Statistical Analysis Plan	
Study Code	D3259C00001
Edition Number	1.0
Date	16/10/2020

A Phase 2b Multinational, Randomised, Double-blind, Parallel-group, 24-week Placebo-controlled Study with 28-week Extension to Investigate the Use of Benralizumab in Patients with Chronic Spontaneous Urticaria Who are Symptomatic Despite the Use of Antihistamines (ARROYO)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ADA	Anti-drug antibodies
AE	Adverse event
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
BLQ	Below the lower limit of quantification
BMI	Body mass index
CI	Confidence intervals
CU-Q2oL	Chronic Urticaria Quality of Life Questionnaire
CSP	Clinical study protocol
CSU	Chronic spontaneous urticaria
DAE	AE leading to discontinuation
DBL	Database lock
DLQI	Dermatology Life Quality Index
DRMI	Dropout reason-based multiple imputation
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-GT
HSS7	Hives severity score over 7 days
LLOQ	Lower limit of quantification
ICF	Informed consent form
IP	Investigational product
IPD	IP discontinuation
ISS7	Itch severity score over 7 days
ITT	Intent-to treat
IVRS	Interactive voice response system
IWRS	Interactive web response system
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities

Abbreviation or special term	Explanation
MI	Multiple imputation
MMRM	Mixed-effect model repeated measurements
MNAR	Missing not at random
MTP	Multiple testing procedure
CCI	[redacted]
PGI-S	Patient Global Impression of Severity
PK	Pharmacokinetics
PRO	Patient-reported outcome
PT	Preferred term
Q1	1st quartile
Q3	3rd quartile
Q4W	Every 4 weeks
Q8W	Every 8 weeks
RNA	Ribonucleic acid
SAP	Statistical analysis plan
SC	Subcutaneous
SD	Standard deviation
SI	System international
SOC	System organ class
TBL	Total bilirubin
UAS7	Urticaria Activity Score over 7 days
UCT	Urticaria control test
ULN	Upper limit of the normal
UPDD	Urticaria patient daily diary
VAS	Visual analogue scale
WOCF	Worst observation carried forward (post-baseline)

AMENDMENT HISTORY

Category*: Change refers to	Date	Description of change	In line with the CSP?	Rationale
N/A				

* 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.

Study objectives

Table 1 Study Objectives

Primary Objective:	**Estimand description / Endpoints**
To determine the clinical efficacy of benralizumab compared to placebo in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment.	• Population: Full Analysis Set: • Endpoint: Change from baseline in ISS7 at Week 12 • Intercurrent events: All data up to Week 12 will be included regardless of randomised treatment adherence or rescue medication received • Summary measure: difference in least squares mean change from baseline in ISS7 at Week 12 between benralizumab and placebo
Secondary Objectives:	**Endpoint/variable:**
To evaluate the effect of benralizumab compared to placebo on supportive measures of clinical efficacy in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment.	• Key secondary endpoint: change from baseline in UAS7 at Week 12[a] • Proportion of responders (UAS7 $\leq$ 6) at Week 12 • Change from baseline in HSS7 at Week 12 • Time to $\geq$ 5 point decrease (clinically relevant decrease) in ISS7 • Proportion of participants with complete UAS7 response (UAS7 = 0) at Week 12 • Measures of angioedema activity at Week 12 in participants with angioedema at baseline • Change from baseline in UCT at Week 12
To evaluate the effect of benralizumab on patient-reported health-related quality of life measures in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment.	• Change from baseline in CU-Q_2oL at Weeks 12 and 24 • Change from baseline in DLQI at Weeks 12 and 24
To assess the PK and immunogenicity of benralizumab 30 mg and 60 mg in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment.	• Serum benralizumab concentration • ADA

Table 1 Study Objectives

To evaluate the longer-term effect of benralizumab compared to placebo at Week 24 on measures of clinical efficacy in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment.	• Change from baseline in ISS7 at Week 24 • Change from baseline in UAS7 at Week 24 • Proportion of responders (UAS7 $\leq$ 6) at Week 24 • Change from baseline in HSS7 at Week 24 • Proportion of participants with complete UAS7 response (UAS7 = 0) at Week 24 • Measures of angioedema activity at Week 24 in participants with angioedema at baseline • Change from baseline in UCT at Week 24
To evaluate the efficacy of administration of benralizumab Q8W versus Q4W up to Week 52 in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment.	Change from baseline in ISS7 at Week 52 Other supportive efficacy assessments at Week 52 in participants on a Q8W dosing regimen compared to those on a Q4W dosing regimen.
Safety Objective:	**Endpoint/variable:**
To assess the safety and tolerability of benralizumab in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment.	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	**Endpoint/variable:**
To evaluate the effect of benralizumab compared to placebo on healthcare resource utilisation due to CSU in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment.	Rate of CSU-related healthcare resource utilisation during the study
To evaluate the effect of benralizumab compared to placebo on overall severity of disease patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment.	Change from baseline in PGI-S score at Week 12 and at Week 24
To evaluate the mechanism of action of benralizumab in patients with CSU refractory to H_1 antihistamine treatment.	CCI [redacted]

Table 1 Study Objectives

CCI [redacted]	[redacted]
To evaluate the effect of benralizumab compared to placebo on general health status in patients with CSU refractory to H_1 antihistamine treatment.	Change from baseline in EQ-5D-5L domain and VAS scores

[a] The key secondary endpoint will use the same treatment policy strategy estimand as outlined for the primary endpoint.
ADA, anti-drug antibodies; AE, adverse event; CU-Q_2oL, Chronic Urticaria Quality of Life Questionnaire; CSU, chronic spontaneous urticaria; DLQI, Dermatology Life Quality Index; EQ-5D-5L, European Quality of Life-5 Dimensions; HSS7, hives severity score over 7 days; ISS7, itch severity score over 7 days; IP, investigational product; PGI-S, Patient Global Impression of Severity; PK, pharmacokinetics; Q4W, every 4 weeks; Q8W, every 8 weeks; RNA, ribonucleic acid; SAP, Statistical Analysis Plan; UAS7, Urticaria Activity Score over 7 days; UCT, Urticaria Control Test; VAS, visual analogue scale.

1.1 Study design

1.1.1 Overall Design

This is a Phase 2b, multinational, randomised, double-blind, parallel-group, 24-week placebo controlled study with a 28-week extension to evaluate the efficacy and safety of benralizumab in participants ≥ 18 years of age with CSU refractory to standard of care treatment, which may include second generation H_1 antihistamines (at approved or up to 4 times approved doses) as monotherapy or in combination with leukotriene receptor antagonists and/or H_2 blockers. The study is designed to evaluate 2 induction doses of benralizumab (60 mg and 30 mg) compared to placebo and to compare 2 maintenance dosing regimens (Q8W versus Q4W) in the 28-week extension period.

The study comprises the following consecutive periods:

- A 10-day to 4-week run-in period,
- A 24-week placebo-controlled, double-blind treatment period (comprising an initial 12-week ‘induction dose’ period followed by an additional 12-week dosing period),
- A 28-week blinded-to-dosing regimen extension period for maintenance treatment.

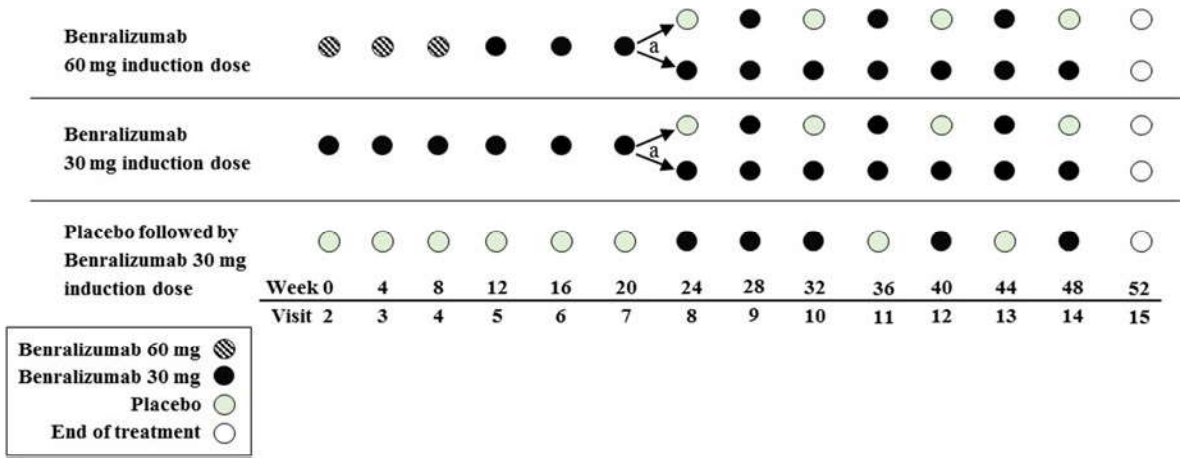
Following informed consent, all eligible participants will enter a run-in period of 10 days to 4 weeks during which inclusion/exclusion criteria will be assessed, medical history taken, and complete physical exam will be conducted. Potentially eligible participants will remain on a stable, locally-approved dose of their H_1 antihistamine treatment throughout the run-in period. Participants will be provided with a handheld device to respond to Patient-Reported Outcome (PRO) questionnaires during the study.

