivity and alternative estimand analysis results.

It is noted that if the primary analysis is statistically significant, it is not necessarily expected that all sensitivity analyses will also give statistically significant results. If the results of the sensitivity analyses provide reasonably similar estimates of the treatment effect to the primary analysis, this will be interpreted as providing assurance that neither the lost information nor the mechanisms which cause the data to be missing have an important effect on primary analysis conclusions. Based on these outputs and the drug's mechanism of action, the plausibility of the assumptions we make about missing data in the different analyses will be considered and described in the clinical study report.

The methodology described above to explore any potential impact of missing data for the primary endpoint may be repeated for other endpoints if required due to amount or pattern of missing data.

8.1.6 References

EMA/CHMP/EWP 2010
Guideline on Missing Data in Confirmatory Clinical Trials 2 July 2010, EMA/CHMP/EWP/1776/99 Rev. 1.

Rubin DB 1987
Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons, Inc.

8.2 Analysis plan for immunogenicity data

Serum samples for ADA assessments will be conducted utilizing a tiered approach (screen, confirm, titre) and ADA data will be collected at scheduled visits shown in the CSP. ADA result from each sample will be reported as either positive or negative. If the sample is positive, the ADA titre will be reported as well.

In general, participants with a missing baseline ADA assessment will be assumed to be ADA negative at baseline as a conservative approach to ensure that all participants are included in all analyses. If a positive ADA titre result is reported as ≤50, then the titre will be imputed as 50 for titre summaries. ADA results from samples collected post-dose instead of pre-dose on an IP administration day are considered unreliable and should be excluded from all derivations.

For each participant, the following ADA responses will be evaluated over the entire on-study period through EOT:

- Participants who are ADA positive at any time during the study, including baseline and/or post-baseline (also generally referred to as ADA positive). The proportion of ADA positive participants in a population is known as ADA prevalence.
- Participants who are ADA negative at all assessments, including baseline and postbaseline (also generally referred to as ADA negative).
- Participants who are ADA positive at baseline only.
- Participants who are ADA positive at baseline and at least one post-baseline assessment.
- Treatment-emergent ADA positive (referred to as ADA incidence). A positive postbaseline result and either of the following statements holds:
  - Baseline is ADA negative and at least one post-baseline assessment is ADA positive. This is called treatment-induced ADA positive.
  - Baseline is ADA positive, and the baseline titre is boosted by greater than the variability of the assay (ie ≥ 4-fold increase) at ≥1 post-baseline timepoint. This is called treatment-boosted ADA positive.
- Participants who are persistently ADA positive, which is defined as ADA negative at baseline and having at least 2 post-baseline ADA positive measurements with ≥ 16 weeks between first and last positive, or an ADA positive result at the last available post baseline assessment.
- Participants who are transiently ADA positive, defined as ADA negative at baseline and at least one post- baseline ADA positive measurement and not fulfilling the conditions for persistently positive.
- Participants who are ADA positive with maximum titre > median of maximum titres. The median of maximum titres will be calculated based on the maximum titre of each ADA positive participant within each treatment group (including both baseline and post-baseline measurements).

The responses above will be summarized as counts and percentages by treatment group. The maximum ADA titre over the on-study period will also be summarized for participants in each of the ADA positive response categories listed above. The maximum titre will be derived based on all available ADA titres reported for each participant, including any unscheduled assessments.

ADA response (positive or negative) and titre will be summarized at baseline and at all scheduled post-baseline visits by treatment group. In the event a participant has more than one result within a given visit window, the maximum ADA titre will be used in the by-visit summary. In addition, the ADA response will be presented cumulatively. The cumulative ADA response is positive for a specific visit if a positive ADA result is detected at any time point up to and including the specific visit. If all ADA result are negative up to the specific visit, then the cumulative ADA response is negative for that visit. A summary of the number and percentage of participants who are ADA positive at a post-baseline assessment for the first time by visit will also be presented. A line plot of the proportion of participants who are ADA positive at each visit will be provided.

The proportion of participants with ADA persistently positive response will be summarized by visit and treatment group.

Key participant information will be listed for participants with positive ADA results, including ADA status, titre, benralizumab serum concentration, and eosinophil level.

All analyses will be conducted on the safety analysis set by treatment group unless otherwise specified. All ADA results will be listed.

