Statistical Analysis Plan	
Study Code	D325AC00002
Edition Number	3.0
Date	18MAY2023

A Multinational, Randomized, Double-blind, Parallel-group, Placebo-controlled Study to Investigate the Use of Benralizumab as a Treatment Option for Patients with Bullous Pemphigoid (FJORD)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AB-QoL	Autoimmune Bullous Disease Quality of Life Questionnaire
ADA	Anti-drug antibodies
AE	Adverse event
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BDR	Blinded Data Review
BMI	Body mass index
BP	Bullous pemphigoid
BPDAI	Bullous pemphigoid disease area index
CCI [redacted]	[redacted]
CI	Confidence interval
COVID-19	Coronavirus Disease 2019
CSP	Clinical Study Protocol
CSR	Clinical Study Report
DAE	AE leading to discontinuation of IP
DBL	Database lock
DL	Direct likelihood
DRMI	Dropout reason-based multiple imputation approach
eCRF	Electronic Case Report Form
ECG	Electrocardiogram
EOT	End of Treatment
EQ-5D-5L	European Quality of Life-5 Dimensions
FAS	Full analysis set
GGT	Gamma-GT
H0	Null hypothesis
H1	Alternative hypothesis
ICF	Informed consent form
IDMC	Independent Data Monitoring Committee
IGA	Investigator Global Assessment
IgE	Immunoglobulin E
IgG	Immunoglobulin G

Abbreviation or special term	Explanation
IP	Investigational product
IPD	IP discontinuation
ITT	Intention-to-treat
IxRS	Interactive Voice Response System/ Interactive Web Response System
LS	Least square
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
MAR	Missing at random
MMRM	Mixed-effect Model for Repeated Measures
MNAR	Missing not at random
NRS	Numeric Rating Scale
OCS	Oral corticosteroids
OLE	Open-label extension
CCI	[redacted]
PK	Pharmacokinetic
PRO	Patient-reported outcome(s)
PT	Preferred Term
Q4W	Every 4 weeks
CCI	[redacted]
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SC	Subcutaneous(ly)
SoA	Schedule of Activities
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TBL	Total bilirubin
ULN	Upper limit of the normal

AMENDMENT HISTORY

Category*: Change refers to	Date	Description of change	In line with the CSP?	Rationale
Primary or secondary endpoints	01/02/2021	• Primary endpoint updated to CR (previously CR+PR) off OCS at week 36 • Addition of IGA score and change from baseline in IGA score at week 36 as secondary endpoints	Yes	• Updated to the most relevant measure of clinical efficacy • To assess efficacy achieved according to the newly developed IGA tool assessing BP disease severity on an ordinal scale
Statistical analysis method for the primary or secondary endpoints	01/02/2021	• Analysis methods for the relapse-free at 36 weeks and relapse-free at 16 weeks endpoints were updated to be assessed as binary proportions at the timepoints of interest. • Estimands for continuous change from baseline endpoints updated to a composite strategy with LOCF imputation for treatment failure intercurrent events, followed by ANCOVA analyses. • Addition of tipping point sensitivity analysis. • Supportive analyses for BPDAI activity score, BPDAI-pruritus and peak pruritus NRS endpoints specified including cumulative responder plots	Yes	• To ensure the analyses of these endpoints is focused on the timepoints of interest and not influenced by timing of achieving disease control/remission before relapsing • To appropriately account for the occurrence of treatment failure intercurrent events. • To define further sensitivity analyses to assess the robustness of results to missing data assumptions • To support interpretation of the clinical relevance of observed changes from baseline in these endpoints

Change in the sample size	18/05/2023	• Reduction in the sample size from 120 patients to 100 patients	Yes	• Due to challenges recruiting patients with this rare disease. The modified sample size will still allow high power (approximately 85%) for statistical significance to be demonstrated under the assumptions outlined in the sample size justification, and the scientific value of the trial is not substantially impacted whilst maintaining robustness of data and interpretability.
Change to futility analysis stopping guidelines	18/05/2023	• Updated predictive power guidelines modified to a 30% threshold and updated in line with the sample size reduction	Yes	• The stopping guidelines for the interim analysis have been updated to a 30% predictive power guideline to increase the chance of stopping for futility if there is truly no treatment difference, without substantially increasing the chance of incorrectly stopping if there is truly a meaningful treatment difference.

* 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.

The current SAP focuses on the double-blind study period with a high-level description of the descriptive analysis for the open-label extension (OLE) analysis. If required, additional details regarding descriptive analyses in the OLE period will be provided in a separate OLE SAP.

The SAP describes the statistical analyses specified in the clinical study protocol (CSP); any changes with regards to what is already specified in the CSP will be described in Section 6.

1.1 Study objectives

Table 1 Objectives and Endpoints

Objective	Estimand descriptions/Endpoints
Primary	
To compare the clinical efficacy of benralizumab with placebo in participants with symptomatic BP	• Population: Full analysis set. • Endpoint: A binary response, whereby a responder is defined as a participant who is in complete remission while off OCS for ≥ 2 months at Week 36. Otherwise, a participant is a non-responder. • Intercurrent events: Participants who receive restricted medications or withdraw from the study will be considered as non-responders from the time such events occur up to Week 36. • Summary measure: Difference in proportion of participants who are responders between benralizumab and placebo at Week 36.
Secondary	
To compare the effect of benralizumab with placebo on supportive measures of clinical efficacy in participants with symptomatic BP up to Week 36	• Key secondary [a]: Proportion of participants who remain relapse-free up to Week 36. • Key secondary [a]: Cumulative OCS exposure (mg/kg) from baseline to Week 36. • Key secondary [a]: Change from baseline in BPDAI activity score at Week 36. • Key secondary [a]: Change from baseline in BPDAI-Pruritus score at Week 36. • Proportion of participants in sustained complete/partial remission on minimal OCS/off OCS for at least 2 months at Week 36. • Cumulative time (weeks) in complete remission off OCS from baseline to Week 36. • Proportion of participants off OCS by Week 36. • IGA score at Week 36. • Change from baseline in IGA score at Week 36.

Table 1 Objectives and Endpoints

Objective	Estimand descriptions/Endpoints
To compare the effect of benralizumab with placebo on clinical efficacy in participants with symptomatic BP up to Week 16	• Key secondary [a] Cumulative OCS exposure (mg/kg) from baseline to Week 16. • Proportion of participants who remain relapse-free up to Week 16. • Proportion of participants with any clinical benefit (eg, partial and complete remission during taper, with no steroid use, or with minimal steroid use [ie, < 0.1 mg/kg/day]) at Week 16. • Time to disease control, OCS dose (mg/kg) at disease control and time to the end of the consolidation phase. • Change from baseline in BPDAI activity score at Week 16. • Change from baseline in BPDAI-Pruritus score at Week 16.
To estimate the PK and immunogenicity of benralizumab in participants with BP	• Serum benralizumab concentration • ADA
Safety	
To compare the safety and tolerability of benralizumab with placebo in participants with symptomatic BP	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
Tertiary/exploratory	
To explore the effect of benralizumab versus placebo on patient-reported, health-related, quality of life, measures in participants with symptomatic BP	• Changes in Peak Pruritus Numeric Rating Scale • Change from baseline at Week 36 for: o AB-QoL o EQ-5D-5L
CCI [redacted]	

Table 1 Objectives and Endpoints

Objective	Estimand descriptions/Endpoints
To explore the effect of benralizumab versus placebo on healthcare resource utilization in participants with symptomatic BP	• Assessment of all-cause and BP-related healthcare resource utilization over 36 weeks including annualized rate of hospitalizations, length of hospital stay, office visits, emergency room visits, tests and procedures.
CCI	
	• Changes in AB-QoL • Changes in EQ-5D-5L
CCI	• Skin histology and immunohistochemistry • CCI and serum and plasma biomarkers
	• Skin histology and immunohistochemistry • CCI and serum and plasma biomarkers
To evaluate the longer-term steroid-sparing benefit of benralizumab	• Assessment of potential glucocorticoid related events, recorded as AEs, and including changes in BMI, glucose tolerance, and blood pressure
CCI	

[a] Estimands for the key secondary endpoints are detailed in Section 3.3.

AB-QoL, Autoimmune Bullous Disease Quality of Life; ADA, anti-drug antibodies; AE, adverse event; BMI, body mass index; BP, bullous pemphigoid; BPDAI, Bullous Pemphigoid Disease Area Index; CCI CCI ; EQ-5D-5L, European Quality of Life-5 Dimensions; IGA, Investigator Global Assessment; IP, investigational product; OCS, oral corticosteroids; CCI ; PK, pharmacokinetics; CCI .

1.2 Study design

This is a multinational, randomized, double-blind, placebo-controlled, parallel-group, Phase 3 study to assess the efficacy and safety of repeat dosing of benralizumab in adult participants with symptomatic bullous pemphigoid (BP). The study consists of a screening period, a randomized double-blind treatment period, and an open-label extension (OLE).

The primary database lock (DBL) will occur when all randomized participants have had the opportunity to complete at least 36 weeks of follow-up. The final DBL will occur when all participants have completed the OLE period. Additional locks, between the primary and final DBLs, may be performed to report data accumulated during the OLE period to support decision making.

For an overview of the study design see Figure 1. For a detailed list of procedures at each period see Schedule of Activities (SoA) in the CSP (Table 2).

1.2.1 Screening period

After obtaining informed consent, all participants will enter a screening period. The screening period is targeted to last 0 to 4 days depending upon the study assessment requirements (eg. there may be delays beyond 4 days awaiting confirmatory biopsy and serology results).

1.2.2 Double-blind treatment period

Approximately 100 eligible participants will be randomized 1:1 at baseline (Visit 2), to receive either benralizumab or placebo every 4 weeks (Q4W) by subcutaneous (SC) injection. The randomization will be stratified by disease severity and time of bullous pemphigoid (BP) diagnosis (participants with newly diagnosed BP; participants with a previous diagnosis of BP who have relapsed). Bullous pemphigoid disease area index (BPDAI) activity score will be used to assess disease severity, whereby participants with a score greater than or equal to 56 points will be classified as severe. The double-blind treatment period for each participant will last for 36 weeks.

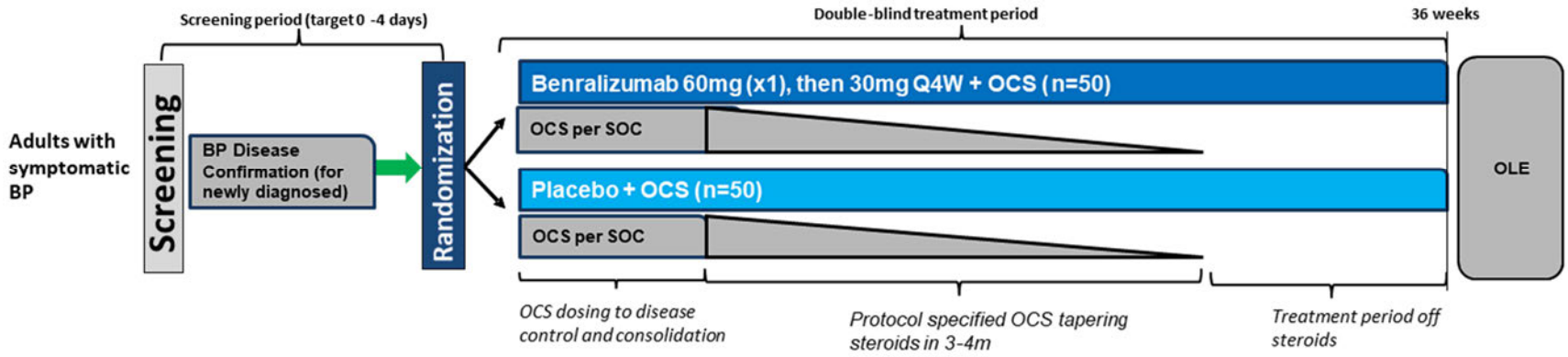
Participants will receive a loading dose of either 60 mg benralizumab or placebo, administered as 2 SC injections at Visit 2. After 4 weeks, participants will receive 30 mg benralizumab or placebo Q4W per their initial randomization until the end of the double-blind treatment period. Starting at Visit 2, participants will also receive oral corticosteroids (OCS) 0.5 mg/kg/day until BP disease control is achieved. In participants who do not achieve disease control, the OCS dose can be increased to achieve disease control. The steroid doses and time necessary to achieve disease control of BP may vary between participants.

After participants have achieved disease control and the end of the consolidation phase has been reached (see Table 4 for definitions of disease control and consolidation phase), participants will begin OCS tapering according to a protocolled taper regimen. The steroid taper will continue according to the protocolled schedule until relapse or until corticosteroid treatment is completely tapered off.

1.2.3 Open-label extension period

After completion of the double-blind treatment period, all participants will have the option of entering an OLE period, during which they will receive benralizumab 30 mg Q4W for at least 1 year. During the OLE period participants can receive corticosteroids (topical and/or oral) for BP relapse at the Investigator's discretion, with the goal of tapering off the steroids to benralizumab monotherapy as soon as clinically indicated.

Figure 1 Study Design



BP, bullous pemphigoid; m, months; n, number of participants randomized; OCS, oral corticosteroids; OLE, open-label extension; Q4W, every 4 weeks; SOC, standard of care.

1.3 Number of subjects

Approximately 140 participants will be enrolled/screened in order to achieve approximately 100 eligible study participants randomly assigned to study intervention.

The sample size calculation was derived for the primary endpoint, the proportion of participants who are in complete remission while off OCS for ≥ 2 months at Week 36. Approximately 50 participants are needed per treatment group to detect a 30% difference between benralizumab and placebo with 85% power. The calculation is based on a 2-sided test and a 5% significance level and assumes a response rate of 50% for benralizumab and 20% for placebo.

There is limited evidence currently available from completed comparative BP trials to date to base the study size assumptions on. However, a mepolizumab Phase 2 trial with a similar participant population, design, and steroid taper rate reported a low proportion of patients relapse free up to 36 weeks (20% versus 33% for mepolizumab versus placebo; Simon et al 2020). From this study, the assumption that an expected remission rate off steroids at Week 36 would be low, at most 20%, was extrapolated. A 30% improvement in this remission rate was recommended by a panel of BP experts to be at the lower end of what may be clinically meaningful, and so was considered to be a conservative assumption.

A non-binding interim futility analysis will be performed by an Independent Data Monitoring Committee (IDMC) when approximately 36 randomized participants (ie, when approximately 36% of the total amount of statistical information is available) have had the opportunity to complete at least 36 weeks of follow-up. By conducting the interim analysis, the overall power will not be affected substantially. Full details are provided in Section 5.

2 ANALYSIS SETS

2.1 Definition of analysis sets

Five analysis sets are defined below: all participant analysis set, full analysis set (FAS), safety analysis set, pharmacokinetic (PK) analysis set and OLE analysis set. Participants must have provided their informed consent. If no signed informed consent is collected, 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 randomized to the study will be considered “screen failures”, unless otherwise specified by the protocol.

Participants who withdraw consent or assent to participate in the study will be included up to the date of their study termination.

2.1.1 All participants analysis set

This analysis set comprises all participants screened for the study, to be used for reporting disposition and screening failures.

2.1.2 Full analysis set (FAS)

This analysis set comprises all randomized participants who received at least 1 dose of investigational product (IP), irrespective of their protocol adherence and continued participation in the study. Participants will be analysed according to their randomized treatment irrespective of whether or not they have prematurely discontinued, according to the intention-to-treat (ITT) principle.

All efficacy analyses will be performed using an ITT approach based on the FAS. Demography and baseline characteristics will be summarized by treatment for the FAS. If there are major differences between the FAS and the safety analysis set, the summaries will be repeated and presented for the safety analysis set.

2.1.3 Safety analysis set

The safety analysis set consists of all participants who have received at least one dose of IP. Participants will be classified according to the treatment they actually received. Erroneously treated participants (eg, those randomized to treatment A but actually given treatment B) are accounted for in this analysis set by assigning them to the treatment they actually received. A participant who has on one or several occasions received active IP is classified as active.

Any deviations from the randomized treatment will be listed and considered when interpreting the safety data.

Safety, and anti-drug antibodies (ADA) data will be based on this analysis set and for whom any post-dose data are available.