Following the 10-day to 4-week run-in period, approximately 160 participants will be randomised at Visit 2 stratified by region, in a ratio of 3:3:3:3:4 (n=30:30:30:30:40), respectively, to one of the following 5 treatment sequences (see also Figure 1 and Figure 2):

- Benralizumab 60 mg Q4W until Week 12, 30 mg Q4W until Week 24, and 30 mg Q8W during the extension period until Week 52 (n = 30)
- Benralizumab 60 mg Q4W until Week 12, 30 mg Q4W until Week 24, and 30 mg Q4W during the extension period until Week 52 (n = 30)
- Benralizumab 30 mg Q4W until Week 12, 30 mg Q4W until Week 24, and 30 mg Q8W during the extension period until Week 52 (n = 30)
- Benralizumab 30 mg Q4W until Week 12, 30 mg Q4W until Week 24, and 30 mg Q4W during the extension period until Week 52 (n = 30)
- Placebo Q4W until Week 24, benralizumab 30 mg Q4W until Week 36, and 30 mg Q8W until Week 52 (n = 40).

Figure 1 presents an overview of IP administration for each randomised treatment group.

Figure 1 Investigational Product Administration

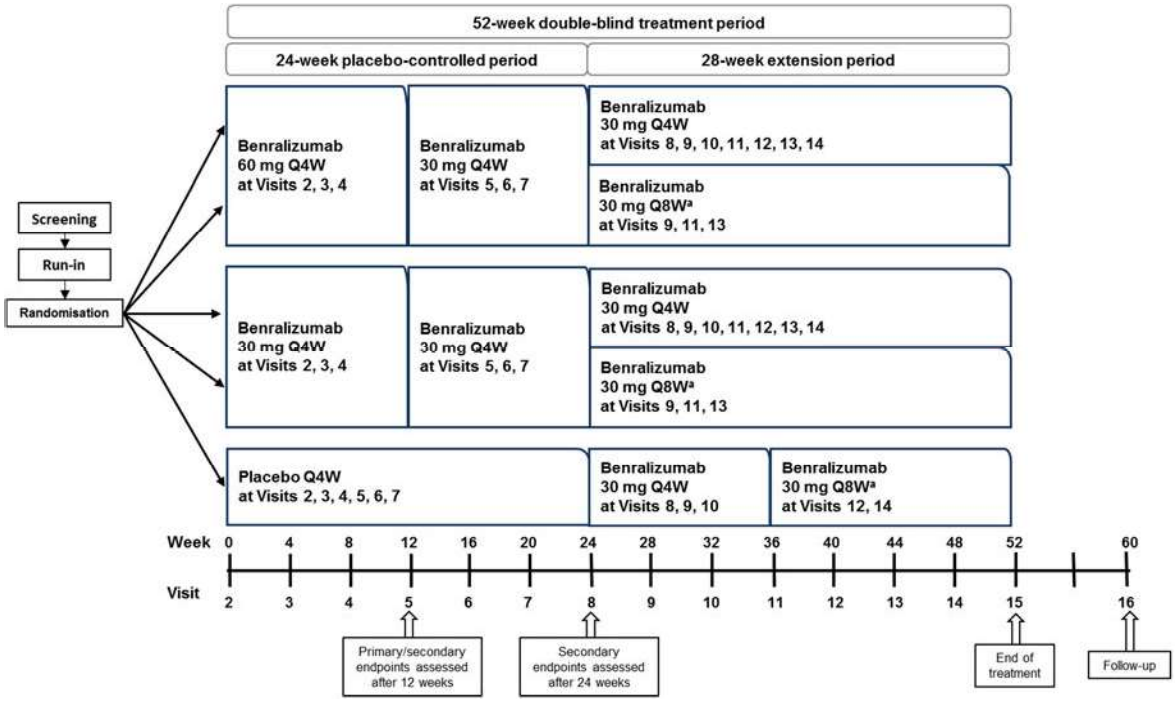


[a] In the extension period, participants will receive benralizumab on a Q8W or Q4W dosing regimen, as predetermined at randomisation (Visit 2). Participants randomised to receive benralizumab Q8W in the extension period will also receive placebo at intervening study visits when they are not receiving benralizumab.
Q4W, every 4 weeks; Q8W, every 8 weeks.

Up to Week 24 analyses at primary database lock (DBL), all analyses will evaluate the three initial dose treatment arms: benralizumab 60 mg (total n = 60), benralizumab 30 mg (total n = 60), and placebo (total n = 40). At final DBL after the extension period, analyses will be evaluating Q4W and Q8W dosing regimens, as well as the five randomised treatment groups.

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 the blind of the dosing regimen in the extension period, participants randomised to receive benralizumab Q8W will receive Q4W investigational product dosing including placebo administered Q8W, 4 weeks after their benralizumab administrations.

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 randomised 3:3:3:3:4 by the interactive voice/web response system (Interactive Voice Response System [IVRS]/Interactive Web Response System [IWRS]) to one of the 5 treatment sequences described in CSP Section 4.1.

Blinded investigational product (IP) (benralizumab or placebo) will be administered by subcutaneous (SC) injection at the investigational site Q4W for up to 48 Weeks. The final dose of benralizumab or placebo will be administered at Week 48 (Visit 14).

During the initial 12-week treatment period, IP will be administered as 2 SC injections at the same anatomical site with a distance of at least 3 cm between the 2 injections. After Week 12, IP will be administered as a single SC injection (CSP Section 6.1).

During the initial 12-week treatment period (Visits 2, 3 and 4) all participants will receive randomised initial dose of study medication Q4W, ie benralizumab 60 mg, benralizumab 30 mg, or placebo (Figure 1 and Figure 2).

After Week 12, all participants randomised to benralizumab (either 60 mg or 30 mg) will receive 30 mg benralizumab Q4W at Visits 5, 6 and 7 (ie, Weeks 12, 16 and 20; Figure 1). Participants receiving placebo will continue to receive placebo Q4W (Figure 1).

In the extension period (ie from Week 24) the dosing regimen of the original 5 randomised treatment sequences will apply Figure 1 and Figure 2). Participants randomised to benralizumab (either 60 mg or 30 mg) will receive benralizumab 30 mg either Q4W or Q8W, at Visits 8 through 14. Participants initially randomised to placebo will receive benralizumab 30 mg Q4W at Weeks 24, 28, and 32 (ie, Visits 8, 9 and 10), followed by benralizumab 30 mg Q8W from Week 36 (ie, Visit 11 through 14, Figure 1 and Figure 2). All participants on Q8W regimen will continue to receive SC injections containing placebo at their Q4W visit intervals (Figure 1), thus maintaining the blinded treatment regimen.

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

1.2 Number of participants

Approximately 240 participants will be enrolled/screened in order to achieve 160 eligible study participants randomly assigned to study intervention.

Up to Week 24, ie, prior to start of the extension period evaluating Q4W and Q8W dosing regimens, participants in both benralizumab 60 mg initial dose treatment arms will be considered as a single treatment group (total n=60), as will participants in both benralizumab 30 mg initial dose treatment arms (total n=60). Together with the 40 participants in the placebo treatment arm the above randomisation will provide a 3:3:2 ratio across the treatment group to assess the primary endpoint at Week 12.

This provides 90% power for statistical significance at the 2-sided 2.5% level within each induction dose comparison versus placebo if the assumed effect is a 4.7 point mean difference in change in ISS7 between benralizumab and placebo, assuming a SD of 6 for the change in ISS7 and a 10% dropout rate. These assumptions are in line with those for previous CSU trials (Saini et al 2015). The proposed sample size also provides a robust dataset to assess the key secondary endpoints, assess potential responders in sub-populations, and make a comparative assessment of the Q4W and Q8W dosing regimens in the maintenance period.