8.2.1 ADA and efficacy

The effects of ADA on the primary and key secondary endpoints calculated through EOT will be evaluated through summary statistics by ADA status (treatment-emergent ADA positive, ADA negative), if enough data are available. However, it could have difficulties to draw solid conclusions due to no comparisons.

8.2.2 ADA and safety

Adverse events and SAEs during the study (separately for on-treatment and on-study periods) will be summarized by ADA status (treatment-emergent ADA positive, ADA negative), if enough data are available. The on-treatment and on-study periods are as defined in Section 3.4.1. AEs and SAEs will also be evaluated by causality as assessed by the Investigator. The potential impact of ADA on hypersensitivity will also be assessed if enough data are available.

8.2.3 ADA and eosinophil levels

Eosinophil levels will be summarized by visit for the following ADA response categories of participants: treatment-emergent ADA positive, ADA negative, ADA persistently positive, and ADA positive with titre > median of maximum titre, if enough data are available. A line plot of eosinophil levels by visit and ADA status will also be presented.

8.2.4 ADA and PK

Benralizumab serum concentrations will be summarized by visit and ADA status (treatment emergent ADA positive, ADA negative, ADA persistently positive, ADA positive with titre > median of maximum titre) for participants in the PK analysis set, if enough data are available.

8.3 Partial dates for adverse events and prior/concomitant medications

Dates missing the day, or both the day and month of the year will adhere to the following conventions to classify TEAEs and to classify prior/concomitant medications:

Adverse events

- The missing start day will be set to:
  - First day of the month of occurrence, if the start YYYY-MM is after the YYYY-MM of first study treatment
  - The day of the first study treatment, if the start YYYY-MM is the same as YYYY-MM of the first study treatment
  - The date of informed consent, if the onset YYYY-MM is before the YYYY-MM of the first study treatment.
- The missing end day will be set to:
  - The last day of the month of the occurrence, if the end YYYY-MM is after the YYYY-MM of the first study treatment.
  - Death date if the participant died in the same month.
  - The day of last study treatment if the YYYY-MM of occurrence is the same as the last study treatment.
- If the start date is missing both the day and month, the start date will be set to:
  - January 1 of the year of occurrence.
  - The date of the first study treatment, if the start year is the same as the year of the first study treatment
- If the end date is missing both the day and month, the date will be set to:
  - December 31 of the year of occurrence.
  - Date of death if the participant died in the same year
  - Last study treatment date if the year of occurrence is the same as the last study treatment date.
- If the start date is null, the date will be set to:
  - The date of first study treatment.
  - January 1 of the same year as the end date, if the end date suggests that the start date could be prior to the date of first study treatment.
- If the end date is null and not recorded as ongoing, the date will be set to:
  - The date of the first study treatment, if the start date is prior to the date of first study treatment.
  - The date of last visit, if the start date is on or after the date of first study treatment.

- If the end date is null and recorded as ongoing, the end date will not be imputed.

Prior/concomitant medication

- The missing day of start date of a therapy will be set to the first day of the month that the event occurred.
- The missing day of end date of a therapy will be set to the last day of the month of the occurrence.
- If the start date of a therapy is missing both the day and month, the onset date will be set to January 1 of the year of onset.
- If the end date of a therapy is missing both the day and month, the date will be set to December 31 of the year of occurrence.
- If the start date of a therapy is null and the end date is not a complete date, then the start date will be set to the earlier of the imputed partial end date and the date of the first study visit.
- If the start date of a therapy is null and the end date is a complete date
  - and the end date is after the date of the first study visit then the start date will be set to the date of the first study visit.
  - otherwise, the start date will be set to the end date of the therapy.
- If the end date of a therapy is null and the start date is not a complete date, then the end date will be set to the study end date.
- If the end date of a therapy is null and the start date is a complete date
  - and the start date is prior to the study end date, then the end date will be set to the study end date.
  - otherwise, the end date will be set to the start date of the therapy.

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d3256c00001-sap-ed-2		
Document Title:	Statistical Analysis Plan Edition 2	
Document ID:	CCI [redacted]	
Version Label:	2.0 CURRENT LATEST APPROVED	
Server Date (dd-MMM-yyyy HH:mm 'UTC'Z)	**Signed by**	**Meaning of Signature**
19-May-2022 11:52 UTC	PPD IQVIA	Author Approval
18-May-2022 17:31 UTC	PPD	Content Approval

Notes: (1) CCI [redacted]