2.1.4 Pharmacokinetic analysis set

This analysis set comprises all participants who received at least one dose of 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 1 quantifiable serum PK observation post first dose.

All PK summaries will be based on this analysis set.

2.1.5 OLE analysis set

This analysis set comprises all participants who start or carry on receiving at least 1 dose of benralizumab after the end of the double-blind period.

2.2 Violations and deviations

Participants who do not meet eligibility criteria but are still randomized 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 (for the double-blind period) and final DBL (for the OLE period) and may include but are not be limited to the following categories:

Eligibility criteria not met (participants incorrectly randomized)
Deviations from key inclusion criteria
Deviations from key exclusion criteria
Deviations from informed consent procedures
Discontinuation criteria for IP met but participant not withdrawn from IP
Deviations from IP management and administration
Received prohibited/restricted concomitant medication
Other important protocol deviations
Scheduled study assessments not done or incorrectly performed out of visit window where the deviation is considered to have significant impact on interpretation of participant efficacy or safety data
Unblinding of treatment assignment for reasons unrelated to participant safety

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

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

2.2.2 Impact on analyses due to Coronavirus Disease 2019 (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 will 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.

COVID-19 related protocol deviations are broadly defined as follows:

Changes to visit schedules, missed visits, changes to study procedures

Discontinuation of IP or changes to the IP supply

This detail will form part of the protocol deviation plan.

During the study, confirmed or suspected cases of COVID-19 will be reported as Adverse Events (AEs) as appropriate. A listing of these cases will be produced as part of the statistical analysis.

3 PRIMARY AND SECONDARY VARIABLES

3.1 General definitions

3.1.1 Baseline definition

For efficacy, the last recorded value on or prior to the date of randomization will serve as the baseline measurement. If time is collected and relevant for the analysis, the assessment performed the same day but at a time prior to randomization will be included in the baseline definition.

For safety and PK, 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.

When time of assessment is not recorded or missing, it is assumed that assessments recorded on the date of randomization (or the date of the first dose of study treatment, depending on the endpoint) were performed prior to randomization (or dosing), except in cases of protocol specified post-dose assessments. If there is no value prior to randomization (or dosing), 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 physical examinations baseline data will be collected at Visit 2 before first dose of IP.

For the Peak Pruritus Numeric Rating Scale (NRS), which is calculated as the average of daily scores over 7 days, change will be assessed by comparison to the assessments on the date

when disease control was recorded and on the 6 days after. Further details on scoring for PROs will be provided in Section 3.6.1.

Additional analyses for the participants who switch from placebo to benralizumab at Week 36 may be performed where the baseline value is set to the last recorded value prior to starting benralizumab (ie, likely the Week 36 measurement) to obtain an assessment of the changes occurring while receiving benralizumab.

3.1.2 Double-blind cut-off

The cut-off for data to be considered as in the double-blind period but not in the OLE is as follows:

If a participant transitions into OLE then the cut-off is the date of the first dose of OLE study treatment, ie, likely Visit 13 (Week 36), unless differently stated. If a participant does not transition into OLE then all the data will be included in the double-blind period.

For all assessments scheduled at Visit 13 (Week 36), the last recorded value on or prior to the first dose of OLE study treatment will be considered as in the double-blind period. If time is collected and relevant for the analysis, for example for PK or Biomarkers, only the assessments performed on the same day but at a time prior to the first dose of OLE study treatment will be included in the double-blind period.

When time of assessment is not recorded or missing, it is assumed that assessments recorded on the date of first dose of OLE study treatment were performed prior to dosing, except in cases of protocol specified post-dose assessments and are included in the double-blind period.

For details about how to classify AEs in the different study periods see Section 3.5.1.

Any data that are to be continued to be captured in the OLE period such as AEs and concomitant medications that were ongoing at the time of this cut-off shall remain as ongoing.

If an AE shows as resolved after (and including) the date of Visit 13 (Week 36) but started prior to the date of Visit 13 (Week 36) it will be shown as ongoing in the double-blind period.

A medication stopped on or before the date of Visit 13 (Week 36) (medication stop date ≤ date of Visit 13) will be regarded as stopped during the double-blind period.

A medication will be regarded as ongoing at the end of the double-blind period if it started prior to the date of Visit 13 (Week 36) and was ongoing after the date of Visit 13 (Week 36).

If a medication was started and stopped on the date of Visit 13 (Week 36), it will be considered as ongoing at the end of the double-blind period.

As per Section 3.1.1, for the participants who switch from placebo to benralizumab at Visit 13 (Week 36), OLE baseline is likely to be from the double-blind period.

3.1.3 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.

3.1.4 Visit windows

For efficacy endpoints that present visit-based data, variables will be summarized based on the scheduled days with adjusted analysis-defined visit windows. The adjusted analysis-defined windows will be based on the collection schedule as per the CSP Table 2 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 randomization) +1

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

The adjusted analysis-defined windows for assessments conducted during the double-blind period at every visit from Visit 3 to Visit 13 are summarized in Table 2.

Table 2 Visit windows for assessments conducted at every visit from Visit 3 to Visit 13 – Double-blind treatment period

Adjusted defined window visit*	Scheduled study day	Maximum windows
Week 2	15	2 ≤ Study day ≤ 21
Week 4	29	22 ≤ Study day ≤ 35
Week 6	43	36 ≤ Study day ≤ 49
Week 8	57	50 ≤ Study day ≤ 70
Week 12	85	71 ≤ Study day ≤ 98
Week 16	113	99 ≤ Study day ≤ 126
Week 20	141	127 ≤ Study day ≤ 154
Week 24	169	155 ≤ Study day ≤ 182
Week 28	197	183 ≤ Study day ≤ 210
Week 32	225	211 ≤ Study day ≤ 237
Week 36	253	238 ≤ Study day ≤ Visit 13 study day or 280 whatever is earlier

*All data are re-windowed apart from any baseline data. This is to account for any delayed IP during Visit 2 (Week 0) to still be assessed as baseline.

The adjusted analysis-defined windows for assessments conducted during the OLE period at every visit from Visit 14 are summarized in Table 3.

Table 3 Visit windows for assessments conducted at every from Visit 14 – OLE

Adjusted defined window visit*	Scheduled study day	Maximum windows
Week 40	281	Visit 13 date or 280 whatever is earlier < Study day ≤ 295
Week 44	309	296 ≤ Study day ≤ 323
Week 48	337	324 ≤ Study day ≤ 351
Week 52	365	352 ≤ Study day ≤ 379
Week 56	393	380 ≤ Study day ≤ 407
Week 60	421	408 ≤ Study day ≤ 435
Week 64	449	436 ≤ Study day ≤ 463
Week 68	477	464 ≤ Study day ≤ 491
Week 72	505	492 ≤ Study day ≤ 519
Week 76	533	520 ≤ Study day ≤ 575
Week 88	617	576 ≤ Study day ≤ 659
Week 100	701	660 ≤ Study day ≤ 743
Every 12 weeks	Scheduled Study Day	744 ≤ Study day ≤ Scheduled Study Day + 42

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 two or more values are located on the same side of the target day at the same date and time and are the closest to the target day, the mean of the values will be included in the analysis.
Note: in summaries of extreme values all on-treatment post baseline values collected are used including those collected at unscheduled visits regardless of whether the value is closest to the target visit day.

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

For endpoints which are not collected at every visit, the above rules will be applied to derive adjusted analysis-defined visit windows based on the protocol defined visit schedule for that endpoint. Similarly, for laboratory parameters not collected at every visit.

When converting the number of days to other units, the following conversion factors will be used: 1 week = 7 days; 1 year = 365.25 days.

3.1.5 Prior/concomitant medications

Medications will be categorized according to the WHO Drug Reference List dictionary which employs the Anatomical Therapeutic Chemical (ATC) classification system. The most recent version that will have been released for execution at AstraZeneca/designee per the Data Management Plan will be used.

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

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

3.1.5.1 OCS usage

During the double-blind period, participants will be treated with IP and OCS until BP disease control is achieved (Figure 1), as detailed in the CSP Section 6.5.

During the OLE period, participants can receive corticosteroids (topical and/or oral) for BP relapse at the Investigator's discretion, with the goal of tapering off the steroids to benralizumab monotherapy as soon as clinically indicated.

The participants will be asked to report daily the dosage of OCS taken over the past 24 hours on a provisioned handheld device at home. Details about OCS usage will then be collected at each scheduled visit by the Investigator. Only data collected in the electronic case report form (eCRF) will be included in the main OCS analysis.

A participant will be considered on OCS on a date if that date is included in the interval

OCS start date ≤ date ≤ OCS end date

Only OCS collected in the OCS eCRF form with dose > 0 will be considered. For any of the OCS with missing dose, the dose will be assumed to be the same of the latest of the previous OCS doses with dose > 0. If end date is missing the OCS will be considered ongoing and the participant assumed to be on OCS until last contact for the participant.

Partial/missing dates for OCS during the study (apart from end date for ongoing OCS) are not allowed.

Any increase in OCS dose after BP disease control is achieved will be considered as OCS rescue use.

3.1.5.2 Restricted medications

Restricted medications are detailed in the CSP Section 6.5.2 and will be selected considering a list of ATC codes that will be agreed and documented before unblinding (for the double-blind period) and final DBL (for the OLE period).

Only certain types of restricted medications used after randomisation during the double-blind period will be considered intercurrent events in the analyses, and only these cases will be handled as described in the estimand approaches detailed in Sections 4.2.6.1 and 4.2.6.2.

The restricted medications that will be considered intercurrent events include the following treatments if given to treat the participants BP, or if given for another condition but where prolonged use is deemed likely to substantially affect their BP outcomes:

- Immunosuppressive medications
- Anti-microbial agents with anti-inflammatory properties (eg. Tetracycline antibiotics, dapsone, erythromycin) – these are specific examples only, but review prior to unblinding will confirm if any others should be considered
- Topical corticosteroids
- Any marketed or investigational biologic (monoclonal or polyclonal antibody)
- Other IPs during the study or within 12 weeks after the last dose of benralizumab/placebo

In addition, for the endpoints where an increase in OCS use post disease control is considered an intercurrent event with consequence, this will only be in the case of OCS use with the aim of treating the participants BP, in a similar approach to the restricted medications described above. Short bursts of OCS use for other conditions (eg. COPD/asthma exacerbation) will not automatically trigger the estimand approach described.

A full review on concomitant medications will be performed prior to unblinding and the final list of intercurrent events will be confirmed and documented then.

3.1.5.3 Extent of Exposure to Study Medication

Participant exposure to study medication (in mg/kg, using baseline weight) will be collected and assessed from baseline to Week 16 and from baseline to Week 36. Exposure will be calculated as (last dose date of IP – first dose date of IP + 1).

In relation to the OCS use, for summary tables, exposure in mg/kg will be converted to prednisone equivalents / kg using the following conversion table:

Table 4. Conversion of Total Corticosteroid to Prednisone Equivalent

Systemic Corticosteroid (SCS)	**Approximate equivalence dose (mg)**
Prednisone	10
Prednisolone	10
Cortisone	50
Hydrocortisone	40
Methylprednisolone	8
Triamcinolone	8
Betamethasone	1.2
Budesonide	2.25
Dexamethasone	1.5
Deflazacort	12

3.1.6 Last contact

Last contact with the participant is defined as the latest of the below dates of assessments:

- Date of visits (on-site or home visits, remote audio or telephone visits)
- Date of randomization
- Start and end dates of dosing
- Date of disease state and BPDAI assessment
- Date of collection of laboratory evaluations
- Date of vital sign testing
- Date of physical examinations
- Date of ECGs
- Start and end dates of concomitant medications
- Start and end dates of hospitalization
- Start and end dates of AE

Dates of contact with the participant or dates of completion/discontinuation are not included. However, if at least one assessment among the ones reported above has occurred on the same date, the date of contact or completion/discontinuation will be indirectly included.

3.2 Primary outcome variable

The primary outcome variable is a binary response whereby participants will be classified as responders if they are in complete remission while off OCS for ≥ 2 months at Week 36. Otherwise, participants will be classified as non-responders.

Investigators will be asked to make a clinical assessment of the participants to determine their BP disease state (ie, active disease, disease control, end of consolidation phase, remission, and relapse) from Visit 2 until the end of the double-blind period, during the OLE period, and then in addition at the IP Discontinuation (IPD), End of Treatment (EOT) and follow-up visit.

At each visit, Investigators will be required to assess a participant's outcome measures, including the states defined in Table 4. The outcome measure will be documented in the eCRF.

Table 4 Definitions for Bullous Pemphigoid Outcome Measures

Outcome measure	Definition [b]
Disease control/beginning of consolidation phase	The time at which new lesions cease to form and established lesions begin to heal or pruritic symptoms start to abate
End of consolidation phase	Time at which no new lesions have developed for a minimum of 2 weeks and approximately 80% of lesions have healed and pruritic symptoms are minimal
Sustained partial remission on minimal steroid [a]/off steroid therapy	The presence of transient new lesions that heal within 1 week while the participant is on minimal steroid therapy or off steroid therapy for at least 2 months
Sustained complete remission on minimal steroid [a] /off steroid therapy	The absence of new or established lesions or pruritus while the participant is on minimal steroid therapy/ off steroid therapy for at least 2 months
Relapse	The appearance of 3 or more new lesions per month (blisters, eczematous lesions, or urticarial plaques) or at least one large (> 10 cm diameter) eczematous lesion or urticarial plaques that do not heal within 1 week, or the extension of established lesions or daily pruritus in participants who have achieved disease control

[b] Minimal steroid therapy is defined as less than or equal to 0.1 mg/kg/day of prednisone or equivalent

[c] Definitions as per Murrell et al 2012

Each BP disease state clinical assessment collected during any scheduled/ unscheduled/ IPD/ EOT/ follow-up visits will be included in the Week 36 visit as per adjusted defined visit window, see Section 3.1.4, and analysed accordingly.

The primary endpoint is the proportion of participants who are in complete remission while off OCS for ≥ 2 months at Week 36.

A participant will be considered in complete remission for ≥ 2 months at Week 36 if a complete remission is recorded in the disease status eCRF at the clinical assessment which is assigned to the Week 36 visit when applying the windowing rules described previously, and a complete remission is recorded at least 55 days prior to this Week 36 date, and all evaluable disease states that are recorded during the 55 days between these assessments (ie, overall 8 weeks) are complete remission.

The disease state at each visit will be recorded on the eCRF form collecting the date of assessment and the disease state. A disease state clinical assessment reported in the eCRF as "Not Evaluable" will be considered as missing. Between two consecutive non-missing disease state assessments it will be assumed that the disease state collected at the first assessment is

current until the day before the second disease state assessment, unless intercurrent events happen. The same rule will be applied for two non-consecutive non-missing disease state assessments at windowed visits if at least one missing assessment is detected between them.

A participant will be considered off OCS for ≥ 2 months at Week 36 if off OCS on the day of the clinical assessment which is assigned to the Week 36 visit when applying the windowing rules and on the 55 days (ie, overall 8 weeks) before that date. Further details on the definition of day off OCS are contained in Section 3.1.5.1.

A participant will be considered in complete remission while off OCS for ≥ 2 months at Week 36 if both the definitions above are met.

Note that regardless of prior OCS use, participants who have been in complete remission while off OCS for ≥ 2 months at Week 36 will be classed as responders. For example, if participants require an increase in OCS dose after previously achieving disease control but they still taper and achieve the remission criteria by Week 36, this will be counted as response in the analysis. It is considered unlikely that this situation would emerge often, but given the clinical relevance of still being able to achieve remission and that OCS use is built into the remission definition, it is considered appropriate to count this situation as response in the analysis.

Intercurrent events for this estimand are unfavorable outcomes and will be handled with a composite strategy. The set of intercurrent events consists of participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 and participants who withdraw from the study. The primary estimand will regard these participants as non-responders from the time such events occur up to Week 36.

Participants who withdraw from the study will be considered non-responders starting after the last available non-missing BP disease state clinical assessment.

Participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 will be considered non-responders starting from the date of start of restricted medications. Missing/partial start dates are not allowed for restricted medications.

Participants with missing data at Week 36 for any other reason will be considered as non-responders.