2 ANALYSIS SETS

2.1 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 (major 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. Potential participants who are screened for the purpose of determining eligibility for the study but are not randomised to the study will be considered “screen failures”, unless otherwise specified by the protocol.

2.1.1 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.

2.1.2 Full analysis set

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

Participants will be analysed according to their randomised 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 analyses will be performed using an ITT approach based on the FAS. For consistency, demographic and baseline characteristics will be presented using the FAS.

2.1.3 Safety analysis set

The safety analysis set will comprise 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 randomised 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. A participant who receives both active dose regimens will be classified as the higher active dose regimen (ie, 60 mg). Any deviations from the randomised treatment will be listed and considered when interpreting the safety data.

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

2.1.4 Pharmacokinetic analysis set

All participants who received benralizumab and from whom PK blood samples are assumed not to be affected by factors such as protocol violations (eg, received wrong dose) and who had at least one quantifiable serum PK observation post first dose.

All PK summaries will be based on this analysis set.

2.1.5 Extension period analysis set

All participants who start or carry on receiving at least 1 dose of benralizumab after the end of the Week 24 placebo-controlled period, and thus entering the extension period.

2.1.6 Placebo-to-benralizumab extension analysis set

All participants who received benralizumab at the start of the extension period who were previously randomised to placebo treatment.

2.2 Violations and deviations

Participants who do not meet eligibility criteria but are still randomised and received at least 1 dose of IP will be analysed according to the analysis sets described in Section 2.1. There is no intention to perform a per-protocol analysis in this study.

2.2.1 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 will include but may not be limited to the following categories:

- Eligibility criteria not met (participants incorrectly randomised)
  - 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 study treatment
- Deviations from IP management and administration
- Received prohibited/restricted concomitant medication
- Other important protocol deviations
  - Scheduled study assessments not done or incorrectly performed at baseline and/or >3 post-baseline visits in the same year (separately for UPDD assessments, physical examination, vital signs, safety lab tests, pregnancy tests)

- 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 summarised and listed in the CSR.

2.2.2 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 may 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. Confirmed or suspected cases of COVID-19 will be listed and included as AEs as appropriate.

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 randomisation will serve as the baseline measurement. If time is collected, the assessment performed the same day but at a time prior to randomisation will be included in the baseline definition.

Generally for safety data, the last recorded value on or prior to the first dose of study treatment 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 study treatment will be included in the baseline definition.

For PROs calculated from daily scores, assessments prior to the date of randomisation will be used for baseline. The weekly scores will be calculated using the 7 days prior to the day of randomisation as baseline. If there is no value prior to the date of randomisation, then the baseline value will not be imputed and will be set to missing. Further details on scoring for daily PROs will be provided in Sections 3.2 and 3.3.

For PROs with recall periods of 1 week or more, the time component of the definition of baseline (i.e. time of assessment relative to randomisation) will not be considered; assessments on the same day as randomisation will be used for baseline. If there is no value prior to randomisation (or 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 Sections 3.3 and 3.5.

When time of assessment is not recorded or missing, it is assumed that assessments recorded on the date of first dose of study treatment were performed prior to dosing, except in cases of protocol specified post-dose assessments. If there is no value prior to randomisation (or the first dose of study treatment, 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 study treatment will be used as baseline. 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 24 may be performed where the baseline value is set to the last recorded value prior to starting benralizumab (i.e. likely the Week 24 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.

3.1.3 Visit windows

For PROs calculated from daily scores, scheduled visit days will be used. The weekly scores will be calculated using the 7 days prior to the day of the scheduled visit assessment. Thus, the primary and key secondary endpoint assessments will be based on the actual 7 days before the Week 12 scheduled visit. No visit window will be applied to the scheduled visit day. Further details on scoring for daily PROs will be provided in Sections 3.2 and 3.3.

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 half way 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 randomisation) +1

By this definition, the day of randomisation will be study day 1 and the planned date of Visit 3 (Week 4) will be study day 29 (=28+1), for example.

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 randomisation (medication stop date ≤ date of randomisation).

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

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

Throughout the study, participants will be required to maintain stable doses of their pre-randomisation H_1 antihistamine treatment (CSP Section 6.5.1). Participants will be permitted to use additional doses (up to 4 x recommended dose per day) of the pre-randomization H_1 antihistamine on days on which symptoms become intolerable, and use should be documented in the ePRO device. A switch of the rescue medication for an individual participant is not permitted (CSP Section 6.5.3).

The number of tablets of rescue medication taken is recorded in the handheld ePRO device as part of the daily Urticaria Patient Daily Diary (UPDD):

- During the past 24 hours, how many tablets of rescue medication did you use in order to control symptoms of your skin condition such as itch or hives?

The weekly number of days of taking rescue medication is calculated by summing the daily use of rescue medication (taken) over the past 7 days. The weekly average amount of anti-histamine medication (including background and rescue) is calculated by summing the recorded multiple of licensed dose taken each day over the past 7 days, and dividing by 7 days. Participants who took no rescue medication and only their daily background medication as per protocol would be presented as 1 time the licensed dose of anti-histamine medication.

For each visit, the 7 daily responses prior to that visit will be used for calculating the weekly scores, thus the day of the visit will not be included into the weekly score. These derivations will be consistent to the definition of the primary endpoint variable.

3.2 Primary outcome variable

The change from Baseline in weekly Itch Severity Score (ISS7) at Week 12 will be used as the primary efficacy variable.

The UPDD will be completed twice daily (morning and evening; CSP Table 3) to capture key measures of urticaria disease activity including the ISS. Compliance with twice-daily completion of the UPDD will be reviewed at every visit (CSP Table 2).

Participants will be asked to document twice a day (morning and evening) the intensity of itch they experienced over the previous 12 hours. The criteria used to define itch severity score (ISS) on a 0- to 3-point scale are shown in Table 2.

Table 2 Assessment of CSU itch activity in study participants

Score	Itch severity
0	None
1	Mild
2	Moderate
3	Severe

Scoring criteria are as defined in omalizumab Study Q4882g (EUDRACT: 2010-022785-27).

The mean of the morning and evening itch scores for each day is taken as the ISS for the day. For daily ISS, if either the morning or evening score is missing, the non-missing score for that day (morning or evening) is used as the daily score. If both morning and evening scores are missing, the daily score is not calculated. The ISS7 is the sum of ISS for the previous 7 days.

If a participant has at least 4 non-missing ISS scores, the ISS7 is defined as the sum of the available daily ISS scores divided by the number of days with non-missing ISS scores, multiplied by 7. If there are less than 4 non-missing daily ISS scores then the weekly ISS7 is missing for that week.

For each visit, the 7 daily scores prior to that visit will be used for calculating the weekly scores, thus the day of the visit will not be included into the weekly score.

The minimum ISS7 is 0 and the maximum is 21. The baseline ISS7 will be the sum of the ISS during the 7 days prior to day of randomisation. The ISS7 at Week 12 will be the sum of the daily ISS during the previous 7 days.

For the primary estimand, all available scheduled visit data up to Week 24 will be included regardless of intercurrent events like randomised treatment adherence, rescue medication received, or early discontinuation. The primary analysis will follow a treatment policy approach with missing at random assumption.

Participants will complete all PRO assessments using a handheld device. The handheld device will be the only accepted source of PRO data. In case of technical issues with the device, a web-based emulator of the device will be available.

3.3 Secondary outcome variables

3.3.1 Urticaria Activity Score (UAS7)

UAS7 is part of the UPDD which will be completed twice daily as described above (Section 3.2).

Participants will be asked to document twice a day (morning and evening) the number of hives they experienced over the previous 12 hours on a 0- to 3-point scale as shown in Table 3.

The mean of the morning and evening hive scores for each day is taken as the hives severity score (HSS) for the day. The sum of the HSS and the ISS (see Section 3.2) for the day is the urticaria activity score (UAS) for the day.

Table 3 Assessment of CSU hives activity in study participants

Score	Intensity (number) of hives
0	None
1	Mild (1 – 6 hives/12 hour)
2	Moderate (7 – 12 hives/12 hour)
3	Intense (> 12 hives/12 hour)

Scoring criteria are as defined in omalizumab Study Q4882g (EUDRACT: 2010-022785-27).

HSS7 is the sum of HSS for the previous 7 days. UAS7 is the sum of UAS for the previous 7 days, ie, the sum of ISS7 and HSS7.

Scoring for HSS/HSS7 and UAS/UAS7 with respect to calculations and missing data will be as described for ISS/ISS7 above.