The amount of missing data for this study is expected to be low, as participants who prematurely discontinue IP will be asked to come in for all visits and complete all study assessments up to the end of the study.

3.3 Key secondary efficacy outcome variables

3.3.1 Proportion of participants who remain relapse-free up to Week 36

The proportion of participants who remain relapse-free up to Week 36 will be analyzed as a binary response variable, where participants who remain relapse-free throughout the study up to Week 36 will be considered responders, but otherwise participants will be considered non-responders.

Relapse is defined as the appearance of: 3 or more new lesions per month (blisters, eczematous lesions, or urticarial plaques); or at least one large (> 10 cm diameter) eczematous lesion or urticarial plaques that do not heal within 1 week; or the extension of established lesions or daily pruritus in participants who have achieved disease control.

Details regarding collection of BP disease state clinical assessments and handling of missing data have been reported in Section 3.2.

Participants who need to increase their OCS dose as a rescue therapy to treat their BP after having previously achieved disease control, those who receive medications deemed to be intercurrent events as per Section 3.1.5.2, and those who withdraw from the study will be considered as not being relapse free up to week 36 (i.e. non responders in the analysis).

A supportive analysis of the time to first BP relapse will also be performed in order to explore the timing of relapse events.

The time (weeks) from randomization to the first BP relapse is defined as follows:

(date of first relapse – date of randomization + 1) / 7

Participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 will be considered with a relapse on the date of start of restricted medications.

Participants who need to increase their OCS dose as a rescue therapy to treat their BP after having previously achieved disease control will be considered with a relapse on the date of OCS dose increase recorded with start date > date of disease control assessment.

For restricted medications/OCS during the study missing/partial start dates are not allowed.

Participants who withdraw from the study or are lost to follow-up before the Week 36 assessment and don't have a relapse or a change in medications as described above will be considered as not-relapse free the day after their last available assessment showing them to be relapse-free.

Participants with missing data for any other reason will be considered as not-relapse free.

The time from randomization to the first BP relapse for participants who do not experience any relapse will be censored at their last available assessment showing them to be relapse-free. The same intercurrent events as detailed above in this section will be applied.

All the data in the double-blind period will be included.

3.3.2 Cumulative OCS exposure (mg/kg) from baseline to Week 36 and from baseline to Week 16

Details about OCS usage will be collected at each scheduled visit by the Investigator as reported in Section 3.1.5.1.

Cumulative OCS exposure (mg/kg) from baseline to Week 36 and from baseline to Week 16 will be calculated for each participant over 4 weekly intervals.

For each visit from Visit 4 (Week 4) until Visit 13 (Week 36), excluding Visit 5 (Week 6), the OCS exposure (mg/kg) will be computed by summing the recorded doses taken during each day over the past 28 days, thus the day of the visit will not be included. It will be calculated using the 28 days prior to the day of the scheduled visit. No visit window will be applied to the scheduled visit day.

For Visit 13 (Week 36), if the 4 weekly interval based on the scheduled visit day is including any days after the actual Visit 13 (Week 36) day (ie, days in the OLE period) the 4 weekly OCS exposure will be computed by summing the recorded doses taken during only the days during the double-blind period and rescaling them into a 4 weekly interval. For example, if the Visit 13 (Week 36) 4 weekly interval includes only 25 days in the double-blind period, it will be computed by summing the recorded doses taken during these 25 days, dividing by 25 and multiplying per 28. If less than 14 days are in the double-blind period, the Visit 13 (Week 36) 4 weekly interval will be considered with missing data of OCS exposure.

More details on OCS daily exposure are reported in Section 3.1.5.1.

To compute the dose in mg/kg for each day of OCS assumption the baseline weight will be considered.

The set of intercurrent events for this estimand consists of participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 and participants who withdraw from the study. These events will be handled with an effectiveness strategy, including all available data in a Mixed-effect Model for Repeated Measures (MMRM) until the occurrence of such events. This avoids making potentially implausible imputation assumptions around future OCS use that a composite strategy would involve. For example, for a participant who discontinues at a high daily OCS dose early in the study, imputing a high daily OCS dose for the remainder of the treatment period would result in an imputed

cumulative OCS dose higher than participants would ever realistically receive. Likewise, using a treatment policy strategy is also not considered appropriate, as the use of restricted medications would confound future OCS use. For example, if OCS were discontinued to receive another treatment due to lack of response or intolerability to OCS, the treatment policy approach would show a low cumulative OCS dose, which is a positive outcome and does not fairly represent the overall picture.

Participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 will be considered with missing OCS exposure (mg/kg) from the date of start of restricted medications. Hence for the 4 weekly interval that includes this date, the OCS exposure will be computed by summing the recorded doses taken during only the days before the start of the restricted medications and rescaling them into a 4 weekly interval. For restricted medications during the study missing/partial start dates are not allowed.

Participants who withdraw from the study will be considered with missing OCS exposure (mg/kg) starting from the last contact with the participant. Hence the 4 weekly interval that includes this, the OCS exposure will be computed by summing the recorded doses taken during only the days before the last contact and rescaling them into a 4 weekly interval.

Note that the list of intercurrent events does not include increases in OCS dose after achieving disease control, and all doses of OCS will be included in the derivation of cumulative OCS exposure from baseline to Week 36/Week 16.

3.3.2.1 Cumulative actual OCS exposure (mg) from baseline to Week 36 and from baseline to Week 16

Cumulative actual OCS exposure (mg) from baseline to Week 36 and from baseline to Week 16 for each participant over 4 weekly intervals, will be also computed as additional endpoint analysed in an exploratory way.

The same derivation rules and intercurrent events described in Section 3.3.2 will be applied.

3.3.2.2 Daily OCS dose (mg/kg/day) from baseline to Week 36

Daily OCS dose (mg/kg/day) will be computed for each participant as average daily dose over weekly intervals and will be analysed descriptively as an additional endpoint.

For each week, from Week 4 to Week 36, the average daily OCS dose (mg/kg/day) will be computed as the mean of the recorded doses taken during the past 7 days, thus the day of the visit will not be included. It will be calculated using the 7 days prior to the day of the scheduled visit. No visit window will be applied to the scheduled visit day.

For Visit 13 (Week 36), if the weekly interval based on the scheduled visit day is including any days after the actual Visit 13 (Week 36) day (ie, days in the OLE period) the weekly

average daily OCS dose will be computed by including only the days during the double-blind period in the mean computation.

More details on OCS daily exposure are reported in Section 3.1.5.1.

To compute the dose in mg/kg for each day of OCS assumption the baseline weight will be considered.

The same derivation rules and intercurrent events described in Section 3.3.2 will be applied.

3.3.3 Change from baseline in BPDAI activity score at Week 36

BPDAI activity score is a continuous endpoint that will be completed at the visits specified in the CSP Table 2 and it is part of the Bullous Pemphigoid Disease Area Index assessment.

The Bullous Pemphigoid Disease Area Index is a clinician completed tool that is used for independent disease severity assessment to measure disease extent in BP.

The Investigator will be required to complete the BPDAI at every visit during double-blind and OLE as specified in the CSP Table 2.

The BPDAI tool computes 2 scores: total BPDAI activity and total BPDAI damage.

The total BPDAI activity score (0 to 360) is the arithmetic sum of the 3 subcomponents – cutaneous blisters/ erosions, cutaneous urticaria/erythema, and mucosal blisters/ erosions. If any of the scores are missing the total BPDAI activity score is considered missing.

The total BPDAI damage score (0 to 12) is the arithmetic sum of the items rated regionally for damage caused by more permanent features such as post-inflammatory hyperpigmentation, scarring and other. If any of the items rate is missing the total BPDAI damage score is considered missing.

The BPDAI quantifies lesion number and size thresholds. Lesions are rated based on the regions affected. BPDAI gives additional weighting to areas of the skin primarily affected in BP, such as the limbs, and less emphasis to scalp and face. The BPDAI total activity and BPDAI damage give an indication of disease activity, with higher scores indicating greater disease activity or damage.

The change from baseline in the total BPDAI activity score at Week 36 will be analysed as key secondary endpoint, while the change in the total BPDAI damage score will be summarized descriptively.

Figure 2 Bullous Pemphigoid Disease Area Index Tool (Murrell et al 2012)

BPDAI					
SKIN	ACTIVITY		ACTIVITY		DAMAGE
Anatomical location	Erosions/Blisters	Number of Lesions if <3	Urticaria/ Erythema / Other	Number of Lesions if <3	Pigmentation / Other
	0 absent		0 absent		Absent 0, present 1
	1 1-3 lesions, none > 1 cm diameter		1 1-3 lesions, none >6 cm diameter		
	2 1-3 lesions, at least one > 1 cm diameter		2 1-3 lesions, at least one lesion > 6 cm diameter		
	3 >3 lesions, none > 2 cm diameter		3 >3 lesions, or at least one lesion > 10 cm		
	5 >3 lesions, and at least one >2 cm		5 >3 lesions and at least one lesion > 25 cm		
	10 >3 lesions, and at least one lesion >5 cm diameter or entire area		10 >3 lesions and at least one lesion > 50 cm diameter or entire area		
Head					
Neck					
Chest					
Left arm					
Right arm					
Hands					
Abdomen					
Genitals					
Back/Buttocks					
Left leg					
Right leg					
Feet					
Total skin	/120		/120		
MUCOSA	Erosions/Blisters				
	1 1 lesion				
	2 2-3 lesions				
	5 >3 lesions, or 2 lesions >2cm				
	10 entire area				
Eyes					
Nose					
Buccal mucosa					
Hard palate					
Soft palate					
Upper gingiva					
Lower gingiva					
Tongue					
Floor of Mouth					
Labial Mucosa					
Posterior Pharynx					
Anogenital					
Total Mucosa	/120				

Any participant with missing data at a specific time point will be considered as having a missing score at that time point.

The set of intercurrent events for this estimand consists of participants who use restricted medications, participants who have their OCS dose increased as rescue therapy after having previously achieved disease control and participants who withdraw from the study.

Participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 will have their last BPDAI activity score observed prior to the date of start of restricted medication carried forward to Week 36. If baseline is the last available assessment prior to the date of start of restricted medication, then it will be used to impute the missing data.

Likewise, participants who need to increase their OCS dose as a rescue therapy to treat their BP after having previously achieved disease control will have their last BPDAI activity score observed prior to the date of OCS dose increase recorded with start date > date of disease control assessment carried forward to Week 36. If baseline is the last available assessment prior to the date of OCS dose increase, then it will be used to impute the missing data.

For restricted medications/OCS during the study missing/partial start dates are not allowed. For participants who withdraw from the study without having rescue or restricted medications, missing BPDAI activity score at Week 36 will be imputed using multiple imputation based on participants who don't have rescue/restricted medication use. The same approach will be applied to those participants with missing data at Week 36 for any other reason. Further details are provided in Section 4.2.7.3. For the summary statistics, missing values will remain missing.

3.3.4 Change from baseline in BPDAI-Pruritus score at Week 36

The BPDAI-Pruritus is a continuous endpoint that will be completed at the visits specified in the CSP Table 2.

The BPDAI-Pruritus is a separate component of the BPDAI that asks the participant to grade the severity of pruritus over the past 24 hours, the past week, and the past month. For each recall period, severity of pruritus is rated on an NRS ranging from 0 for no itch to 10 for maximal itching.

The BPDAI-Pruritus score will be computed as the sum of the three components ranging from 0 to 30.

The change from baseline in the BPDAI-Pruritus score at Week 36 will be analysed as a key secondary endpoint.

Any participant with missing data at a specific time point will be considered as having missing severity of pruritus at that time point.

The set of intercurrent events, the estimand approach and the imputation rules are the same as described in Section 3.3.3 for the change from baseline in BPDAI activity score at Week 36.

3.4 Secondary efficacy outcome variables

3.4.1 Proportion of participants in sustained complete/partial remission on minimal OCS/off OCS for at least 2 months at Week 36

A supportive responder analysis will be performed where any participants who achieve complete or partial remission while on minimal OCS or off OCS for at least 2 months at Week 36 will be defined as responders.

The same derivation rules and the same intercurrent events described for the primary analysis will be applied.

3.4.2 Cumulative time (weeks) in complete remission off OCS from baseline to Week 36

Cumulative time in complete remission while off OCS will be derived for each participant over the initial 36-week treatment period.

Cumulative time in complete remission while off OCS will be computed summing the days included in all the time intervals between each remission and the following relapse according to the below rules.

If there is no OCS usage from and inclusive of the date when complete remission is assessed until the day before the date of relapse, then cumulative time in complete remission while off OCS will be computed in weeks as:

(date of relapse – date of complete remission) / 7

If there is OCS usage on the date of complete remission and the complete remission is confirmed at a later assessment after the OCS usage stop date, then cumulative time in complete remission while off OCS will be computed in weeks as:

(date of relapse – first day off OCS after remission and before relapse) / 7

If the OCS usage stop date is after the last assessment of complete remission then the time will be imputed as equal to zero.

If there is OCS usage after complete remission off OCS that is before or on the day prior to the date of relapse then, the cumulative time in complete remission while off OCS will be computed in weeks as:

(last day off OCS after remission and before OCS restart – date of complete remission + 1) / 7

For participants who do not have a relapse after a complete remission, the time interval from the complete remission until the date of the BP disease state allocated to Week 36 will be considered for computing the cumulative time in the interval.

For more details about definition of days on or off OCS see Section 3.1.5.1.

These times will then be assigned to a set of ordered categories:

- Category Value = 0 for cumulative time = 0 days
- Category Value = 1 for 0 < cumulative time ≤ 56 days (8 weeks)
- Category Value = 2 for 56 days (8 weeks) < cumulative time ≤ 112 days (16 weeks)
- Category Value = 3 for 112 days (16 weeks) < cumulative time ≤ 168 days (24 weeks)
- Category Value = 4 for cumulative time > 168 days (24 weeks)

Depending on the distribution of the data, categories might be condensed or expanded. Any such decisions will be documented prior to unblinding.

The set of intercurrent events for this estimand consists of participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 and participants who withdraw from the study.

Participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 will be considered with a relapse on the date of start of restricted medications. For restricted medications during the study missing/partial start dates are not allowed.

Participants who withdraw from the study will be considered with a relapse starting after the last available BP disease state clinical assessment.

Note that the list of intercurrent events does not include increases in OCS dose after achieving disease control, and all doses of OCS will be included in the derivation of cumulative time in complete remission while off OCS.

3.4.2.1 Cumulative time (weeks) in complete or partial remission off OCS from baseline to Week 36

Cumulative time (weeks) in complete or partial remission off OCS from baseline to Week 36 will be also computed, as additional endpoint analysed in an exploratory way.

Derivation rules and intercurrent events described in Section 3.4.2 for the cumulative time in complete remission while off OCS will be applied.

3.4.3 Proportion of participants off OCS by Week 36

Participants will be defined as responders if they are off OCS by Week 36. Otherwise, participants will be classified as non-responders.

A participant will be considered off OCS at Week 36 if the participant is off OCS prior to the date of the BP disease state allocated to Week 36. Further details on the definition of days off OCS are contained in Section 3.1.5.1.

The set of intercurrent events for this estimand consists of participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 and participants who withdraw from the study. The primary estimand will regard these participants as non-responders from the time such events occur up to Week 36.

Participants who withdraw from the study will be considered non-responders starting from the last contact with the participant.

Participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 will be considered non-responders starting from the date of start of restricted medications. For restricted medications during the study missing/partial start dates are not allowed.

3.4.4 IGA score at Week 36 and change from baseline in IGA score at Week 36

An IGA scale uses clinical characteristics to assess overall disease severity at any given time point. This scoring system allows clinicians to rate skin disease activity by general overall impressions. This IGA scale uses key signs of BP with ordinal levels of severity. The IGA is a 5-point categorical scale ranging from 0 (clear) to 4 (severe) with higher scores indicating greater disease activity or damage (Table 5). The IGA uses clinical characteristics based on the number of lesions, blisters, erosions, erythema and the number of anatomical locations. The IGA scale will be completed at the site by the investigator at the visits specified in the CSP Table 2.