The minimum UAS7 is 0 and the maximum is 42. The baseline UAS7 will be the sum of the UAS during the 7 days prior to day of randomisation. The UAS7 at Week 12 will be the sum of the daily UAS during the previous 7 days.

The key secondary endpoint variable is change from baseline of UAS7 score at Week 12. The primary analysis of the key secondary endpoint will use the same treatment policy strategy estimand as outlined for the primary endpoint.

The change from baseline in UAS7 at Week 24 will be analysed as secondary endpoint.

The change from baseline in UAS7 at Week 52 will be analysed after the extension period.

3.3.2 Proportion of responders of UAS7

The UAS7 is defined for the key secondary endpoint variable above (Section 3.3.1).

The proportion of responders with UAS7 ≤ 6 at Week 12 and Week 24 will be analysed. The proportion of participants with complete UAS7 response (UAS7 = 0) at Week 12 and Week 24 will be analysed.

The proportion of responders with UAS7 ≤ 6 at Week 52 and the proportion of participants with complete UAS7 response (UAS7 = 0) at Week 52 will be analysed after the extension period.

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.

If UAS7 response is missing at a respective analysis visit (e.g. Visit 12 or 24) then the participant will be treated as a non-responder at that visit. Participants with a missing baseline UAS7 response will have missing responder status.

3.3.3 Hives severity score (HSS7)

HSS7 is defined for the key secondary endpoint variable above (Section 3.3.1).

The minimum HSS7 is 0 and the maximum is 21. The baseline HSS7 will be the sum of the HSS during the 7 days prior to day of randomisation. The HSS7 at Weeks 12 and 24 will be the sum of the daily HSS during the previous 7 days.

The change from baseline in HSS7 will be analysed at Weeks 12 and 24.

The change from baseline in HSS7 score at Week 52 will be analysed after the extension period.

3.3.4 Itch Severity Score (ISS7)

The ISS7 is defined for the primary endpoint variable above (Section 3.2).

The change from baseline in ISS7 will be analysed at Week 24 as secondary endpoint.

The change from baseline in ISS7 at Week 52 will be analysed after the extension period.

3.3.5 Time to ≥ 5 point decrease in ISS7

The ISS7 is defined for the primary endpoint variable above (Section 3.2). The time to ≥ 5 point decrease (clinically relevant decrease) in ISS7 will be summarised.

This will be defined as change from baseline of the first ≥ 5 point decrease in ISS7. Time from randomisation to the first clinically relevant decrease is derived as follows:

(Date of first ≥ 5 point decrease − Date of randomisation + 1).

The time to ≥ 5 point decrease for participants who do not experience a ≥ 5 point decrease during the treatment period will be censored at EOT visit for participants who complete the study. Participants who withdraw from the study or are lost to follow-up before EOT visit will be censored at the last visit date at which a decrease in ISS7 could be assessed.

The time to ≥ 5 point decrease in ISS7 will be analysed over the double blind period up to Week 24, and also over the whole study period up to Week 52 after the extension period.

3.3.6 Measures of angioedema activity

The UPDD includes a daily yes/no question asking whether the participant experienced angioedema during the past 24 hours. If yes, the participant is asked a follow-up question about how they treated the swelling:

0 = Did nothing
1 = Took some prescription or non-prescription medication,
2 = Called my doctor, nurse or nurse practitioner,
3 = Went to see my doctor, nurse or nurse practitioner,
4 = Went to the emergency room at the hospital,
5 = Was hospitalized.

The percentage of angioedema free days will be calculated over the past 7 days.

*(number of angioedema free days / number of non-missing responses) * 100*

The number of non-missing responses for angioedema during the past 24 hours is the sum of days with response of yes or no, and excludes missing responses.

Summaries of the percentage of angioedema free days over the past 7 days will be produced. In addition, presence of angioedema in each 7 day period will be calculated where a participant will be considered to have had angioedema if they have at least one day with angioedema present in the 7 day period.

For the follow-up question, the frequency of each response over the past 7 days will be calculated for each participant.

For each visit, the 7 daily responses prior to that visit will be used for calculating the weekly proportion of angioedema free days over the past 7 days, and for the presence of angioedema in the past 7 days, thus the day of the visit will not be included into the weekly assessment. This approach is consistent with the derivations of the primary endpoint.

Measures of angioedema activity at Weeks 12 and 24 will be assessed in all participants and in the subgroups of participants with angioedema at baseline, and those with history angioedema.

Measures of angioedema activity at Week 52 will be assessed in all participants and in the subgroups of participants with angioedema at baseline, and those with history angioedema, after the extension period.

3.3.7 Urticaria Control Test (UCT)

Urticaria disease control will be assessed by the Urticaria Control Test (UCT) using the ePRO device. The UCT is the first valid and reliable tool to assess disease control in participants with chronic urticaria. It has a retrospective approach using a recall period of 4 weeks and responses on 5-point Likert scales. The UCT questions, responses and scoring are shown in Table 4.

Table 4 UCT questions and the scores

1. How much have you suffered from the physical symptoms of the urticaria (itch, hives (welts) and/or swelling) in the last four weeks?				
0 = very much	1 = much	2 = somewhat	3 = a little	4 = not at all
2. How much was your quality of life affected by the urticaria in the last 4 weeks?				
0 = very much	1 = much	2 = somewhat	3 = a little	4 = not at all
3. How often was the treatment for your urticaria in the last 4 weeks not enough to control your urticaria symptoms?				
0 = very often	1 = often	2 = sometimes	3 = seldom	4 = not at all
4. Overall, how well have you had your urticaria under control in the last 4 weeks?				
0 = not at all	1 = a little	2 = somewhat	3 = well	4 = very well

A score between 0 and 4 is assigned to every answer option. Subsequently, the scores for all 4 questions are summed up. Accordingly, the minimum and maximum UCT scores are 0 and 16, with 16 points indicating complete disease control. The UCT score will be calculated when all four questions are answered, otherwise the UCT will be missing. A UCT score of <12 shows poorly controlled urticaria, a UCT score of ≥12 presents well controlled urticaria.

UCT is assessed at site at Visits 1 and 2 and at home every 28 ± 3 days after Visit 2.

Change from baseline in UCT at Weeks 12 and 24 will be analysed.

Change from baseline in UCT at Weeks 52 will be analysed after the extension period.

3.3.8 Chronic Urticaria Quality of Life Questionnaire (CU-Q2oL)

The Chronic Urticaria Quality of Life Questionnaire (CU-Q2oL) is a 23-item assessment of CSU-specific health-related quality of life (Baiardini et al 2005). Participants are asked to rate their CSU symptoms and the impact of their symptoms over the last two weeks on several domains: Pruritus, Swelling, Impact on Life Activities, Sleep Problems, Limits, and Looks. Both the overall score and domain scores will be calculated.

The questions are scored as 1 = not at all, 2 = a little, 3 = moderately, 4 = very much, 5 = extremely. Item and scale scores are such that higher scores indicate worse QoL. To give meaningful scores, the overall and domain scores will be transformed into percentages of maximum possible score. For pruritus and swelling, if one question is unanswered, the domain will be missing. For other domains and overall score, if two or more questions are unanswered, the percentage score is not calculated. The minimum possible score is defined as 0 and the maximum possible score is defined as 100. The domain scores are shown in Table 5.

Table 5 CU-Q2oL domains

Domain	Item	Question
Pruritus	Pruritus / Itching	Q1
	Wheals / Hives	Q2
Swelling	Eyes swelling	Q3
	Lip swelling	Q4
Impact on life activities	Urticaria interferes with my work	Q5
	Urticaria interferes with my physical activities	Q6
	Urticaria interferes with my sleep	Q7
	Urticaria interferes with my spare time	Q8
	Urticaria interferes with my social relationship	Q9
	Urticaria interferes with my eating behavior	Q10
Sleep problems	Do you have difficulties in falling asleep?	Q11
	Do you wake up during the night?	Q12
	Do you feel tired during the day because of your bad night sleep?	Q13
	Do you have difficulties in keeping concentration?	Q14
	Do you feel nervous?	Q15
Limits	Do you feel in a bad mood?	Q16
	Do you have to put some limit in choosing your food?	Q17
	Does urticaria limits your sport activities?	Q22
Looks	Are you troubled by drugs' side effects?	Q23
	Are you embarrassed due to urticaria symptoms?	Q18
	Are you embarrassed in going to public places?	Q19

	Do you have any problems in using cosmetics?	Q20
	Do you have any limits in choosing clothes material?	Q21

The CU-Q2oL is assessed at site at Visits 1 and 2 and at home every 28 ± 3 days after Visit 2. Changes from baseline in CU-Q2oL at Weeks 12 and 24 will be analysed.