Table 5 IGA Scores and Descriptions

Score	Description
0 – Clear	No skin lesions
1 – Almost Clear	Very few small skin lesions
2 – Mild	Few skin lesions
3 – Moderate	Multiple lesions in > 1 anatomical area OR any large erosive area
4 - Severe	Extensive skin lesions/erosive areas

Note: Examples of skin lesions include blisters, small round erosions, and erythema.

IGA score achieved (clear, almost clear, mild, moderate, severe) will be summarized using descriptive statistics by treatment group at each scheduled visit up to Week 36.

For the summary statistics the following approach will be applied to handle intercurrent events and missing data.

- Participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 will have their last observed value prior to the date of start of restricted medication carried forward to subsequent visits. If baseline is the last available assessment prior to the date of start of restricted medication, then it will be used to impute the missing data.

- Participants who need to increase their OCS dose as a rescue therapy to treat their BP after having previously achieved disease control will have their last observed value prior to the date of OCS dose increase recorded with start date > date of disease control assessment carried forward to subsequent visits. If baseline is the last available assessment prior to the date of OCS dose increase, then it will be used to impute the missing data.

- For participants who withdraw from the study without having rescue or restricted medications, missing values will remain missing. The same approach will be applied to those participants with missing data for any other reason.

A responder analysis will be performed where participants will be classified as responders if they achieve IGA 0/1 (ie, clear or almost clear) at Week 36. Otherwise, participants will be classified as non-responders.

The same derivation rules and the same intercurrent events as described for the primary analysis will be applied whereby patients with the intercurrent events will be considered as non responders in the analysis.

3.4.5 Proportion of participants who remain relapse-free up to Week 16

The same derivation rules and the same intercurrent events described for the key secondary analysis on the proportion of participants who remain relapse-free up to Week 36 will be applied for the analysis of the proportion of participants who remain relapse-free up to Week 16.

3.4.6 Proportion of participants with any clinical benefit at Week 16

Participants will be defined as responders if they are in partial remission or complete remission at Week 16 as assessed by Investigator. Otherwise, participants will be classified as non-responders.

Responders will be further categorized based on their disease status and OCS data into:

- Partial remission
    - On minimal steroid (less than or equal to 0.1 mg/kg/day of OCS)
    - Off steroid therapy
- Complete remission
    - On minimal steroid (less than or equal to 0.1 mg/kg/day of OCS)
    - Off steroid therapy

3.4.7 Time to disease control, OCS dose at disease control and time to the end of the consolidation phase

As detailed in Section 3.2, Investigators will be asked to make a clinical assessment of the participants to determine their BP disease state.

Disease control is achieved at the time when new lesions cease to form, and established lesions begin to heal, or pruritic symptoms start to abate.

The time (weeks) from randomization to disease control is defined as follows:

(date of first assessment of disease control – date of randomization + 1) / 7

Participants who do not reach disease control will be reported as having missing time to disease control.

For the Kaplan Meier analyses, the time from randomization to disease control for participants who do not reach disease control will be censored at the date of the Week 36 assessment for participants who complete the double-blind period. Participants who withdraw from the study

or are lost to follow-up before the Week 36 assessment will be censored at the last visit date after which disease state could not be assessed (ie, at the last non-missing disease state clinical assessment).

OCS dose (mg/kg) at disease control will be calculated for each participant as the dose recorded by the Investigator as per Section 3.1.5.1 on the day before the clinical assessment of disease control.

To compute the dose in mg/kg the baseline weight will be considered.

The handling of partial/missing dose or dates for OCS is detailed in Section 3.1.5.1.

The end of the consolidation phase is defined as the time at which no new lesions have developed for a minimum of 2 weeks and approximately 80% of lesions have healed and pruritic symptoms are minimal, as per Investigator assessment.

The time (weeks) from randomization to end of the consolidation phase is defined as follows:

(date of end of the consolidation phase – date of randomization + 1) / 7

For the Kaplan Meier analyses, the time from randomization to the end of the consolidation phase for participants who do not reach the end of the consolidation phase will be censored at the date of the Week 36 assessment for participants who complete the double-blind period. Participants who withdraw from the study or are lost to follow-up before the Week 36 assessment will be censored at the last visit date after which disease state could not be assessed (ie, at the last non-missing disease state clinical assessment).

All the data in the double-blind period will be included.

3.4.8 Change from baseline in BPDAI activity score at Week 16

The same derivation rules and the same intercurrent events described for the key secondary analysis on the change from baseline in BPDAI activity score at Week 36 will be applied for the analysis of the change from baseline in BPDAI activity score at Week 16.

3.4.9 Change from baseline in BPDAI-Pruritus score at Week 16

The same derivation rules and the same intercurrent events described for the key secondary analysis on the change from baseline in BPDAI-Pruritus score at Week 36 will be applied for the analysis of the change from baseline in BPDAI-Pruritus score at Week 16.

3.4.10 Serum benralizumab concentration

All PK samples will be collected according to the CSP Table 2. Pharmacokinetic samples will be collected before administration of IP, where applicable. Pharmacokinetic samples will also be collected at disease control if no IP administration has occurred.

The following samples will be analysed:

Double-blind treatment period:

- PK samples from participants assigned to the benralizumab treatment group

OLE period:

- PK samples at Visit 19 (Week 60) in the OLE period in all groups.

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.

3.4.11 Anti-drug antibodies (ADA)

Blood samples for determination of ADA in serum will be collected as detailed in the CSP Table 2 and prior to IP administration where applicable. Anti-drug antibodies variables, such as ADA responses, will be generated and analysed as per the details in Appendix 8.2.

3.5 Safety outcome variables

Safety and tolerability will be evaluated in terms of: reported AEs (including serious adverse events [SAEs]), clinical laboratory assessments, vital signs, physical examination and 12-lead electrocardiogram (ECG).

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

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

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

3.5.1 Adverse Events

All AEs will be collected starting from screening (Visit 1), throughout the treatment period, and including the follow-up period.

Serious AEs will be recorded from the time the participant signs the informed consent form (ICF), throughout the duration of the study.

Any non-serious AEs with a start date later than 12 weeks after the last dose of IP will be excluded from the analysis.

Adverse events will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA) that will have been released for execution at AstraZeneca/designee per the Data Management Plan.

A per Section 9.4.3.2 of CSP, the following events are considered treatment emergent:

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

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

The handling of partial/missing dates for AEs is detailed in Appendix 8.3.Treatment Emergent AEs 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 follow-up visit or date of last dose + 12 weeks, 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 post-treatment period are defined as those with onset date after the on-treatment period defined above.
AEs in the on-study period are defined as those with onset on or after the day of first dose of study treatment.

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

AEs in the on-treatment double-blind period are defined as those with an onset date between day of first dose of study treatment and day prior to first dose of study treatment in OLE period. For participants who do not transition into OLE all the AEs in the on-treatment period are included in the on-treatment double-blind period.

AEs in the on-treatment OLE period are defined as those with an onset date between the day of first dose of study treatment in the OLE period and the day of the scheduled follow-up visit or date of last dose + 12 weeks, 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.

The Investigator will assess the causal relationship between IP and each AE, and answer 'yes' or 'no' to the question 'Do you consider that there is a reasonable possibility that the event may have been caused by the investigational product?'. For SAEs, the causal relationship will also be assessed for other medication and study procedures. Note that for SAEs that could be associated with any study procedure, the causal relationship is implied as 'yes'.

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

The Investigator will assess the severity of each AE as "Mild", "Moderate" or "Severe". Adverse events that have missing severity will be assumed to be severe.

3.5.2 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 Table 2 during the double-blind and OLE periods. The parameters outlined in the CSP Table 7 will be collected. Local laboratory results will not be included in the analysis.

In summaries, figures, and listings, laboratory 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. The maximum or minimum value post-baseline will be calculated over the double-blind period for all participants and during the OLE period for the participants who are included in the OLE 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.

3.5.3 Vital signs and physical examination

Pre-dose vital signs (pulse, systolic blood pressure, diastolic blood pressure, respiration rate, body temperature) 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.

Physical examinations (complete or brief), including height and weight, will be conducted in accordance with the schedule provided in the CSP Table 2. Baseline measurements will be assessed at screening (Visit 1) or the randomization visit (Visit 2) before first dose of IP. Any new findings or aggravated existing abnormalities, judged as clinically significant by the Investigator, will be reported as an AE as described in the CSP Section 8.3.5.

The complete physical examination will include an assessment of the following: height and weight, 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 following: weight, general appearance, skin, abdomen, cardiovascular, and respiratory system. For the brief physical examination, only information on whether the assessment was performed or not is to be recorded. Weight measurements will be performed in light clothing and without shoes and will be recorded in kilograms.

Body Mass Index will be calculated from the height and weight as follows:

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

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

Absolute values will be compared to the reference ranges in Table 6 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 6 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

3.5.4 ECG

A single 12-lead ECG will be obtained locally during screening using an ECG machine that automatically calculates the heart rate and measures PR, QRS, and QT intervals. The ECG results will be interpreted locally. Any findings will be recorded in the eCRF.

3.6 Tertiary/exploratory outcome variables

3.6.1 Peak Pruritus Numeric Rating Scale

The Peak Pruritus NRS is a one-item assessment that asks participants to rate the intensity of their worst itch during the past 24 hours. The scale ranges from 0 (no itch) to 10 (worst itch imaginable).

Participants will be asked to record their Peak Pruritus NRS assessment using a handheld device at site at Visit 2, then at home every evening starting the day after Visit 2 thereafter

until Week 72 (Visit 22). Note that this endpoint is not collected during the screening period and so a baseline score will not be available. All changes will be assessed in an exploratory manner, relative to disease control.

Weekly mean scores will be computed for each week starting from the 7th day after disease control. Change will be assessed by comparison to the assessments on the date when disease control was recorded and on the 6 days after as per Section 3.1.1.

For each participant, weekly mean scores will be calculated by summing the daily score over the past 7 days. If a daily assessment is missing, the corresponding daily score is missing. If a participant has at least 4 non-missing daily assessments included in the calculation of the weekly score, the weekly mean score is defined as the sum of the available daily assessments in that week, divided by the number of days with non-missing daily assessments. If there are less than 4 non-missing daily assessments included in the calculation of the weekly score, then the weekly mean score is missing for that week.

Binary endpoints will be derived using weekly mean scores, whereby participants will be classified as a responder or a non-responder for each week using pre-specified clinically meaningful thresholds.

As no threshold for a clinically relevant within-person response definition has been established for BP, the ≥ 2–4-point change threshold proposed for atopic dermatitis by Yosipovitch et al 2019 will be used for this study. Atopic dermatitis is an appropriate comparison point as it is another intensely pruritic inflammatory skin condition associated with eosinophilia (Jenerowicz et al 2007). Therefore, in the current study, a participant will be considered to have an improvement for a given week if a ≥ 2-point reduction versus disease control is observed, and will be considered to have a worsening for a given week if a >=2 point increase versus disease control is observed for the weekly mean score. However, depending on distribution of the data, other thresholds for response may be explored. Different approaches might also be explored using a response definition based on achievement of an absolute score. The responder definition will be finalized and documented before unblinding.

The set of intercurrent events for this estimand consists of participants who need to increase their OCS dose as a rescue therapy to treat their BP after having previously achieved disease control, those who receive medications deemed to be intercurrent events as per Section 3.1.5.2, and those who withdraw from the study.

Participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 will be considered as non-responders from the date of start of restricted medications.

Participants who need to increase their OCS dose as a rescue therapy to treat their BP after having previously achieved disease control will be considered as non-responders from the date of OCS dose increase recorded with start date > date of disease control assessment.

For restricted medications/OCS during the study missing/partial start dates are not allowed.

Participants who withdraw from the study or are lost to follow-up and don't have a change in medications as described above will be considered as non-responders from their last available assessment.

Participants with missing data for any other reason will be considered as non-responders.

For the summary statistics of the weekly mean scores the same approach described for the IGA score in Section 3.4.4 will be applied to handle with intercurrent events and missing data.

Peak Pruritus NRS assessment compliance will be also evaluated for each week until Week 72. Peak Pruritus NRS assessment compliance is defined as the total number of actual completed diary entries divided by the number of diary entries expected for a given time period. For weeks in which a visit is planned, the 7 daily scores prior to that visit will be used for calculating the weekly compliance rates, thus the day of the visit will not be included into the weekly compliance rating.

3.6.2 Autoimmune Bullous Disease Quality of Life Questionnaire (AB-QoL)

The AB-QoL is a 17-item assessment of the health-related quality of life impact of autoimmune bullous diseases, including BP. The recall period is one week. The participant will complete the AB-QoL on a tablet at the site as per the CSP Table 2.

Table 7 AB-QoL questions and scores

Please choose an option from the right hand column which most closely correlates to how you felt within the last week.	
Questions	**Scores**
1. In regards to your blistering disease, does your skin burn, sting or hurt in any way?	o All the time = 3 o Sometimes = 2 o Occasionally = 1 o Never = 0
2. In regards to your blistering disease, does your skin itch?	o All the time = 3 o Sometimes = 2 o Occasionally = 1 o Never = 0
3. Have you had to change your clothing because of your blistering disease?	o I have to be very careful with how tight my clothing is and what materials they are made of – I have had to change what I wear all the time = 3 o I have had to change most of the things I wear = 2 o I have had to change some of the things I wear = 1 o I have never had to change what I wear = 0
4. Do you notice your skin heals slowly?	o I notice this all the time = 3 o I notice this sometimes = 2 o I notice this occasionally = 1 o I have never had this problem = 0
5. Do you have difficulty bathing or showering because of your blistering disease?	o All the time = 3 o Sometimes = 2 o Occasionally = 1 o Never = 0
6. In regards to your blistering disease, does your mouth have erosions which are painful?	o All the time = 3 o Sometimes = 2 o Occasionally = 1 o Never = 0
7. In regards to your blistering disease, do your gums bleed easily?	o All the time = 3 o Sometimes = 2 o Occasionally = 1 o Never = 0
8. Does your blistering disease result in you having to avoid food or drinks that you enjoy?	o I can no longer eat any of the foods I used to enjoy = 3 o I can eat some of the foods I enjoy = 2 o I can eat most of the foods I enjoy = 1 o I can eat anything I like = 0

Please choose an option from the right hand column which most closely correlates to how you felt within the last week.	
Questions	**Scores**
9. As a result of your blistering disease, are you embarrassed about your appearance?	o All the time = 3 o Sometimes = 2 o Occasionally = 1 o Never = 0
10. Do you feel depressed or angry because of your blistering disease?	o All the time = 3 o Sometimes = 2 o Occasionally = 1 o Never = 0
11. Do you feel anxious or cannot relax as a result of your blistering disease?	o All the time = 3 o Sometimes = 2 o Occasionally = 1 o Never = 0
12. Do you worry that friends and family find your blistering skin condition tiresome?	o All the time = 3 o Sometimes = 2 o Occasionally = 1 o Never = 0
13. Is your blistering disease causing sexual difficulties?	o All the time = 3 o Sometimes = 2 o Occasionally = 1 o Never = 0
14. Does your blistering disease affect relationships with friends or loved ones?	o I have had to end a relationship because of my disease OR I cannot have a relationship because of my disease = 3 o Relationships are very difficult = 2 o Relationships are a little difficult = 1 o This has not affected my relationships = 0
15. Does your blistering disease affect your social life?	o I cannot go out to socialize any more = 3 o I can only go to some social events = 2 o I can go to most social events = 1 o My social life is not affected = 0
16. Does your blistering disease affect your work or study?	o Yes, I can no longer work or study = 3 o Yes, I find it difficult to work or study = 2 o Yes, it is a little harder than before to work or study = 1 o No, I am not affected OR this is not applicable = 0
17. Do employers discriminate against you because of your blistering disease?	o I cannot find a job due to my blistering disease = 3 o I have had to change jobs due to my blistering disease = 2 o I still have my job but it is more difficult than before = 1 o My employers are completely understanding OR this is not applicable = 0

The AB-QoL total score for the assessment is computed summing the scores of the 17 questions.

The set of intercurrent events for this estimand consists of participants who use restricted medications, participants who have their OCS dose increased after disease control and participants who withdraw from the study. For the summary statistics the same approach described for the IGA score in Section 3.4.4 will be applied to handle intercurrent events and missing data.

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

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

3.6.6 Healthcare resource utilization

At randomization, retrospective BP-related healthcare resource utilization information will be collected with a one-year recall period. At all subsequent visits (see CSP Table 2), BP-related healthcare resource utilization information will be collected with a recall period of 'since last scheduled visit'.

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

- Number and duration of medical care encounters, including surgeries, and other selected procedures (inpatient and outpatient);
- Duration of hospitalization (total days or length of stay, including duration by wards [eg, intensive care unit]);
- Outpatient medical encounters and interventions (including specialist visits, General Practitioner visits, home health care visits, and emergency room visits, etc.);
- Number and type of diagnostic and therapeutic tests and procedures.

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

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

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

3.6.7 Biomarkers

3.6.7.1 Serum and plasma biomarkers

Serum and plasma samples will be collected according to the CSP Table 2 to evaluate the pharmacology of benralizumab as well as biomarkers of eosinophil recruitment, activation, and survival (eg, absolute eosinophil count, IL-5, EDN, eotaxin).

Additional markers of cellular inflammation and/or activation may be assessed including, but not limited to, those associated with humoral autoimmunity (total serum IgE, IgG directed against auto-antigens BP-180 and/or BP-230), T cell subsets (eg, IL-13, IFN-gamma, IL-17), eosinophil granule proteins (eg, MBP, eosinophil peroxidase), epithelial cell damage (eg, TSLP, IL-25, IL-33), itch (IL-31), and pain (substance P). The results from such studies will not be reported in the CSR but in separate reports or publications as appropriate. The corresponding analyses are not covered in this SAP.

3.6.7.2 Tissue histology and immunostaining

Skin histology and immunohistochemistry data will be analysed separately and not form part of this SAP.

CCI

3.6.9 OLE

The effect of benralizumab will be explored in participants in the OLE period by assessing

CCI

[redacted]

Peak Pruritus NRS at each visit up to Week 72
AB-QoL at each visit up to Week 72
EQ-5D-5L at each visit up to Week 72

These endpoints will be derived in the OLE period using the same rules, assumptions and intercurrent events applied for the double-blind period analysis.

CCI

[redacted]

4 ANALYSIS METHODS

4.1 General principles

The primary analysis will be based on the initial 36-week, double-blind, placebo-controlled phase of the study, and will compare all participants randomized to benralizumab until Week 36 against all participants randomized to placebo until Week 36.

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

The current SAP focuses on the double-blind study period with a high-level description of the descriptive analysis for the OLE analysis. If required, additional detail regarding descriptive analyses in the OLE period will be provided in a separate OLE SAP.

The primary CSR will focus on data from the double-blind period as defined in Section 3.1.2. Final data collected during the OLE period will be reported separately.

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 2.1.

Summary data will be presented in tabular format by treatment group. Categorical data will be summarized using frequency and percentages, where the denominator for calculation is the underlying analysis set population unless otherwise stated. Continuous data will be summarized with descriptive statistics of number of available observations, mean, standard deviation, 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.

Hypothesis testing for the primary and key secondary endpoints will be based on a 2-sided test and carried out using a 5% significance level. If a p-value is less than 0.05, the treatment effect favours benralizumab and the decision will be to reject the null hypothesis (H0) and the alternative hypothesis (H1) will be accepted, subject to the testing strategy outlined in Section 4.1.1 below. Any p-values presented for endpoints other than those included in the testing strategy will be nominal (i.e. not multiplicity adjusted).

All analyses for the OLE endpoints will be descriptive, as no placebo-control data will be available for this period. Consequently, no hypothesis testing will be performed.

Sensitivity analyses will be explored to assess the robustness of treatment effects for the primary and key secondary endpoints. Different missing data mechanisms (eg, values missing at random and values missing not at random) and tipping point analyses will be considered to assess the robustness of the treatment effect for these endpoints. The use of multiple imputation approaches will be considered. Full details are reported in Appendix 8.1.

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 8 Primary and key secondary efficacy and main safety estimands

Statistical category	Treatment condition	Endpoint (Population)	Intercurrent event strategy	Population level summary (Analysis)	
Primary Objective: To compare the clinical efficacy of benralizumab with placebo in participants with symptomatic BP					
Primary	Treatment with benralizumab versus placebo, regardless of IP compliance; restricted medication use indicates treatment failure	Proportion of participants who are in complete remission while off OCS for ≥ 2 months at Week 36 (FAS)	• Treatment discontinuation – treatment policy • Rescue with OCS[a] – treatment policy • Restricted medication use – composite (non-responder) • Study withdrawal – composite (non-responder)	Odds ratio and difference in proportion of participants who are responders at Week 36 (logistic regression model)	4.2.6.1
Secondary Objective: To compare the effect of benralizumab with placebo on supportive measures of clinical efficacy in participants with symptomatic BP up to Week 36/up to Week 16					
Key secondary	Treatment with benralizumab versus placebo, regardless of IP compliance; rescue with OCS and restricted medication use indicate treatment failure	Proportion of participants who remain relapse-free up to Week 36 (FAS)	• Treatment discontinuation – treatment policy • Rescue with OCS[a] – composite (not relapse free) • Restricted medication use – composite (not relapse free) • Study withdrawal – composite (not relapse free)	Odds ratio and difference in proportion of participants remaining relapse free up to Week 36 (logistic regression model)	4.2.7.1
	Treatment with benralizumab versus placebo, regardless of IP compliance	Cumulative OCS exposure (mg/kg) from baseline to Week 36 and from baseline to Week 16 (FAS)	• Treatment discontinuation – treatment policy • Restricted medication use – effectiveness (includes OCS dosing up to occurrence of event) • Study withdrawal – effectiveness (includes OCS dosing up to occurrence of event)	Difference in mean cumulative OCS exposure (mg/kg) from baseline to Week 36 and from baseline to Week 16 (MMRM)	4.2.7.2

	Treatment with benralizumab versus placebo, regardless of IP compliance; rescue with OCS and restricted medication use indicate treatment failure	Change from baseline in BPDAI activity score at Week 36 (FAS), change from baseline in BPDAI-Pruritus score at Week 36 (FAS)	• Treatment discontinuation – treatment policy • Rescue with OCS[a] – composite (LOCF) • Restricted medication use – composite (LOCF) • Study withdrawal – hypothetical (multiple imputation)	Difference in mean change from baseline in BPDAI activity score/BPDAI-Pruritus score at Week 36 (ANCOVA)	4.2.7.3/ 4.2.7.4
Safety Objective: To compare the safety and tolerability of benralizumab with placebo in participants with symptomatic BP					
Safety	Treatment with benralizumab versus placebo, regardless of restricted medications.	• AEs • SAEs • Vital Signs • Physical examination (Safety)	• Remained adherent to intervention (on-treatment)	• Descriptive	4.2.9

[a] Rescue with OCS is defined as needing to increase OCS dose after previously having achieved disease control.
ANCOVA, analysis of covariance; BP, bullous pemphigoid; BPDAI, bullous pemphigoid disease area index; FAS, full analysis set; IP, investigational product; LOCF, last observation carried forwards; MMRM, mixed-effect model for repeated measures; OCS, oral corticosteroids.

4.1.1 Testing strategy to account for multiplicity considerations

A hierarchical testing procedure will be used to control the overall Type I error rate at 5% across the primary and key secondary endpoints. The variables will be tested sequentially in the following order using a 2-sided test and a 5% significance level:

1 Proportion of participants who are in complete remission while off OCS for ≥ 2 months at Week 36 (primary endpoint)
2 Proportion of participants who remain relapse-free up to Week 36 (key secondary endpoint)
3 Cumulative OCS exposure (mg/kg) from baseline to Week 36 (key secondary endpoint)
4 Change from baseline in BPDAI activity score at Week 36 (key secondary endpoint)
5 Change from baseline in BPDAI-Pruritus score at Week 36 (key secondary endpoint)
6 Cumulative OCS exposure (mg/kg) from baseline to Week 16 (key secondary endpoint)

If the treatment comparison for (1) is statistically significant, testing will proceed to (2). If the treatment comparison for (2) is statistically significant, testing will proceed to (3). If the treatment comparison for (3) is statistically significant, testing will proceed to (4). If the treatment comparison for (4) is statistically significant, testing will proceed to (5). If the treatment comparison for (5) is statistically significant, testing will proceed to (6). If, at any point in this hierarchy, a null hypothesis cannot be rejected at the 5% significance level in favour of benralizumab, further testing will stop, and no subsequent null hypotheses will be rejected.

4.1.2 Stratification and subgroup variables

Statistical models will include stratification factors to properly handle the possible variation of treatment effect within the strata, as detailed in Section 4.2.

Exploratory analyses will be also carried out to explore the uniformity of the detected overall treatment effect across key, pre-defined, subgroups. These analyses will focus on the primary endpoint, and they may be performed on key secondary endpoints if deemed appropriate. If a small number of participants are in a subgroup, only descriptive statistics will be reported, and formal comparisons might not be performed.

The following stratification and/or subgroup variables are defined:

Disease severity (Moderate, Severe)

The BPDAI activity score at baseline will be used to assess disease severity, whereby participants with a score greater than or equal to 56 points will be classified as severe.

Time of BP diagnosis (Participants newly diagnosed, Participants previously diagnosed and relapsing)

A participant is defined as newly diagnosed if the participant is enrolled during the first episode of BP, and this episode hasn't been controlled prior to the study. The number of relapses in the past 2 years as collected in the eCRF is equal to zero.

A participant is defined as previously diagnosed and relapsing if the participant is enrolled during a relapse of BP. In this case the number of relapses in the past 2 years as collected in the eCRF is ≥ 1.

Time of BP diagnosis and recurrence treatment (Participants newly diagnosed, Participants previously diagnosed and relapsing on OCS, Participants previously diagnosed and relapsing off OCS)

Participants newly diagnosed will be identified as detailed above, while participants previously diagnosed will be further classified into:

Previously diagnosed and relapsing on OCS, if the participant most recent relapse occurred whilst the participant was on OCS, during tapering, or within 2 months of tapering off OCS.

Previously diagnosed and relapsing off OCS, if the participant most recent relapse occurred whilst the participant had been off OCS for more than 2 months.

Age group (at informed consent, < 65 years, ≥ 65 years and < 75 years, ≥ 75 years)

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 eCRF will be used.

Gender (male, female)

Race (White, Asian, Other)

Region (Europe, North America, Rest of World)

OCS dose at randomization (≤ 0.5 mg/kg/day, > 0.5 mg/kg/day; if suitable distribution of data)

Baseline blood eosinophils (< median, ≥ median)

OCS dose at disease control (≤ 0.5 mg/kg/day, > 0.5 mg/kg/day; if suitable distribution of data)

OCS dose at disease control will be calculated for each participant as the dose recorded by the Investigator as per Section 3.1.5.1 on the day before the clinical assessment of disease control.

Stratification and subgroup variables will be based on eCRF data, and not data entered in IxRS. If eCRF data is missing for disease severity or time of BP diagnosis, the IxRS data will be used in the analysis.

4.2 Analysis methods

4.2.1 Participant disposition

Participant disposition will be summarized using the all participants analysis set. The total number of participants will be summarized for the following groups: those who enrolled and

those who were not randomized (and reason). The number and percentage of participants within each treatment group will be presented by the following categories: randomized, received treatment with study drug, completed treatment with study drug, discontinued treatment with study drug (and reason), discontinued treatment with study drug but completed study follow-up, completed the double-blind period, and withdrawn from study (and reason). This will be presented for both the double-blind 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 summarized by treatment group and scheduled visit, separately for participants in the FAS.

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

4.2.2 Demography data and participant characteristics

Demography and baseline characteristics will be summarized by treatment group and overall in the FAS, using frequency and percentages (for categorical variables) and descriptive statistics of n, mean, standard deviation, median, minimum, and maximum, and where appropriate first quartile and third quartile (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 double-blind treatment period at primary DBL.

Various baseline characteristics will also be summarized. These include participant characteristics (ie, age at informed consent [as continuous and categorical variable in 3 categories < 65 years, ≥ 65 years and < 75 years, ≥ 75 years], weight, height, BMI, gender, ethnicity, race, region) and disease characteristics (ie, disease severity, time of BP diagnosis, current active disease duration at screening [as collected in the eCRF in 5 categories ≤ 1 month, >1-≤ 3 months, >3-≤ 6 months, >6-≤ 12 months, > 12 months], number of BP relapses the last 2 years before screening [as collected in the eCRF in 4 categories 0, 1, 2, ≥ 3], clinical presentations [including presence of blisters, pruritus, erosions, erythema, urticaria plaques], anatomic location, OCS dose at randomization [as continuous and categorical variable in 4 categories ≤ 0.25 mg/kg/day, > 0.25 mg/kg/day and ≤ 0.50 mg/kg/day, > 0.50 mg/kg/day and ≤ 0.75 mg/kg/day, > 0.75 mg/kg/day], baseline blood eosinophil count [as continuous and categorical variable in 4 categories < 150 cells/µL, ≥ 150 cells/µL and < 300 cells/µL, ≥ 300 cells/µL and < 450 cells/µL, ≥ 450 cells/µL], baseline total serum IgG directed against auto-antigens BP-180, baseline total serum IgG directed against auto-antigens BP-230).

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 eCRF will be used.

Medical and surgical histories will be coded using the most recent version of the MedDRA that will have been released for execution at AstraZeneca/designee per the Data Management Plan and will be summarized by MedDRA Preferred Term (PT) within MedDRA System Organ Class (SOC).

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 restricted concomitant medications (refer to CSP Section 6.5.2) will be summarized by treatment group and overall. OCS will be included in the above summaries, but also reported separately. The frequency and percentage of participants reporting OCS usage as rescue medication will be summarized by treatment group and overall for scheduled visits.

The concomitant medications will be summarized over the double-blind period and over the OLE period, respectively. A summary table of background medications at baseline will also be provided.

Concomitant medications will be classified according to the WHO Drug Reference List dictionary. The summary tables will present data by generic term within Anatomical Therapeutic Chemical code.

4.2.4 Study treatments

Exposure to IP administration will be calculated in days as:

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

and will be summarized descriptively by treatment group for the safety analysis set. Exposure will be computed overall, during double-blind period and during OLE.

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

<16 weeks
≥16 weeks and <24 weeks
≥24 weeks and <36 weeks
≥36 weeks and <52 weeks
≥52 weeks and <76 weeks
≥76 weeks and <104 weeks
≥104 weeks

4.2.5 Compliance

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

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

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

Treatment compliance during the double-blind period will be calculated using the sum of administrated injection doses divided by the total number of injection doses until Week 36/IPD visit (excluding the first dose of OLE IP).

4.2.6 Primary outcome variable

4.2.6.1 Primary analysis

The primary estimand will be based on participants in the FAS and will be used for the analysis of the primary endpoint, a binary response, whereby responders are defined as participants who are in complete remission while off OCS for ≥ 2 months at Week 36. Otherwise, participants are non-responders. More details on derivation rules and intercurrent events are given in Section 3.2.

For the first primary endpoint of proportion of patients achieving a histological response at Week 36, the null hypothesis is that the odds of responding on breanlizumab 30mg Q4W is equal to the odds of responding on placebo (this can be interpreted as the proportion of responders on benralizumab 30 mg Q4W is equal to the proportion of responders on placebo). The alternative hypothesis is that the odds of responding on benralizumab 30mg Q4W is not equal to the odds of responding on placebo, i.e.:

H_0: *Odds ratio (benralizumab 30mg / Placebo) = 1*

H_1: *Odds ratio (benralizumab 30mg / Placebo) ≠ 1*

If a p-value is less than 0.05, the treatment effect favours benralizumab and the decision will be to reject the null hypothesis (H0) and the alternative hypothesis (H1) will be accepted.

For the primary analysis, a logistic regression model will be fitted to the primary endpoint using a logit link function. The model will include treatment, baseline disease severity (moderate, severe) and time of BP diagnosis (participants with newly diagnosed BP, participants with a previous diagnosis of BP who have relapsed) as categorical covariates.

Results will be presented in terms of an odds ratio, where >1 will favour benralizumab treatment, and associated 2-sided 95% CIs. Adjusted response rates and the difference in response rates (benralizumab - placebo) and associated 2-sided 95% CIs will also be derived from the model using marginal standardization methods (Bartlett 2018).