Changes from baseline in CU-Q2oL at Week 52 will be analysed after the extension period.

3.3.9 Dermatology Life Quality Index (DLQI)

The Dermatology Life Quality Index (DLQI) is a 10-item assessment of dermatology-specific health-related quality of life (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 overall score and domain scores will be calculated.

Each question /except question 7) is scored on a four-point Likert scale: Very much = 3, A lot = 2, A little = 1, Not at all = 0, Not relevant = 0. 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 DLQI is calculated by summing the score of each question resulting in a maximum of 30 and a minimum of 0. The higher the score, the more quality of life is impaired.

The effects of life scores are calculated as follows:

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

The domain subscales are derived as shown in Table 6.

Table 6 Subscales for DLQI

Section	Questions	Score
Symptoms and feelings	Questions 1 and 2	Maximum 6
Daily activities	Questions 3 and 4	Maximum 6
Leisure	Questions 5 and 6	Maximum 6

Work and school	Question 7	Maximum 3
Personal relationships	Questions 8 and 9	Maximum 6
Treatment	Question 10	Maximum 3

The domain subscales for each section (Table 6) will be presented as a percentage of the maximum score (either 6 or 3). The ePRO does not allow for questions to be unanswered, no consideration is required for missing data.

The DLQI is assessed at site at Visits 1 and 2 and at home every 28 ± 3 days after Visit 2. Changes from baseline in DLQI at Weeks 12 and 24 will be analysed.

Changes from baseline in DLQI at Week 52 will be analysed after the extension period.

3.3.10 Serum benralizumab concentration

All PK samples will be collected before administration of study medication according in accordance with the visit schedule provided in Tables 1 and 2 of the CSP.

Concentration results below the lower limit of quantification (BLQ) will be set to 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.

3.3.11 Anti-drug antibodies (ADA)

Samples for anti-drug antibodies (ADA) will be collected before administration of study medication according (CSP Table 2). ADA variables, such as ADA responses, will be generated and analysed as per the details in Appendix 8.2.

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

3.4 Safety outcome variables

Safety and tolerability will be evaluated in terms of: reported adverse events (AEs) (including serious adverse events (SAEs)), vital signs 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 summarised by means of descriptive statistics and qualitative summaries.

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

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.

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

Treatment emergent 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 study treatment.
- AEs in the on-treatment period are defined as those with onset date between day of first dose of study treatment and the day of the scheduled EOT visit or date of last dose + 30 days, whichever is earlier. If the upper limit is after the end of on-study period, then the upper limit is set to the end of on-study period.
- AEs in the on-study period are defined as those with onset on or after the day of first dose of study treatment.
- AEs in the post-treatment period are defined as those with onset date after the on-treatment period defined above.

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.

On-treatment AEs in the double-blind period are defined as those with an onset date between the day of first dose of double-blind period treatment and day prior to first dose in extension period, or last dose in double blind period + 30 days for those who do not enter the extension period. On-treatment AEs in the extension period are defined as those with an onset date on or after day of first dose in extension period up to day of EOT visit or last dose + 30 days.

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, haematology 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 7 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 haematology 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 analysed 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 entire post-baseline period, including the post-treatment period.

For the liver function tests: aspartate Aminotransferase (AST), alanine Aminotransferase (ALT), alkaline phosphatase (ALP), gamma-GT (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.

Participants who meet any of the following criteria at any point during the study will be flagged:

- AST ≥ 3x ULN
- ALT ≥ 3x ULN
- TBL ≥ 2xULN

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

3.4.3 Vital signs and physical examination

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 the CSP Table 2. 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 7 and classified as low (below range), normal (within range or on limits) or high (above range). All values (absolute and change) falling outside the reference ranges will be flagged.

Table 7 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 (only at baseline) and weight, will be conducted in accordance with the schedule provided in CSP Table 2. 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 the brief physical examination, only information on whether the assessment was performed or not is to be recorded.

3.5 Tertiary/exploratory outcome variables

3.5.1 Patient Global Impression of Severity (PGI-S)

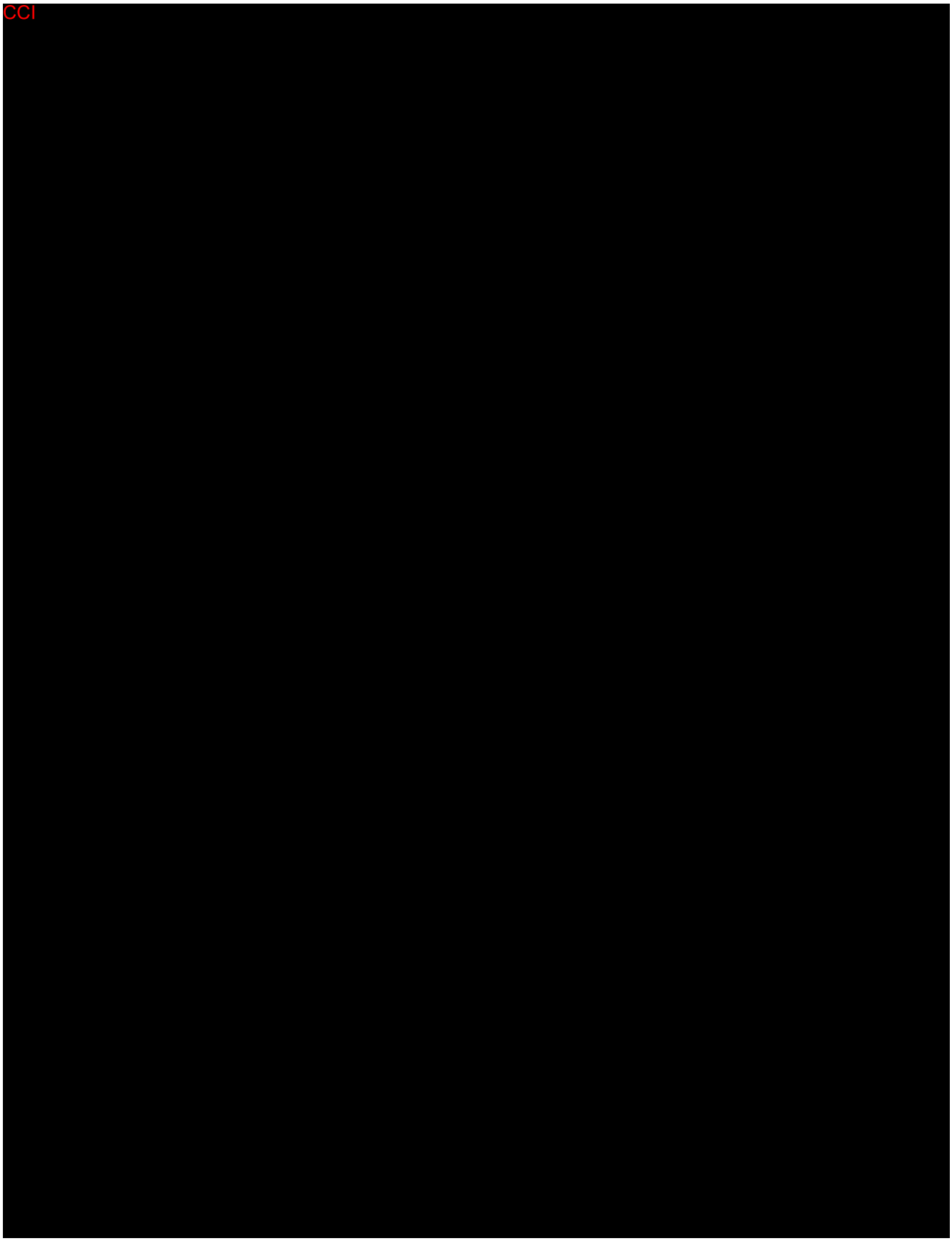
Patient Global Impression of Severity (PGI-S) is a single item designed to capture the participant's perception of overall symptom severity over the past 7 days, using a 6-point categorical response scale (no symptoms, very mild, mild, moderate, severe, very severe).

The responses will be scored from 0 = no symptoms to 5 = very severe. Missing responses will not be imputed.

The PGI-S is collected at site at Visits 1 and 2 and at home every 7 ± 3 days after Visit 2.

The change from baseline in PGI-S score will be analysed at Week 12, Week 24 and at Week 52 after the extension period.

CCI



CCI

[redacted]

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

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

3.5.6 CSU-related healthcare resource utilisation

CSU-related healthcare resource utilisation will be analysed separately and not form part of this SAP.