A p-value, corresponding to a 2-sided test, will be presented to compare the benralizumab and placebo treatment groups.

Note that if the logistic regression model does not converge to a solution due to low response rates in certain categories, the endpoint will be analysed using a Cochran–Mantel–Haenszel test.

To support this analysis, the observed proportion of responders will be summarized using descriptive statistics by treatment group and each scheduled visit up to Week 36.

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 stratification factors listed in Section 4.1.2.

For each of the subgroup factors in turn, a separate logistic regression model will be fitted using the same model terms as used for the primary analysis (described in Section 4.2.6.1), with additional terms for the subgroup main effect and the treatment by subgroup interaction. Disease severity and time of BP diagnosis will not be a covariate when used as subgroup. If any model does not converge, respective sub-groups may be collapsed appropriately.

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

Results for the subgroups based on OCS dose at disease control will be reported separately, since OCS dose at disease control is not a baseline variable.

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.

4.2.7 Key Secondary efficacy outcome variables

4.2.7.1 Proportion of participants who remain relapse-free up to Week 36

The proportion of participants who remain relapse-free up to Week 36 will be analyzed as a binary response variable, where participants who remain relapse-free throughout the study up to Week 36 will be considered responders, but otherwise participants will be considered non-responders. More details on derivation rules and intercurrent events are given in Section 3.3.1.

The null hypothesis is: odds of responding on breanlizumab is equal to the odds of responding on placebo (this can be interpreted as there is no difference in the proportion of participants who remain relapse-free up to Week 36 between benralizumab and placebo). Whereas, the alternative hypothesis is: the odds of responding on benralizumab is not equal to the odds of responding on placebo. That is:

H_0: *Odds ratio (benralizumab 30mg / Placebo) = 1*

H_1: *Odds ratio (benralizumab 30mg / Placebo)* $\neq 1$

The proportion of participants who remain relapse-free up to Week 36 will be analyzed using the same method described for the primary endpoint as detailed in Section 4.2.6.1

As a supportive analysis to the proportion of participants who remain relapse-free, time to first BP relapse will be analysed using a log-rank test (performed using the LIFETEST procedure with a TEST statement) stratified by baseline disease severity and time of BP diagnosis. A p-value, corresponding to a 2-sided test, will be presented to compare the benralizumab and placebo treatment groups.

The hazard ratio and its 95% CI will be estimated from a stratified Cox Proportional Hazards model with baseline disease severity and time of BP diagnosis as covariates. 95% CI will be calculated using a profile likelihood approach.

The tie handling recommended approach for the log-rank test is the Breslow method and for the Cox proportional hazards model is the Efron method.

An unadjusted analysis (an unstratified log-rank test and a Cox proportional hazards model adjusting for treatment group only) may be conducted as a sensitivity analysis. If there are convergence issues when the study is unblinded (for example due to a stratum experiencing zero events), an unadjusted analysis will replace the analysis outlined above.

The median time to first relapse will be estimated from the Kaplan-Meier curves for the benralizumab and placebo treatment groups, where calculable.

4.2.7.2 Cumulative OCS exposure (mg/kg) from baseline to Week 36 and from baseline to Week 16

Cumulative OCS exposure (mg/kg) from baseline to Week 36 and from baseline to Week 16 will be calculated for each participant over 4 weekly intervals. More details on derivation rules and intercurrent events are given in Section 3.3.2.

The null hypothesis is: the mean for benralizumab is equal to the mean for placebo. Whereas, the alternative hypothesis is: the mean for benralizumab is not equal to the mean for placebo. That is:

H_0: *Difference in means (benralizumab – placebo)* $= 0$

H_1: *Difference in means (benralizumab – placebo)* $\neq 0$

This endpoint will be analysed by a MMRM model.

The cumulative exposure over the 36 (and 16) weeks will be estimated as the sum over the relevant 4-week periods from the MMRM model. This analysis method was selected to utilize all available data in the situation of any participants dropping out prior to 36 weeks. This ensures that any such participants can still contribute to the analysis for the period they remain in the study, without needing to make potentially implausible imputation assumptions around future OCS exposures in these situations (eg, carrying forward a high last observation of 0.75 mg/kg/d, which would not be a feasible dose to maintain for a 36-week period).

The model will include cumulative OCS exposure (mg/kg) as the dependent variable; treatment, 4 weekly interval, baseline disease severity and time of BP diagnosis as categorical covariates, treatment-by-4 weekly interval as an interaction term.

An unstructured variance-covariance matrix will be used to model within-participant errors, and the Kenward-Roger approximation will be used to estimate denominator degrees of freedom and to adjust standard errors. If the model fails to converge, alternative variance-covariance models will be tried in the following order: Toeplitz, first-order regressive, compound symmetry.

For each 4 weekly interval and the whole of the 36-week treatment period, and for the initial 16-week treatment period, the model will be used to estimate the mean cumulative OCS exposure (mg/kg) for each treatment group and the difference versus placebo, with corresponding 95% confidence limits.

Results will be presented in terms of least square (LS) means of cumulative OCS exposure (mg/kg), treatment differences in LS means of cumulative OCS exposure (mg/kg), 95% CI and p-values for all scheduled visits. The cumulative OCS exposure (mg/kg) at Week 36 and Week 16 will be of primary interest. The LS means will be calculated using the OM option in the LSMEANS statement and the treatment-by-visit interaction to provide a LS means estimate for each scheduled visit. The LS means with 95% CIs will also be presented for each treatment group by scheduled visit.

P-values, corresponding to 2-sided tests, will be presented to compare the benralizumab and placebo treatment groups for each interval, including the whole of the 36-week treatment period and the initial 16-week treatment period.

The following SAS code is given as an example:

```
lsmestimate Treatment * 4WeeklyInterval

'Treatment 1: 16 weeks'    0  0  0  0  0  0  0  0  0
                           1  1  1  1  0  0  0  0  0,
'Treatment 0: 16 weeks'    1  1  1  1  0  0  0  0  0
                           0  0  0  0  0  0  0  0  0,
'Diff: 16 weeks'          -1 -1 -1 -1  0  0  0  0  0
                           1  1  1  1  0  0  0  0  0,
'Treatment 1: 36 weeks '   0  0  0  0  0  0  0  0  0
                           1  1  1  1  1  1  1  1  1,
'Treatment 0: 36 weeks '   1  1  1  1  1  1  1  1  1
                           0  0  0  0  0  0  0  0  0,
'Diff: 36 weeks'          -1 -1 -1 -1 -1 -1 -1 -1 -1
                           1  1  1  1  1  1  1  1  1
/ om cl;
```

To support this analysis, observed means will be summarized using descriptive statistics by treatment group and 4 weekly interval, including the whole of the 36-week treatment period and the initial 16-week treatment period. Observed means will be also plotted by treatment group and 4 weekly interval.

To explore the uniformity of the detected overall treatment effect, the model might also be fitted considering OCS dose at randomization and its interaction with treatment as additional covariates.

The set of intercurrent events for this estimand and the imputation approach are detailed in Section 3.3.2. Since any BP relapses are expected to be treated with OCS (which is incorporated in the endpoint definition), the use of restricted medications which would result in exclusion of data beyond that point are expected to be infrequent. Therefore, the amount of missing data is expected to be low and is considered not likely to impact the results.

Sensitivity analyses might be conducted based on different missing data mechanism assumptions, including those expected to be more conservative, such as MNAR, to explore the robustness of any treatment effect, if the amount of missing data warrants further investigation. This might include utilizing multiple imputation approaches including the dropout reason-based multiple imputation where dropouts related to treatment are imputed using a jump to reference approach. Details using multiple imputation techniques are specified in Appendix 8.1.

Cumulative actual OCS exposure (mg) will be also summarized in an exploratory way using descriptive statistics by treatment group and 4 weekly interval, including the whole of the 36-week treatment period and the initial 16-week treatment period.

Daily OCS dose (mg/kg/day) will be also summarized in an exploratory way using descriptive statistics by treatment group and weekly interval from Week 2 to Week 36.

4.2.7.3 Change from baseline in BPDAI activity score at Week 36

The change from baseline in the total BPDAI activity score at Week 36 will be analysed by an ANCOVA model. More details on derivation rules and imputation rules to handle intercurrent events are given in Section 3.3.3.

The null hypothesis is: the mean for benralizumab is equal to the mean for placebo. Whereas, the alternative hypothesis is: the mean for benralizumab is not equal to the mean for placebo. That is:

H0: Difference in means (benralizumab – placebo) = 0

H1: Difference in means (benralizumab – placebo) ≠ 0

The model will include change from baseline at Week 36 as the dependent variable, baseline BPDAI activity score as a continuous covariate and treatment and time of BP diagnosis as categorical covariates. Only patients with baseline and at least one post-baseline BPDAI activity score will be included in the model.

The model will be used to estimate, in terms of LS means, the mean change from baseline at Week 36 for each treatment group and the difference versus placebo, with corresponding 95% confidence limits. A p-value, corresponding to a 2-sided test, will be presented to compare the benralizumab and placebo treatment groups.

For participants who receive medications deemed to be intercurrent events as per Section 3.1.5.2 or need to increase their OCS dose as a rescue therapy will be imputed using a last observation carried forward (LOCF) method, as detailed in Section 3.3.3.

For participants who withdraw from the study without having rescue or restricted medications, missing BPDAI activity score at Week 36 will be imputed using multiple imputation based on outcome of participants who don't have rescue/restricted medication use. The same approach will be applied to those participants with missing data at Week 36 for any other reason. Distributions will be modelled considering treatment group, time of BP diagnosis and values of BPDAI activity score observed at baseline and prior to the withdrawal/occurrence of intercurrent events.

Specifically, for a participant with missing observations after visit t, imputed observations value at visit t+n will be based on a model estimated from all participants who have an outcome observed at visit t+n conditioning on that participant's outcomes up to and including visit t.

Then, the overall estimate of treatment effect is obtained by using Rubin's rules [Rubin 1987] to combine the imputed outcomes using the ANCOVA model above specified.

In summary, the pattern mixture models will be implemented by multiple imputation via the following steps;

Step 1: Non-monotone missing data (usually relatively infrequent) will be imputed using the MCMC statement of SAS PROC MI under the MAR assumption. The following variables will be included in the imputation model: treatment group, time of BP diagnosis and BPDAI activity score from baseline to Week 36. Note: Dummy variables will be used for treatment and time of BP diagnosis. This will create a monotone missing pattern in the dataset within each treatment group. This step will create 100 imputed datasets.

Step 2: For participants who withdraw from the study without having rescue or restricted medications the values for each pattern will be imputed via the sequential regression method, using the MONOTONE REG statement of SAS PROC MI. The imputation model will be estimated from the observed data in participants without intercurrent events or need to increase their OCS dose as a rescue therapy and it will be used to impute missing observations in both treatment groups using the MNAR statement of SAS PROC MI. The following variables will be included in the imputation model: treatment group, time of BP diagnosis and BPDAI activity score from baseline to Week 36.

Step 3: For participants who receive medications deemed to be intercurrent events or need to increase their OCS dose as a rescue therapy the missing values for each pattern will be replaced by the last observation carried forward.

Step 4: The imputed data from Step 1, Step 2 and Step 3, along with observed data will be combined to create 1000 complete dataset which will then be analysed using the main ANCOVA statistical model.

Step 5: The estimates from multiple imputed datasets will be combined using Rubin's combination rules for statistical inference (using PROC MI ANALYZE).

The seed used will be 88281 for both steps 1 and 2.

To support this analysis, observed means and mean changes from baseline will be summarized at each scheduled visit up to Week 36 using descriptive statistics by treatment group. The same descriptive approach will be used to summarize results for the total BPDAI damage score.

To explore the uniformity of the detected overall treatment effect, the model might also be fitted considering baseline variables and their interaction with treatment as additional covariates.

In addition, the cumulative distribution function of absolute changes from baseline in BPDAI activity score at Week 36 will also be plotted, to allow the treatment difference at various response thresholds to be explored. Responder analyses using minimal clinically important differences presented in the literature (Wijayanti et al 2017) may be presented, if appropriate.

4.2.7.4 Change from baseline in BPDAI-Pruritus score at Week 36

The change from baseline in the BPDAI-Pruritus score at Week 36 will be analysed by a ANCOVA model. More details on derivation rules and intercurrent events are given in Section 3.3.4.

The statistical analysis described in Section 4.2.7.3 for BPDAI activity score will be carried out for this endpoint using 88381 as the seed. In the MI models however, baseline disease severity will be added as categorical covariate.

To support this analysis, observed means and mean changes from baseline will be summarized at each scheduled visit up to Week 36 using descriptive statistics by treatment group.

In addition, the cumulative distribution function of absolute changes from baseline in BPDAI activity score at Week 36 will also be plotted, to allow the treatment difference at various response thresholds to be explored.

4.2.7.5 Sensitivity analyses

Sensitivity analyses for the primary and key secondary endpoints will be used to explore the robustness of any treatment effect, if the amount of missing data warrants further investigation.

Alternative estimand approaches for handling intercurrent events will be explored including treatment policy approaches, or alternative imputation techniques for handling of treatment failure intercurrent events for continuous endpoints, if relevant.

For missing data due to study withdrawal or incomplete data for other reasons, different missing data mechanisms (eg, values missing at random [MAR] and values missing not at random [MNAR]) will be considered to assess the robustness of the treatment effect for these endpoints. The use of multiple imputation approaches, including dropout reason-based multiple imputation will be considered.

To examine the impact of missing data, including the impact of non-responder/not relapse free imputations due to study withdrawal, for the primary and first key secondary endpoint respectively, tipping point analyses may be performed. These analyses will systematically vary the assumptions about outcomes among the subsets of participants on the treatment arms who withdraw from the study early or have missing data at Week 36 for any other reason. Tipping point analyses are intended to identify the point at which the results would tip from statistically significant to not statistically significant. Thus, the tipping point analyses will only be performed if an endpoint achieves a nominally statistically significant result (ie, nominal p-value < 0.05).

Details using multiple imputation techniques are specified in Appendix 8.1.

To explore the primary analysis estimand approach, a sensitivity analysis may be performed whereby an increase in OCS dose after previously achieving disease control will be considered an intercurrent event, and the participant being considered a non-responder in the analysis.

4.2.8 Secondary efficacy outcome variables

4.2.8.1 Proportion of participants in sustained complete/partial remission on minimal OCS/off OCS for at least 2 months at Week 36

More details on derivation rules and intercurrent events are given in Section 3.4.1.

The statistical methodology described for the primary analysis will be carried out for this endpoint.

The p-value will be nominal, as secondary objectives are not controlled for multiplicity and do not appear in the testing hierarchy.

Summaries of the different combinations of complete and partial remission, each on minimal OCS or off OCS, will also be produced to understand the distribution of these clinical benefit categories at visits throughout the study up to Week 36.

4.2.8.2 Cumulative time (weeks) in complete remission off OCS from baseline to Week 36

Cumulative time in complete remission while off OCS will be derived for each participant over the initial 36-week treatment period. More details on derivation rules and intercurrent events are given in Section 3.4.2.

Cumulative time in complete remission while off OCS will be categorized as per Section 3.4.2.

An ordered logistics regression model will be fitted to these categories, assuming proportional odds. The model will include cumulative time as the dependent variable and treatment, baseline disease severity and time of BP diagnosis as categorical covariates. The model will be used to estimate an odds ratio, with 95% CIs.

The p-value will be nominal, as secondary objectives are not controlled for multiplicity and do not appear in the testing hierarchy.

Cumulative time (weeks) in complete or partial remission off OCS from baseline to Week 36 will be also analysed in an exploratory way using same methodology.

4.2.8.3 Proportion of participants off OCS by Week 36

Participants will be defined as responders if they are off OCS by Week 36. Otherwise, participants will be classified as non-responders. More details on derivation rules and intercurrent events are given in Section 3.4.3.

The statistical methodology described for the primary analysis will be carried out for this endpoint.

The p-value will be nominal, as secondary objectives are not controlled for multiplicity and do not appear in the testing hierarchy.

To support this analysis, the observed proportion of responders will be summarized using descriptive statistics by treatment group and each scheduled visit up to Week 36.

4.2.8.4 IGA score at Week 36 and change from baseline in IGA score at Week 36

IGA score achieved (clear, almost clear, mild, moderate, severe) will be summarized using descriptive statistics by treatment group at each scheduled visit up to Week 36.

A responder analysis will be performed where participants will be classified as responders if they achieve IGA 0/1 (ie, clear or almost clear) at Week 36. Otherwise, participants will be classified as non-responders. The statistical methodology described for the primary analysis will be carried out for this endpoint (the p-value will be nominal, as secondary objectives are not controlled for multiplicity and do not appear in the testing hierarchy).

Number of points change from baseline in IGA score will be summarized, and a shift table comparing IGA score at baseline against IGA score achieved at Week 36 (and other visits throughout the study, as relevant) will be produced.

4.2.8.5 Proportion of participants who remain relapse-free up to Week 16

Referring to the statistical analysis for the proportion of participants who remain relapse-free up to Week 36, the proportion of participants who remain relapse-free up to Week 16 will be analyzed as a binary endpoint with a logistic regression model similar to the one applied to the analysis at Week 36.

4.2.8.6 Proportion of participants with any clinical benefit at Week 16

To assess signs of clinical efficacy during the first 16 weeks of the study, the observed proportion of participants showing partial remission or complete remission at Week 16 will be summarized using descriptive statistics by treatment group. Results will be also presented within OCS usage as reported in Section 3.4.6.

If a consistent message can be drawn across these categories and treatment groups, a supportive analysis will be carried out, whereby a participant will be classified as a responder if they show any clinical benefit (ie, partial remission or complete remission) during taper, with no steroids or with minimal steroid use at Week 16. Otherwise, a participant will be classified as a non-responder.

This binary variable will then be analysed using a logistic regression model and applying the same statistical methodology as described for the primary analysis in Section 4.2.6.1.

The p-value will be nominal, as secondary objectives are not controlled for multiplicity and do not appear in the testing hierarchy.

4.2.8.7 Time to disease control, OCS dose at disease control and time to the end of the consolidation phase

Time to disease control, OCS dose at disease control and time to the end of the consolidation phase will be summarized using descriptive statistics by treatment group.

Statistical analyses and Kaplan Meier plots may be performed if appropriate.

4.2.8.8 Change from baseline in BPDAI activity score at Week 16

Referring to the statistical analysis for change from baseline in BPDAI activity score at Week 36, the same analysis method will be used for the change from baseline at Week 16 endpoint.

The p-value at Week 16 will be nominal, as secondary objectives are not controlled for multiplicity and do not appear in the testing hierarchy.

4.2.8.9 Change from baseline in BPDAI-Pruritus score at Week 16

Referring to the statistical analysis for change from baseline in BPDAI-Pruritus score at Week 36, the same analysis method will be used to provide an estimate at Week 16.

The p-value at Week 16 will be nominal, as secondary objectives are not controlled for multiplicity and do not appear in the testing hierarchy.

4.2.8.10 Serum benralizumab concentration

The PK analyses will be performed at or under the guidance of AstraZeneca Research and Development. Benralizumab serum concentrations will be summarized using descriptive statistics at each scheduled visit for participants in the PK analysis set.

4.2.8.11 Anti-drug antibodies (ADA)

Anti-drug antibodies to benralizumab will be summarized using descriptive statistics by treatment group and each visit specified in the CSP Table 2. The ADA titres-time profiles of benralizumab may 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.

4.2.9 Safety Analyses

Safety analyses will be performed using the safety analysis set.

Summaries will be presented for the double-blind period and the OLE period. A small number of key tables might also be presented for data in the overall study period including only participants who received benralizumab during the double-blind period.

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. More details are reported in Section 3.1.1.

Safety data will be presented using descriptive statistics. No safety data will be imputed.

4.2.9.1 Adverse Events

Treatment Emergent Adverse Events (TEAEs) will be summarized separately for the on-treatment, on-study, and post-treatment periods, as defined in Section 3.5.1. The primary period of interest for safety summaries will be the on-treatment period, unless differently specified.

Adverse events in the pre-treatment period (with start date prior to the first dose of IP) will be listed for the double-blind treatment period at primary database lock only.

Non-serious AEs with an onset date after the on-treatment period will be listed but excluded from the summaries.

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

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).

The rate of TEAEs, SAEs, AEs with outcome of death and DAEs per person-years at risk will be calculated for the on-treatment period as (number of participants reporting the TEAE) / (total time with participants at risk of TEAE). For SAEs and AEs with outcome of death the rate per person-years at risk will be calculated also for the on-study period as (number of participants reporting the TEAE) / (total time with participants at risk of TEAE). The total period at risk for each participant will be the duration of the on-treatment and on-study periods as defined in Section 3.5.1. Rates will be expressed in terms of events per 100 participant-years.

The post-treatment AEs will be listed.

Treatment-emergent AEs, AEs with outcome of death, SAEs and DAEs will be presented for each treatment group for the on-treatment and on study periods by SOC and PT, including the number and percentage of participants reporting at least one event (ie, multiple occurrences of an AE for a participant will only be counted once), number of events and exposure-adjusted rates (where applicable).

A summary of the most common (frequency of >5%) TEAEs will be presented by PT.

Treatment-emergent AEs and SAEs causing discontinuation of the study treatment and SAEs causing discontinuation from the study will also be summarized.

Treatment-emergent AEs, SAEs and DAEs will be summarized 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 worst-case Investigator's causality assessment (the order being not related and related) will be taken as Investigator's causality assessment and the highest recorded maximum intensity (the order being mild, moderate, and severe) will be taken as intensity.

Other significant TEAEs will include but may not be limited to injection site reactions and hypersensitivity events. Treatment-emergent AEs of injection site reactions (high level term of injection site reactions) and hypersensitivity (standardized MedDRA query of hypersensitivity [narrow]) will be summarized by preferred term. The injection site reactions will be summarized 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, SAEs, and DAEs.

Separate listings of participants with AEs, AEs with outcome of death, SAEs, or DAEs will be presented.

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

4.2.9.2 Clinical laboratory safety assessments

All protocol-specified continuous laboratory parameters will be summarized descriptively by absolute value at each visit by treatment group, together with the corresponding changes from baseline, separately for the double-blind period and the OLE period. All parameters will be summarized in SI units, except for blood eosinophil counts which will be summarized in both SI and conventional units. Results reported by the central laboratory in conventional units will be converted to SI units for reporting.

For continuous data, the summary statistics presented will be the n, mean, standard deviation, median, minimum value and maximum value.

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). Total bilirubin 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.

4.2.9.3 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 and separately for the double-blind period and the OLE period. Summary statistics for continuous variables will include n, mean, standard deviation, median, minimum value and maximum value.

All vital sign and physical examination data will be listed.

4.2.9.4 ECG

The Investigator's overall evaluation of ECG (normal or abnormal) will be listed for all participants, detailing whether any abnormalities were clinically significant or not.

4.2.10 Tertiary/exploratory outcome variables

4.2.10.1 Peak Pruritus Numeric Rating Scale

The Peak Pruritus NRS will be summarized by descriptive statistics for observed means and mean changes from disease control of the weekly mean scores and the observed proportion of responders as per Section 3.6.1 by treatment group and week.

Peak Pruritus NRS assessment compliance will be also evaluated reporting descriptive summaries for each week.

4.2.10.2 Autoimmune Bullous Disease Quality of Life Questionnaire (AB-QoL)

The AB-QoL total score will be summarized presenting summary statistics for observed means and mean changes from baseline by treatment group and scheduled visit.

To support these summaries, an MMRM model may be fitted to this outcome variable if deemed appropriate, using the same statistical methodology detailed in Section 4.2.7.3 for the BPDAI activity score.

4.2.10.3 Serum and plasma biomarkers

Serum and plasma samples will be collected according to the CSP Table 2 to evaluate the pharmacology of benralizumab as well as biomarkers of eosinophil recruitment, activation, and survival (eg, absolute eosinophil count, IL-5, EDN, eotaxin). Summaries and plots of absolute levels and changes from baseline will be produced as appropriate.

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

Scatterplots of baseline biomarker levels versus efficacy measures at Week 36, with locally weighted smoothing (LOESS) lines by treatment as appropriate

Subgroup analyses of proportion of participants in complete remission while off OCS for ≥ 2 months at Week 36 (and other efficacy endpoints if appropriate) by subgroups defined by baseline biomarker levels. Groupings of baseline biomarker levels may include splitting biomarkers at the median, or at quartiles if n large enough to warrant 4 groups, or other clinically relevant thresholds of interest to be pre-defined prior to unblinding. Results would be presented in forest plots of treatment effect in each level of the biomarker subgroups.

Additional markers of cellular inflammation and/or activation may be assessed including, but not limited to, those associated with humoral autoimmunity (total serum IgE, IgG directed against auto-antigens BP-180 and/or BP-230), T cell subsets (eg, IL-13, IFN-gamma, IL-17), eosinophil granule proteins (eg, MBP, eosinophil peroxidase), epithelial cell damage (eg, TSLP, IL-25, IL-33), itch (IL-31), and pain (substance P). The results from such studies will not be reported in the CSR but in separate reports or publications as appropriate. The corresponding analyses are not covered in this SAP.

CCI

[redacted]

CCI

[redacted]

4.2.11 OLE

The effect of benralizumab will be explored in participants in the OLE period by assessing endpoints as per Section 3.6.9.

These endpoints will be summarized using descriptive statistics for the OLE analysis set. Summaries will be presented for the overall population, and by prior randomized treatment (benralizumab or placebo).

In addition, selected efficacy and safety data may be integrated for those participants randomized to benralizumab, to describe efficacy and safety data over the entire study follow-up period, and on those participants who switch from placebo to benralizumab to describe efficacy and safety data over the benralizumab treatment period.

CCI

[redacted]

4.2.13 Relationship between IGA score and other measures of BP disease severity

The relationship between the IGA score and other measures of BP disease severity / activity will be explored via scatterplots and tabulations of endpoints. The types of outputs that may be produced include the following:

- Scatterplots of IGA score versus measures of disease activity such as physician global assessment and BPDAI score, at baseline, weeks 8, 16 and 36

- Cross tabulations of IGA score versus BP disease states at weeks 8, 16 and 36
- Cross tabulations of IGA score versus BPDAI severity (using BPDAI threshold of <24, ≥24 to <56 and ≥56 to as indicators of mild, moderate and severe disease) at baseline and weeks 8, 16 and 36
- Summaries of change from baseline in IGA score by PGI-C groups where PGI-C is categorised into groups of worsening, no change and improved

5 INTERIM ANALYSES

A non-binding futility analysis will be carried out by the IDMC when approximately CCI randomized participants (ie, when approximately CCI of the total amount of statistical information is available) have had the opportunity to complete at least 36 weeks of follow-up. Recruitment will remain ongoing while the futility analysis is performed. In the futility analysis, a predictive power-based guideline will be used for the primary endpoint (remission whilst off OCS for at least 2 months by Week 36) such that, given the data accumulated to date, if there is low chance (CCI of achieving statistical significance at the final analysis if the study were to continue, then a stop decision for futility should be considered. CCI [redacted]. If a decision is made that the study should continue, the primary analysis will be performed when all randomized participants have had the opportunity to complete at least 36 weeks of follow-up. The overall power is approximately CCI after accounting for the inclusion of the futility analysis.

6 CHANGES OF ANALYSIS FROM PROTOCOL

Section 2.1.5 OLE analysis set has been added to identify all participants who enter into the OLE period.

In Section 3.1.5.2 further detail around which restricted medications are considered intercurrent events has been provided.

In Section 3.2 the missing data handling strategy for the primary endpoint has been detailed to cover occurrences of missing disease state assessments between two non-missing assessments. This is to integrate and not to change the approach specified in the CSP that states: "Any participant with missing data at a specific time point will be considered as a non-responder at that time point." This statement in the CSP is meant to be applied to any missing disease state assessment after the last non-missing assessment.

In Section 4.2.7.3 the ANCOVA model specified for the analysis of the change from baseline at Week 36 in BPDAI activity score has been changed removing baseline disease severity as categorical covariate. This categorical variable is derived based on the baseline BPDAI activity score which is already included in the model as continuous covariate.

7 REFERENCES

Bartlett 2018
Bartlett JW. Covariate adjustment and estimation of mean response in randomized trials. Pharmaceutical Statistics. 2018; 17:648–666.

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

Guy 1976
Guy W. The Clinical Global Impression Scale. In: ECDEU Assessment Manual for Psychopharmacology Revised. Rockville, MD. US Dept. of Health, Education & Welfare, ADAMHA, NIMH Psychopharmacology Research Branch, 1976:218-22.

Jenerowicz et al 2007
Jenerowicz D, Czarnecka-Operacz M, Silny W. Peripheral blood eosinophilia in atopic dermatitis. Acta Dermatovenerol Alp Pannonica Adriat. 2007;16(2):47-52.

Murrell et al 2012
Murrell DF, Daniel BS, Joly P, Borradori L, Amagai M, Hashimoto T, et al. Definitions and outcome measures for bullous pemphigoid: recommendations by an international panel of experts. J Am Acad Dermatol 2012 Mar;66(3):479-85.

Rosenbach et al 2009
Rosenbach M, Murrell DF, Bystryn JC, Dulay S, Dick S, Fakharzadeh S, et al. Reliability and convergent validity of two outcome instruments for pemphigus. J Invest Dermatol 2009 Oct;129(10):2404-10.

Simon et al 2020
Simon D, Yousefi S, Cazzaniga S, Bürgler C, Radonjic S, Houriet C, et al. Mepolizumab failed to affect bullous pemphigoid: a randomized, placebo-controlled, double-blind phase 2 pilot study. Allergy 2020 Mar;75(3):669-72.

Wijayanti et al 2017
Wijayanti A, Zhao CY, Boettiger D, Chiang YZ, Ishii N, Hashimoto T, et al. The reliability, validity and responsiveness of two disease scores (BPDAI and ABSIS) for bullous pemphigoid: which one to use? Acta Derm Venereol 2017;97:24-31.

Yosipovitch et al 2019
Yosipovitch G, Reaney M, Mastey V, et al. Peak Pruritus Numerical Rating Scale: psychometric validation and responder definition for assessing itch in moderate-to-severe atopic dermatitis. *Br J Dermatol*. 2019;181(4):761-769. doi:10.1111/bjd.17744

8 APPENDIX

8.1 Accounting for missing data

Missing data can occur due to intercurrent events, like usage of restricted medications and OCS dose increased, or early withdrawn. The amount of missing data is minimized in this study as participants are encouraged to remain in the study after premature discontinuation of IP and complete visits according to the protocol.

This section 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 9). Dependent on these outputs additional exploratory analyses may be produced as deemed necessary to further understand the pattern of missing data.

8.1.2 Primary analysis and key secondary main analyses

The primary analysis for the primary endpoint of proportion of participants who are in complete remission while off OCS for ≥ 2 months at Week 36 will be performed under a conservative assumption. Participants who experience any of the intercurrent events, including early withdrawn, are considered as non-responders. The analysis will include all the data up to Week 36, regardless of if they discontinue from randomised treatment.

The main analysis for the first key secondary endpoint of proportion of participants who remain relapse-free up to Week 36 will be analysed with a statistical model similar to the one described for the primary endpoint, under the same conservative assumption.

The analysis for the key secondary continuous endpoint of cumulative OCS exposure (mg/kg) from baseline to Week 36 and from baseline to Week 16 is performed imputing as missing data after the occurrence of intercurrent events, such as usage of restricted medications and OCS dose increased, or early withdrawal. The analysis uses the MMRM method and assumes that missing data is MAR and is a direct likelihood (DL) approach.

The analysis for the remaining key secondary endpoints, i.e. change from baseline in BPDAI activity score at Week 36 and change from baseline in BPDAI-Pruritus score at Week 36, will be performed with a composite strategy: participants who need to increase their OCS dose as a

rescue therapy to treat their BP after having previously achieved disease control will have their last BPDAI activity score observed carried forward. For participants who withdraw from the study without having rescue or restricted medications, missing BPDAI activity score at Week 36 will be imputed using multiple imputation based on participants who don't have rescue/restricted medication use. The same approach will be applied to those participants with missing data at Week 36 for any other reason.