4 ANALYSIS METHODS

The primary DBL is targeted to occur when all participants have completed the 24-week double-blind 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 IPD/EOT visit.

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, 12-week placebo-controlled induction period, including data up to Week 24. Efficacy endpoints will be analysed using the FAS, the analysis of safety endpoints will be based on the safety analysis set. Analysis sets are defined in Section 2.

Summary data will be presented in tabular format by treatment group. Categorical data will be summarised using frequency and percentages, where the denominator for calculation is the underlying analysis set population unless otherwise stated. Continuous data will be summarised with descriptive statistics of number of available observations, mean, standard deviation (SD), median, minimum and maximum, and quartiles where more appropriate. 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.

The primary estimand, analysed using the treatment-policy strategy, is the difference in mean change from baseline at Week 12 in ISS7 between each benralizumab dose and placebo, using the full analysis set, including all data up to week 12 regardless of discontinuation of IP or whether use of rescue medication has or has not occurred. As all participants who prematurely discontinue from IP in the study are asked to come in for all visits and study assessments up to Week 52, the amount of missing data expected in the analysis is assumed to be low and this estimand strategy is considered appropriate. However, sensitivity analyses may be performed to assess the robustness of the efficacy results to the missing data assumptions as described in Appendix 8.1.

As described above, in general analyses will follow an ITT principle; while changes to background medications for CSU and related treatments are discouraged, if any do occur, data following any changes to background therapies, or post missed doses or discontinuation of IP will still be included in the analyses. However, sensitivity analyses may be performed to assess any impact of changes in study and background medications if deemed appropriate (see Section 4.2.6.3 and Appendix 8.1).

All analyses across the 24-week double-blind period will include 3 treatment groups comprising, respectively, all participants randomised to benraliuzmab 60 mg initial dose (n=60), all participants randomised to benraliuzmab 30 mg initial dose (n=60) and all participants randomised to placebo initial dose (n=40).

For Week 52 analyses, data over the entire 52-week treatment period will be presented in summaries and data listings with the 5 individual treatment groups to initial dose and maintenance regimen. If it is suitable to pool the data across the initial dose groups to enable one comparison of Q4W versus Q8W with 60 participants per arm, then the model will

include 3 treatment groups (those randomised to benralizumab then Q4W maintenance, those randomised to benralizumab then Q8W maintenance and those randomised to placebo followed by benralizumab treatment). If pooling of treatment groups for Q4W versus Q8W dosing is deemed not appropriate, then the model will include the 5 individual treatment groups to initial dose and maintenance regimen.

Key efficacy tables will also summarise data from start of benralizumab, key safety tables will be produced from start of benralizumab to explore changes from starting active study medication.

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 10 Primary and key secondary efficacy and main safety estimands

Statistical Category	Estimand[1]				Section
	Treatment Condition[1]	Endpoint (Population)	Intercurrent Event Strategy[1]	Population Level Summary[1] (Analysis)	
Primary Objective: To determine the clinical efficacy of benralizumab compared to placebo in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment					
Primary	Treatment with benralizumab versus placebo, regardless of compliance, regardless of rescue medication	• Change from baseline in ISS7 at Week 12 (FAS)	• All data up to Week 12 will be included • Treatment discontinuation – treatment policy • Rescue medication – treatment policy	• Difference in least squares mean change from baseline in ISS7 at Week 12 between benralizumab and placebo (MMRM)	4.2.6.1
Key secondary Objective: To evaluate the effect of benralizumab compared to placebo on supportive measures of clinical efficacy in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment					
Key secondary	Treatment with benralizumab versus placebo, regardless of compliance, regardless of rescue medication	• Change from baseline in UAS7 at Week 12 (FAS)	• All data up to Week 12 will be included • Treatment discontinuation – treatment policy • Rescue medication – treatment policy	• Difference in least squares mean change from baseline in UAS7 at Week 12 between benralizumab and placebo (MMRM)	4.2.7.1
Safety Objective: To assess the safety and tolerability of benralizumab in patients with CSU who are symptomatic despite the use of H_1 antihistamine treatment					
Safety	Treatment with benralizumab versus placebo, regardless of compliance, regardless of rescue medication	• AEs • SAEs • Vital Signs • Physical examination (Safety)	Remained adherent to intervention (on-treatment)	• Descriptive	4.2.8

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

4.1.1 Testing strategy to account for multiplicity considerations

To control the overall type I error rate to be ≤ 0.05 for the multiple comparisons in the primary and the key secondary endpoints, a Multiple Testing Procedure (MTP) with Bonferroni-Holm over groups, and Fixed-Sequence within groups will be applied.

Specifically, there will be 2 groups defined by the doses of 30 and 60 mg, and within each group there will be a pre-defined fixed-sequence MTP of comparisons of the primary and key secondary endpoints versus placebo (ie, change from baseline in ISS7 at Week 12, then change from baseline in UAS7 at Week 12) at level of α/2.

If the null hypotheses for 1 dose group can be rejected entirely (ie, significant difference between active versus placebo for both endpoints at α=0.025), one will increase the level to α (ie, 0.05) for the other group. This amounts to using Bonferroni-Holm over groups, and fixed-sequence within groups. This multiplicity adjustment method follows the general results described by Bretz (Bretz et al 2009) and by Burman (Burman et al 2009).

4.2 Analysis methods

4.2.1 Participant disposition

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

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

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 summarised by treatment group and scheduled visit, separately for participants in the FAS.

The number of participants randomised by country and centre will also be summarised by treatment group in the FAS.

4.2.2 Demography data and participant characteristics

Demography and baseline characteristics will be summarised 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. These summaries will be presented for reporting of the placebo-controlled treatment period at primary database lock.

Age will be derived from the date of informed consent-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 electronic case report form (eCRF) will be used.

Various baseline characteristics will also be summarised, including participant characteristics (age (as continuous and categorical variable in 3 categories: <50 years / ≥50 to <65 years / ≥ 65 years), weight, height, BMI, ethnicity, race, region (randomisation stratification variable)). Medical and surgical histories will be summarised separately for past and current conditions.

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 summarised by treatment group and overall. Background medications (e.g. type of antihistamine) will be summarised 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 (taken, not taken) over the past 7 days will be summarised by treatment group and overall for scheduled visits. The frequency and percentage of recorded multiple of licensed dose of antihistamine taken each day over the past 7 days will be summarised by treatment group and overall for scheduled visits.

4.2.4 Study treatments

Exposure to investigational product administration will be calculated in days as:

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

and will be summarised descriptively by treatment group for the safety analysis set.

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

- <12 weeks
- ≥12 weeks and <20 weeks
- ≥20 weeks and <36 weeks

- ≥36 weeks and <52 weeks
- ≥52 weeks.

4.2.5 Compliance

Study treatment compliance with IP administration will be summarised 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.

4.2.6 Primary outcome variable

4.2.6.1 Primary analysis

For the primary endpoint, change from baseline in ISS7 score to Week 12, for each dose of benralizumab the null hypothesis (H0) is that the change in ISS7 at Week 12 on benralizumab is equal to the change in ISS7 at Week 12 on placebo. The alternative hypothesis is that the change in ISS7 with benralizumab treatment is not equal to the change in ISS7 for placebo:

- H0: Change from baseline ISS7 at Week 12 (benralizumab 60 mg – Placebo) = 0
- H1: Change from baseline ISS7 at Week 12 (benralizumab 60 mg – Placebo) ≠ 0
- H0: Change from baseline ISS7 at Week 12 (benralizumab 30 mg – Placebo) = 0
- H1: Change from baseline ISS7 at Week 12 (benralizumab 30 mg – Placebo) ≠ 0

Hypothesis testing for the primary analyses will be performed according to the multiple testing procedure described in Section 4.1.1. If the p-value is less than 0.025 (or 0.05 if alpha can be recycled from the other dose group), the null hypothesis (H0) will be rejected and the alternative hypothesis (H1) will be accepted.

The change from baseline in ISS7 score at Week 12, will be compared between each benralizumab initial treatment group (60 mg and 30 mg) with placebo using a mixed-effect model repeated measurements (MMRM) analysis. The dependent variable in this model will be the change from baseline in ISS7 at post-baseline protocol-specified visits (data included up to the Week 24 visit). Treatment group will be included as an explanatory variable along with the baseline ISS7 score, region, visit and treatment-by-visit interaction as fixed effects. An unstructured covariance structure will be used to model the within-participant errors. If the model fails to converge, the following structures will be attempted in the specified order: Toeplitz, first-order autoregressive, compound symmetric, and variance components.

Results will be presented in terms of least square means (LSMEANS), treatment differences in LSMEANS, 95% CI and p-values for all scheduled visits. The results at Week 12 will be of primary interest. The LSMEANS will be calculated using the OM option in the LSMEANS

statement and the treatment-by-visit interaction to provide an LSMEAN estimate for each scheduled visit. The LSMEANS with 95% CIs will also be presented for each treatment group by scheduled visit.

The missing at random assumption for the analysis is considered appropriate and justified as the primary estimand includes all data collected up to Week 24 in the analysis regardless of discontinuation of randomised treatment, as described in CSP Section 9.4.1, and so the amount of missing data is expected to be low and is considered not likely to impact the primary results.

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

Descriptive summary statistics for scores and change from baseline in ISS7 will be produced by treatment group and visit.

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; as used in randomisation),
- Gender (male, female),
- Age group categories (<65, ≥65 years),
- BMI (≤30 kg/m^2, >30 kg/m^2)
- Race (White, Asian, Other),
- Thyroid function defined by TSH (normal thyroid function, low thyroid function),
- Antihistamine use prior to study entry (labelled dose of H_1 antihistamine, higher than labelled dose/combination treatment),
- Baseline ISS7 (<13, ≥13),
- Baseline UAS7 (<median, ≥median),
- Baseline HSS7 (<median, ≥median),
- Screening period hives over last 7 days (any days without hives, hives on all 7 days*; categories to be confirmed prior to unblinding*),
- Duration of CSU at enrolment (<2, 2-10, >10 years),
- Prior omalizumab use (Omalizumab naïve, Omalizumab failure, Omalizumab prior responder),
- Angioedema at baseline (presence, absence),
- Baseline blood eosinophils (<median, ≥median),
- BHRA status (BHRA+, BHRA-*; if suitable data will be available*).

Low thyroid function is defined as TSH above normal range or participants taking thyroid supplement.

For each of the subgroup factors in turn, a separate MMRM 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 and the treatment by subgroup interaction. Region and baseline ISS7, respectively, will not be a covariate when used as subgroup. 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 difference in lsmeans 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, and as such these analyses are considered as exploratory.

4.2.6.3 Sensitivity analyses for the primary outcome variable

Sensitivity analyses for the primary and key secondary endpoints based on different missing data mechanism assumptions including those expected to be more conservative, such as missing not at random (MNAR), 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 using multiple imputation techniques are specified in Appendix 8.1.

To explore alternative approaches other than the primary analysis treatment policy strategy for exploring the impact of rescue medication use, the following sensitivity analyses may be performed for the change from baseline in ISS7 at Week 12 primary endpoint, and may also be repeated for other secondary endpoints if appropriate:

- Rescue medication use: ISS7 treated as missing for 1 day after the day of any post baseline rescue therapy use (but keep values recorded on the day of rescue use), however any ISS7 recorded as 3 (worst score) will not be set to missing. All data after use of restricted medication will be treated as missing. MMRM analysis as described for the primary model above will be performed on the remaining available data whereby as in the primary analysis, weekly scores will only be derived if there remains at least 4 days with evaluable data.
- Substantial rescue medication use: ISS7 treated as missing after substantial use of rescue medication beyond Week 6, or after any use of restricted medication at any time post baseline. Substantial use will be defined as >2*licensed dose of pre-randomisation antihistamine every day or every two days, for at least 5 consecutive days. Those set to missing data will be imputed by WOCF method. Multiple imputation (MI) techniques (see Appendix 8.1) will be used to impute data missing for other reasons at Week 12 only.

ISS7 will be analysed using an Analysis of covariance (ANCOVA) model, containing the same covariates and factors as the primary analysis model.

4.2.7 Secondary outcome variables

4.2.7.1 Change from baseline UAS7

The key secondary endpoint of change from baseline in UAS7 at Week 12 for both benralizumab 60 mg and benralizumab 30 mg treatment groups will be analysed using an MMRM model in a similar way to that described for the primary endpoint analyses (Section 4.2.6.1).

The dependent variable will be the change from baseline in UAS7 at post-baseline protocol-specified visits up to Week 24, and baseline UAS7 will be included as a covariate as well as the other covariates listed for the primary endpoint analysis.

Descriptive summary statistics for scores and change from baseline in UAS7 will be produced by treatment group and visit.

Sensitivity analyses for the key secondary endpoint may be considered if appropriate for the amount rescue medication usage or missing data, respectively. Details are described above for the primary endpoint.

The secondary endpoint of change from baseline in UAS7 at Week 24 will be analysed using the same method as the key secondary endpoint.

The change from baseline in UAS7 at Week 52 will be analysed with MMRM model as described above, comparing effects on the Q4W and Q8W maintenance regimens.

4.2.7.2 Proportion of responders of UAS7

The proportion of responders (UAS7 ≤ 6) at Weeks 12 and 24 will be compared between the benralizumab 60 mg and placebo treatment groups, and the benralizumab 30 mg and placebo treatment groups, respectively, using logistic regression. The model will include terms for treatment group, region and baseline in UAS7. If the model does not converge due to small proportions, only descriptive statistics will be presented.

The results of the analyses will be presented as an odds ratio, together with its associated 95% CIs and 2-sided p-value. Results will also be transformed into a difference in proportions for ease of interpretation. The number and percentage of responders will also be summarised by randomised treatment.

The proportion of participants with complete UAS7 response (UAS7 = 0) at Weeks 12 and 24 will be analysed using logistic regression in a similar way as described above.

The responder analyses described above use a treatment policy strategy where all responses count even if after use of rescue therapy. Sensitivity analyses at Week 12 using a composite strategy may be explored for the responder endpoints if appropriate. Any substantial use of rescue therapy after Week 6, where substantial is as defined in Section 4.2.6.3, or use of restricted medication at any time post-baseline will impute the participant as a non-responder at Week 12.

The proportion of responders (UAS7 ≤ 6) at Week 52 and the proportion of participants with complete UAS7 response (UAS7 = 0) at Week 52 will be analysed as described above, comparing effects on the Q4W and Q8W maintenance regimens.

4.2.7.3 Hives severity score (HSS7)

The change from baseline in HSS7 at Weeks 12 and 24 will be analysed using the same MMRM method as the key secondary endpoint (Section 4.2.7.1).

The change from baseline in HSS7 at Week 52 will be analysed with MMRM model as described above, comparing effects on the Q4W and Q8W maintenance regimens.

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

4.2.7.4 Itch Severity Score (ISS7)

The change from baseline in ISS7 score at Week 24 will be analysed using the same MMRM method as the key secondary endpoint (Section 4.2.7.1).

The change from baseline in ISS7 at Week 52 will be analysed with MMRM model as described above, comparing effects on the Q4W and Q8W maintenance regimens.

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

4.2.7.5 Time to ≥ 5 point decrease in ISS7

The time to ≥ 5 point decrease (clinically relevant decrease) for ISS7 will be summarised by treatment group.

For time to first clinically relevant decrease, Kaplan Meier summary statistics will be presented by treatment group. The number of participants with at least one ≥ 5 point decrease, number of participants censored, and 25th percentile, median and 75th percentile with 95% CIs (calculated using the complementary log-log transformation) will be presented if calculable. Results will be displayed in Kaplan Meier plots.

4.2.7.6 Measures of angioedema activity

Measures of angioedema activity will be assessed in the overall population and repeated in participants with angioedema at baseline and in those with history angioedema.

Descriptive summaries of proportion of angioedema free days over the past 7 days, and proportion of patients with angioedema in the past 7 days, will be produced by treatment group and visit. A statistical model may also be used to analyse this data if appropriate (MMRM analyses, see Section 4.2.7.1).

For the follow-up question, the frequency and percentages of each response over the past 7 days will be presented for number of participants and for number of responses per participant, presented by treatment group and visit.

The summaries will also be presented for durations of Week 4-12 and Week 4-24.

4.2.7.7 Urticaria Control Test (UCT)

Change from baseline in UCT at Weeks 12 and 24 will be analysed using the same MMRM method as the key secondary endpoint (see Section 4.2.7.1).

The change from baseline in UCT at Week 52 will be analysed with MMRM model as described above, comparing effects on the Q4W and Q8W maintenance regimens.

The UCT score and changes from baseline will be summarised by treatment group and visit. The number and percentage of participants with UCT score of <12 (poorly controlled urticaria) and ≥12 (well controlled urticaria) will be presented by treatment group and visit.

4.2.7.8 Chronic Urticaria Quality of Life Questionnaire (CU-Q2oL)

Change from baseline in CU-Q2oL overall score at Weeks 12 and 24 will be analysed using the same MMRM method as the key secondary endpoint (see Section 4.2.7.1).