8.1.3 Multiple Imputation Analyses

To examine the sensitivity of the results of the primary analysis to departures from the underlying assumptions, additional analyses may be performed using controlled multiple imputation method introduced in Keene et al 2014, Wan et al 2015, Gottlow et al 2015 and further developed by the European Medicines Agency (Guideline on Missing Data in Confirmatory Clinical) and AstraZeneca (AZ guidance), which allows for different underlying assumptions to be used.

As with the primary analysis, the sensitivity analyses include all data until participants withdraw from the study regardless of if they discontinue from randomized treatment.

Participants who use restricted medications and early withdrawals from the study will not be imputed as non-responders as for the primary analysis but will be considered as missing. Participants who use restricted medications will be considered as study withdrawers for severe non-compliance to protocol.

An underlying Bernoulli distribution for the status of response (responder or non-responder) is assumed and post study withdrawal events status of response will be imputed conditional upon the treatment group and the observed covariates. This allows various assumptions about the missing data to be analysed by modifying the post-withdrawal model assumption.

The method involves first fitting the logistic regression model to the observed data and then imputing post-withdrawal status of response by calculating the conditional probability of response relating the observed covariates based on various assumptions.

$$logit\ (P(Y_{ij} = 1)) = j + Covariates_i \qquad (1)$$

Here i denotes the participant identifier and j denotes the treatment group (benralizumab or placebo). $Covariate_i$ denotes the baseline disease severity and the time of BP diagnosis of the ith participant. Furthermore,

$$p_{ij} = \frac{exp(j + Covariates_i)}{1 + exp(j * Covariates_i)} \qquad (2)$$

The imputed status of response that would have been seen is then combined with the observed responses and data will be analysed using the primary analysis methodology. This analysis is repeated multiple times and the results combined using Rubin's rules [Rubin 1987, Bernard et al 1999].

The following default assumptions that will be used to impute the missing data who withdraw early from the study are as follows:

(a) MAR: Missing response status in each treatment group are imputed assuming the expected response rate within that treatment group.
(b) Dropout Reason-based Multiple Imputation (DRMI): Missing status will be imputed differently depending on the reason for dropout; statuses for participants in the benralizumab treatment group who dropped out for a treatment related reason are imputed based on the expected response rate in the placebo treatment group, whereas the remaining participants who have dropped out are imputed assuming MAR. Treatment related reasons include (1) AEs, (2) Death and (3) Development of study specified reasons to stop active treatments, and (4) severe non-compliance of protocol.

Some reasons for withdrawal are clearer to determine as treatment related (AEs, death, development of study-specific discontinuation criteria) or non-treatment related (participant lost to follow up, eligibility criteria not fulfilled). Other reasons are less clear such as severe non-compliance of protocol, participant decision and 'Other'; a review of each participant who withdraws from the study will therefore be carried out prior to unblinding the study. The review will include assessment of the reason for discontinuation of randomised treatment for those participants who discontinued randomised treatment and then withdrew from the study and also free text for when the reason for withdrawal or discontinuation of randomised treatment is recorded as participant decision or other. Based on this review the default assumptions for DRMI as described in b) and Table 9 may be changed. A list of these participants and the assumptions made under DRMI will be documented prior to unblinding of the study.

A summary of reasons for participants withdrawing from the study and the corresponding treatment group used to calculate the imputation response rate under MAR and DRMI is given in Table 9.

Table 9 Summary of reasons for participants withdrawing from the study and the corresponding treatment group used to calculate the imputation response rate 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
Other	Benralizumab	Based on review prior to study unblinding

Note all participants on response rate in the placebo treatment group are imputed using the placebo treatment group rate

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

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

8.1.4 Tipping point analyses

To examine the impact of missing data, including the impact of non-responder/not relapse free imputations due to study withdrawal, for the primary and first key secondary endpoint respectively, tipping point analyses may be performed. These analyses will systematically vary the assumptions about outcomes among the subsets of participants on the treatment arms who withdraw from the study early or have missing data at Week 36 for any other reason. Tipping point analyses are intended to identify the point at which the results would tip from statistically significant to not statistically significant. Thus, the tipping point analyses will only be performed if an endpoint achieves a nominally statistically significant result (ie, nominal p-value < 0.05).

The analyses will be performed following below steps:

- The proportions of participants non-responder/not relapse free will be analysed using a Pearson's chi-squared test, thus the stratification factors that are used in the main analysis will be disregarded. Since the strata will be balanced in respect of treatment assignment by virtue of the stratified randomisation scheme used, the only impact of this simplification should be that the inferences from the unstratified analysis will be somewhat conservative.

- For the primary analysis, participants who withdraw from the study early or have missing data at Week 36 for any other reason are by definition imputed as non-responders. For this sensitivity analysis, these participants will be altered from non-responder to responder in an iterative manner.

- At each step of the analysis one of these participants switches from not being responder to being responder, and the Pearson's chi-squared test is re-run. The results (statistical significance) are presented in a grid where the x-axis and the y-axis represent the number of participants assumed to be responder for placebo and benralizumab, respectively. The region where the conclusion changes, will be considered as the tipping point.

- The grid will be divided in 3 regions, limited in the top by the expected number of participants with missing values in the benralizumab treatment group which would have been responder if observed at Week 36 (based on the proportion of responders in participants who did not withdraw from the study early or have missing data at Week 36 for any other reason in the benralizumab treatment group):

    - Likely – Its right limit is the expected number of participants with missing values in the placebo arm which would have been responder if observed at Week 36 (based on the proportion of responders in participants who did not withdraw from the study early or have missing data at Week 36 for any other reason in the placebo treatment group);

    - Uncertain – its right limit is the expected number of participants with missing values in the placebo treatment group which would have been responder if observed at Week 36 (based on the proportion of responders in participants who did not withdraw from the study early or have missing data at Week 36 for any other reason in the benralizumab treatment group);

    - Unlikely – It is the region to the right of the uncertain region.

8.1.5 Missing Not At Random for MMRM

In addition, an imputation sensitivity analysis will test the assumption of MAR that is made by the MMRM analysis foreseen for the analysis of the cumulative OCS exposure (mg/kg) from baseline to Week 36 and from baseline to Week 16.

For this purpose, a model that assumes MNAR will be used, whereby assumptions for the missing data "stress test" the MAR assumptions of the primary analysis, by positing outcomes that, while clinically plausible, are likely to be worse for the benralizumab group than the outcomes assumed by MAR. A plausible stress test could assume that the outcome for the benralizumab treatment is somewhat worse than would be predicted by MMRM from the study data. Such a stress test would appropriately disadvantage the estimate of treatment effect, compared to the estimate of the main analysis.

The proposed MNAR sensitivity analysis assumes that the trajectory after the intercurrent event from the benralizumab group, regardless of the type of event, can be modelled by that of placebo participants. This assumption will tend to result in smaller estimates of difference between the treatment groups, compared to those under the MAR assumption, and as such constitutes a reasonable stress test of the MAR assumption. Note that this approach constitutes a pattern mixture model as defined by Molenberghs 2007. The patterns here are defined by the time of the intercurrent event, with the imputed values for both treatment groups modelled by observed outcomes from the placebo group only. Specifically, for a participant with missing observations after visit t, imputed value for the change from baseline at visit t+n will be based on a model estimated from all in the control group who have an outcome observed at visit t+n conditioning on that participant's outcomes up to and including visit t. Then, the overall estimate of treatment effect is obtained by using Rubin's rules [Rubin 1987] to combine the imputed outcomes. Note that in this analysis, placebo missing are imputed assuming MAR and here follows the pattern of observed placebo observations, while missing observations for the benralizumab group are assumed MNAR. This strategy for imputation is commonly referred to as the Copy-Reference approach. In summary, the pattern mixture models will be implemented by multiple imputation via the following steps;

Step 1: Non-monotone missing data (usually relatively infrequent) will be imputed using the MCMC statement of SAS PROC MI under the MAR assumption. The following variables will be included in the imputation model: treatment, baseline disease severity, time of BP diagnosis and cumulative OCS exposures from baseline to Week 36 by 4-weekly intervals. Note: Dummy variables will be used for treatment, baseline disease severity and time of BP diagnosis. This will create a monotone missing pattern in the dataset within each treatment group. This step will create 1000 imputed datasets.

Step 2: Then the values for each pattern will be imputed via the sequential regression method, using SAS PROC MI statement MONOTONE REG. The imputation model will be estimated

from the observed data in the placebo group and it will be used to impute missing observations in both treatment groups using the MNAR statement of SAS PROC MI. The following variables will be included in the imputation model: baseline disease severity, time of BP diagnosis and cumulative OCS exposures from baseline to Week 36 by 4-weekly intervals.

Step 3: The imputed data from Step 1 and Step 2, along with observed data will be analyzed using the same MMRM model as for the main analysis. The estimates from multiple imputed datasets will be combined using Rubin's combination rules for statistical inference (using PROC MI ANALYZE).

The seed used will be 88281 for both steps 1 and 2.

8.1.6 On-treatment analyses

In addition to the primary and sensitivity analyses described previously, two alternative estimands may be estimated for primary and key secondary endpoints, using only the data from participants while on initial randomised treatment:

- Primary on-treatment estimand: This will be estimated using the primary method but including only data from participants whilst being on initial randomized treatment. Participants who drop from randomized treatment are imputed as non-responders.

- Efficacy estimand: This will be estimated using the MAR method above described but including only data from participants whilst being on initial randomized treatment.

- Effectiveness estimand: Assumed loss of effect post discontinuation of benralizumab: This will be estimated using the DRMI controlled imputation approaches including only data from participants whilst on treatment.

Therefore, the primary analyses and sensitivity analyses will be repeated including only data from participants whilst being on initial randomised treatment, i.e., excluding data once participants discontinue from randomised treatment (including on-treatment period to IPD visit or last contact if the participant is lost to follow-up).

A summary of reasons for participants withdrawing from the benralizumab treatment group and the corresponding treatment group used to calculate the imputed response probability under MAR and DRMI are given in Table 10. As for participants who withdraw from the study, a review of each participant who discontinued randomised treatment will be carried out prior to unblinding the study where the default assumptions for DRMI as described in Table 9 may be changed. Again, a list of these participants and the assumptions made under DRMI will be documented prior to unblinding of the study.

Table 10 Summary of reasons for participants withdrawing from the benralizumab treatment and the corresponding treatment group used to calculate the imputation response rate 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
Participant lost to follow up	Benralizumab	Benralizumab
Participant decision	Benralizumab	Based on review prior to study unblinding
Other	Benralizumab	Based on review prior to study unblinding

Note all participants on response rate in the placebo treatment group are imputed using the placebo treatment group rate

Using on treatment data is easier to interpret as it is not impacted by any subsequent pattern of alternative treatments once participants discontinue from randomised treatment. Sensitivity analyses under DRMI allow for alternative assumptions to be made based on reasons for discontinuation.

8.1.7 Treatment policy estimand

A pure treatment policy strategy may also be applied in another sensitivity analysis performed for primary and key secondary endpoints without considering restricted medications/ OCS dose increase as a rescue therapy, as applicable based on the endpoint, as intercurrent events.

8.1.8 Overall summary of sensitivity analyses to account for missing data

A summary of the different sensitivity analyses for the primary endpoint to be carried out under different estimands and assumptions are described in Table 11.

Sensitivity analyses and multiple imputation analyses might be performed on key secondary outputs if deemed appropriate. Full details of the sensitivity analyses will be finalized as an update to the Statistical Analyses Plan (SAP) before primary database lock.

Table 11 Summary of the analyses to be carried out under different estimands and assumptions

	On-treatment + post-discontinuation of randomised treatment				On-treatment		
Assumption	No missing data	MAR		MNAR	No missing data	MAR	MNAR
Estimand	Primary	Treatment policy	MAR MI	DRMI	On-treatment	MAR MI	DRMI
Responder/Non-responder status for imputation in benralizumab treatment group[a]	Participants who receive restricted medications and participants who withdraw from the study are imputed as non-responders	Participants who withdraw from the study are imputed as non-responders.	Benralizumab rate for all reasons for withdrawal assumed to be on benralizumab	Placebo rate for AEs, death, development of study specified reasons to stop active treatments and severe non-compliance to protocol assumed to be on, otherwise benralizumab or based on review prior to study unblinding	Participants who receive restricted medications and participants who withdraw from the study or the randomized treatment are imputed as non-responders	Benralizumab rate for all reasons for withdrawal assumed to be on benralizumab	Placebo rate for AEs, death, development of study specified reasons to stop active treatments and severe non-compliance to protocol assumed to be on, otherwise benralizumab or based on review prior to study unblinding
Default definition for p_{ij} based on formula (2)[b]			Formula (2) for all treatment groups. j = B or P.	Formula (2) using j = P for reasons above. Otherwise, j = B or P (same as MAR)		Formula (2) for all treatment groups. j = B or P.	Formula (2) using j = P for reasons above. Otherwise, j = B or P (same as MAR)

[a] All participants in the placebo treatment group are imputed using the placebo treatment group rate (i.e., j = P)
[b] Note can be over written by review prior to study unblinding
B – Benralizumab; P – Placebo

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.

8.1.9 References

AZ guidance
AZ guidance (clinical OPI): Guidance on Minimizing the Loss of Patient Data in AstraZeneca Clinical Trials, ed 2.0. (LDMS_001_00102309)

Bernard et al 1999
Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. Biometrika 1999; 86:948-955.1

Gottlow et al 2015
Gottlow M, Hollis S, Wan R, Hirsch I, Darilay A, Weissfeld L, France L. A Simulation study of a controlled imputation approach for analyzing missing data in recurrent events due to early discontinuations. PSI Annual Conference 2015.

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

Keene et al 2014
Keene ON, Roger JH, Hartley BF, Kenward MG. Missing data sensitivity analysis for recurrent event data using controlled imputation. Pharmaceutical Statistics 2014, 13 258–264.

Molenberghs et al 2007
Molenberghs and M.G Kenward. Missing Data in Clinical Studies, p221. John Wiley and Sons, 2007.

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

Wan et al 2015
Wan R, Hirsch I, Gottlow M, Hollis S, Darilay A, Weissfeld L, France L. Controlled imputation approach for analyzing missing data in recurrent events due to early discontinuations. DIA/FDA Statistics Forum 2015.

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 post-baseline result and either of the following statements holds:

Baseline is ADA negative and at least one post-baseline assessment is ADA positive. This is called treatment-induced ADA positive.

Baseline is ADA positive, and the baseline titre is boosted by greater than the variability of the assay (ie, ≥ 4-fold increase) at ≥1 post-baseline timepoint. This is called treatment-boosted ADA positive.

Participants who are persistently ADA positive, which is defined as ADA negative at baseline and having at least 2 post-baseline ADA positive measurements with ≥ 16 weeks between first and last positive, or an ADA positive result at the last available post baseline assessment.

Participants who are transiently ADA positive, defined as ADA negative at baseline and at least one post- baseline ADA positive measurement and not fulfilling the conditions for persistently positive.

Participants who are ADA positive with maximum titre > median of maximum titres. The median of maximum titres will be calculated based on the maximum titre of each ADA positive participant within each treatment group (including both baseline and post-baseline measurements).

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

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

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

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

8.2.2 ADA and safety

Adverse events and SAEs during the study (separately for on-treatment and on-study periods) will be summarized by ADA status (treatment-emergent ADA positive, ADA negative), if enough data are available. The on-treatment and on-study periods are as defined in Section

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

8.2.4 ADA and PK

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

8.3 Partial dates for adverse events and prior/concomitant medications

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

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:

The date of the first study treatment, if the start year is the same as the year of the first study treatment

January 1 of the year of occurrence.

If the end date is missing both the day and month, the date will be set to:

Last study treatment date if the year of occurrence is the same as the last study treatment date.

Date of death if the participant died in the same year.

December 31 of the year of occurrence.

If the start date is null, the date will be set to:

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.

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	D325AC00002
Document Name	Statistical Analysis Plan
Document Version	2.0
Document Date	11 February 2021

[redacted]

PPD

PPD

PPD

Signature Page – Wet Ink

Study Biostatistician: PPD

Principal Biostatistician

PPD

PPD

Signature Page – Wet Ink

Approved by:

PPD

AstraZeneca

PPD

Signature **Date**