Both the overall score and domain scores and changes from baseline will be summarised by treatment group and visit.

4.2.7.9 Dermatology Life Quality Index (DLQI)

Change from baseline in DLQI overall score at Weeks 12 and 24 will be analysed using the same MMRM method as the key secondary endpoint (see Section 4.2.7.1).

Both the overall score, summed effects on life scores and domain scores, will be summarised by treatment group and visit.

4.2.7.10 Serum benralizumab concentration

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

4.2.7.11 Anti-drug antibodies (ADA)

Anti-drug antibodies to benralizumab will be summarised 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.

4.2.8 Safety Analyses

Safety analyses will be performed using the safety analysis set. Summaries for the first 24 weeks will be presented by the 3 induction dose treatment groups as noted in Section 4.2. Summaries for the overall study period will be presented by the 5 randomised treatment sequences, including totals for the Q4W and Q8W dose regimens.

A small number of key tables will also be presented for data in the extension period alone to explore the safety profile in the second 6 months of study treatment were a direct comparison of Q4W versus Q8W dosing can be made for patients previously receiving benralizumab, and to summarise time on benralizumab alone for patients previously receiving placebo.

Participants will be analysed 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 summarised 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 and this is the only period that will be used for the placebo controlled first 24 weeks summaries. Limited summaries will be produced for the on-study and post-treatment periods for the whole study summaries only. AEs in the pre-treatment period (with start date prior to the first dose of IP) will be listed for the placebo-controlled treatment period at primary database lock 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 IP exposure 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 1 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 and preferred term, 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 summarised by system organ class (SOC) and preferred term (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 study treatment and SAEs causing discontinuation from the study will also be summarised.

TEAEs, SAEs and DAEs will be summarised by preferred term 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 administration and injection site) and hypersensitivity (standardized MedDRA query of hypersensitivity) will be summarised by preferred term. The summary of injection site reactions will be summarised by injection site location and number of IP administrations. The summary of TEAEs of hypersensitivity will be presented overall and 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, serious AEs, and AEs leading to discontinuation of IP. 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.

4.2.8.2 Clinical Laboratory Safety Assessments

All protocol-specified continuous laboratory parameters will be summarised descriptively by absolute value at each visit by treatment group, together with the corresponding changes from baseline. All parameters will be summarised in SI units, except for blood eosinophil counts

which will be summarised in both SI and conventional units. Results reported by the central laboratory in conventional units will be converted to SI units for reporting.

For continuous data, the summary statistics presented will be the minimum, Q1, median, Q3, maximum, mean and SD.

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 upper limit of normal (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.

Data for participants who meet the biochemical criteria for Hy's law (ALT or AST ≥3 x ULN and TBL≥2 x ULN) will be presented, which will include all visits for this subset of participants. A line plot of liver biochemistry test results (including ALP, ALT, AST, TBL, and GGT) over time will also be presented for this subset of participants. For all participants who meet the biochemical criteria for confirmed Hy's law, a SAE narrative will be produced.

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.

4.2.8.3 Vital Signs and Physical Examination

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, shift tables from normal to abnormal between baseline and post-baseline visits will be evaluated.

All vital sign and physical examination data will be listed.

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 Patient Global Impression of Severity (PGI-S)

The change from baseline in PGI-S score at Week 12, Week 24 and Week 52 will be analysed using MMRM model as described for secondary endpoint scores above (Section 4.2.7.1).

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

CCI

CCI

[redacted]

5 INTERIM ANALYSES

No formal interim analysis is planned for the study.

The first analysis will be the primary analysis reported once the last participant has reached 24 weeks of follow-up in the study. The final analysis will be performed when all participants have completed the extension period of the study.

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

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

6 CHANGES OF ANALYSIS FROM PROTOCOL

Section 2.1.5 Extension period analysis set is added to identify all participants who enter into the extension period.

Sections 3.3.6 and 4.2.7.6 measures of angioedema activity, analyses added inclusion of all participants and subgroup of participants with history angioedema, in addition of those with angioedema at baseline. This will provide consistency with other studies and greater understanding of the disease and IP treatment.

7 REFERENCES

Baiardini et al 2005
Baiardini I, Pasquali M, Braido F, Fumagalli F, Guerra L, Compalati E, et al. A new tool to evaluate the impact of chronic urticaria on quality of life: chronic urticaria quality of life questionnaire (CU-QoL). Allergy 2005;60:1073-8.

Bretz et al 2009
Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Statist Med 2009;28:586-604.

Burman et al 2009
Burman CF, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. Statist Med 2009;28:739-61.

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.

Saini et al 2015
Saini SS, Bindslev-Jensen C, Maurer M, Grob J-J, Baskan EB, Bradley MS, et al. Efficacy and safety of omalizumab in patients with chronic idiopathic/spontaneous urticaria who remain symptomatic on H1 antihistamines: A Randomized, placebo-controlled study. J Invest Dermatol 2015;135:67–75.

8 APPENDIX

8.1 Accounting for missing data

Missing data can occur due to intercurrent events like non-adherence to randomised treatment, or early discontinuation, or non-adherence of daily UPDD usage.

This section summarises 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 randomised treatment as well as for withdrawal from the study will be presented by treatment to describe why participants discontinue from randomised treatment or withdraw from the study. The time to discontinuation of randomised 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 11). 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 variable ISS7 and key secondary endpoint variable UAS7, the summaries of number and percentage of participants with missing/non-missing daily assessments will be presented by treatment group and visit. This will summarise the amount of daily responses used in the calculation of the weekly scores.

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 24, on participants with missing data not due to reason of interest (e.g. lost-to-follow-up).
2 For each of the imputed datasets obtained in step 1, the remaining missing data due to reason of interest (e.g. AE) up to Week 24, 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 WOCF approach, and then be analysed 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 Weeks 12 and 24.

8.1.3 Primary analysis under the Treatment Policy Estimand using the Missing at Random (MAR) assumption

The primary analysis for the primary endpoint of ISS7 at Week 12 allows for differences in outcomes over the study treatment period up to 12 weeks. This primary analysis includes all data until participants withdraw from the study or up to Week 24, whichever is earlier, regardless of if they discontinue from randomised treatment. The primary analysis uses the MMRM method, treatment group will be included as an explanatory variable along with the baseline ISS7 score, visit and treatment*visit interaction, and region, as explanatory variables, 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 24) from the study regardless of if they discontinue from randomised 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 following 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. [The primary analysis already implements this approach. It will not be repeated.].

(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 has 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 11.

Table 11 Parameters for calculating the imputation under MAR and DRMI

Reason for withdrawal	MAR	DRMI
Adverse Event	Benralizumab	Placebo
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. Based on this review the default assumptions for DRMI as described in b) and Table 11 may be changed. A list of these participants and the assumptions made under DRMI will be documented prior to unblinding of the study.

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 ISS7 change at Week 12 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 (i.e. ≥ 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 summarised as counts and percentages by treatment group. The maximum ADA titre over the on-study period will also be summarised 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 summarised 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 summarised by visit and treatment group.

Key participant information will be listed for participants with positive ADA results, including ADA status, titer, 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

Since ADA will not be tested in the placebo group, no statistical comparisons of benralizumab versus placebo by ADA status (positive/negative) are planned. 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 summarised 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 3.6.1. 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.

8.2.3 ADA and eosinophil levels

Eosinophil levels will be summarised 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 summarised 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:

- 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.

SIGNATURE PAGE – WET INK

Protocol	D3259C00001
Document Name	Statistical Analysis Plan
Document Version	1.0
Document Date	16 October 2020
Wet Ink Signature	
Name	PPD
Position	Document Author, IQVIA Biostatistician
Signature Date	16 October 2020
Signature	PPD

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: d3259c00001-sap-ed-1		
Document Title:	Statistical Analysis Plan Edition 1	
Document ID:	CCI [redacted]	
Version Label:	1.0 CURRENT LATEST APPROVED	
Server Date (dd-MMM-yyyy HH:mm 'UTC'Z)	**Signed by**	**Meaning of Signature**
23-Oct-2020 13:22 UTC	PPD	Content Approval
22-Oct-2020 12:04 UTC	PPD	Content Approval

CCI [redacted]

Electronic Signature Manifestation

This page is a manifestation of the electronic signature(s) used in compliance with the organization's electronic signature policies and procedures.

Signer Full Name	Meaning of Signature	Date and Time
PPD	Document Approval (I certify that I have the education, training and experience to perform this task)	09 Nov 2021 13:53:43 UTC