–

STATISTICAL ANALYSIS PLAN	
Study Code	D3258C00001
Edition Number	1.0
Date	2-Oct-2023

A Multi-center, Randomized, Double-blind, Parallel-group, Placebo-controlled 3-Part Phase 3 Study to Demonstrate the Efficacy and Safety of Benralizumab in Patients with Eosinophilic Gastritis and/or Gastroenteritis

(The HUDSON GI Study)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

List abbreviations and definitions of specialized or unusual terms, measurements, or units. Examples are provided below. These can be modified at study level.

Abbreviation or special term	Explanation
ADA	Anti-drug antibodies
AE	Adverse Event
ALT	Alanine aminotransferase
ANCOVA	analysis of covariance
AST	Aspartate aminotransferase
Benra	Benralizumab
BMI	Body mass index
CFB	Change from baseline
CGI-C	Clinical Global Impression of Change
CGI-S	Clinical Global Impression of Severity
CI	Confidence interval
CMH	Cochran-Maentel-Haenszel
CRF	Case Report Form (electronic/paper)
CSP	Clinical Study Protocol
CSR	Clinical Study Report
DAE	Discontinuation of investigational product due to adverse event (AE)
DB	Double-Blinded
DILI	Drug-induced liver injury
DL	Direct Likelihood approach
DRMI	Dropout Reason-based Multiple Imputation
DSQ	Dysphagia Symptom Questionnaire

Abbreviation or special term	Explanation
EAP	Exploratory Analysis Plan
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EG	Eosinophilic Gastritis
EGE	Eosinophilic Gastroenteritis
EG-REFS	Eosinophilic Gastritis Endoscopic Reference System
EoD	Eosinophilic Duodenal
EoE	Eosinophilic esophagitis
eos	Eosinophils
EOT	End of Treatment
FAS	Full Analysis Set
GGT	Gamma-Glutamyl Transferase
hpf	high powered field
HRU	Healthcare Resource Utilization
IP	Investigational Product
IPD	Investigational Product Discontinuation (visit)
ITT	Intent to treat
LOCF	Last observation carried forward
LSMD	Least Square Mean Difference
MAR	Missing At Random
MCS	Mental health Component Summary
MCP	Multiple Comparison Procedure
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed effect Models for Repeated Measures
MNAR	Missing Not At Random

Abbreviation or special term	Explanation
M-N	Miettinen and Nurminen
nAb	neutralizing antibodies
OLE	Open-Label Extension
PCS	Physical health Component Summary
PAGI-QoL	Patient Assessment of Upper Gastrointestinal Disorders-Quality of Life
PAGI-SYM	Patient Assessment of Gastrointestinal Disorders Symptom Severity Index
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PK	Pharmacokinetics
PRO	Patient Reported Outcome
PT	Preferred Term
Q4W	Every 4 weeks
SAGED	Symptom Assessment for Gastrointestinal Eosinophilic Diseases
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SF-36 v2	36-Item Short Form Survey, Version 2.0, acute recall
SOC	System Organ Class
TBL	Total bilirubin
TEAEs	Treatment-Emergent Adverse Events
ULN	Upper Limit of Normal
WHO	World Health Organization
WOCF	Worst observation carried forward
WPAI+CIQ	Work Productivity and Activity Impairment questionnaire plus Classroom Impairment Questions

AMENDMENT HISTORY

CATEGORY Change refers to:	Date	Description of change	In line with CSP?	Rationale
Data presentation	27-Aug-2021	Section 7.2 Updated the imputation rules for partial date AE/CM	Yes	Previous rules have some issues which doesn't work for prior and concomitant meds.
Derivation rules for TEAE	27-Aug-2021	Section 4.6.2.1 Updated the end date rule for TEAEs in the on-treatment period	Yes	The visit window should be 3 days, not 7 days.
Data presentation	24-Sep-2021	Adding forest plots for subgroup analysis	Yes	Visualize the results
	24-Sep-2021	Adding line plots for secondary efficacy endpoints	Yes	Visualize the results
Choose an item.	Click or tap here to enter text.			

1 INTRODUCTION

The purpose of this document is to give details for the statistical analysis of study D3258C00001 supporting the clinical study report (CSR). There will be a separate exploratory analysis plan (EAP) prepared to describe the exploratory analyses to be reported outside of the CSR.

The reader is referred to the clinical study protocol (CSP) and the case report form (CRF) for details of study conduct and data collection.

1.1 Study objectives

The study objectives and endpoints are listed below.

Table 1 Objectives and Endpoints

Objectives	Estimands descriptions / endpoints
Primary	
• Part A/B: To compare the effect of benralizumab 30 mg every 4 weeks (Q4W) with placebo on histologic signs and gastrointestinal symptoms in patients with eosinophilic gastritis and/or gastroenteritis	Dual primary endpoints including histology-based and symptom-based endpoints: Histology-based dual primary endpoint • Population: Full analysis set • Histologic Endpoint: Proportion of patients achieving a histological response (defined as ≤6 eosinophils/high power field (hpf) in the stomach and/or, ≤15 eosinophils/hpf in the duodenum) at Week 24 • Intercurrent events: A composite approach whereby participants who receive restricted[a] or rescue medications, or discontinue randomized therapy prior to Week 24 will be considered as histologic non-responders at Week 24

	• Summary measure: Odds ratio and difference in response rates between benralizumab and placebo
	Symptom-based dual primary endpoint • Population: Full analysis set • Symptom Endpoint: Absolute change from baseline in SAGED score at Week 24 • Intercurrent events: A composite approach whereby participants who receive restricted[a] or rescue medications, or discontinue randomized therapy prior to Week 24 will be considered as treatment failures at Week 24. Such patients will have their worst observation prior to or at the intercurrent event carried forwards to Week 24. • Summary measure: Difference in least squares mean change from baseline in SAGED score at Week 24 between benralizumab and placebo
Secondary	
• Part A/B: To compare the effect of benralizumab 30 mg Q4W with placebo on clinical features of eosinophilic gastritis/ gastroenteritis and disease activity	• Key[b]: Percentage change from baseline in tissue eosinophils (stomach and/or duodenum) at Week 24 • Key[b]: Proportion of patients who achieve treatment response: tissue remission (≤ 6 eosinophils/hpf in the stomach and/or ≤ 15 eosinophils/hpf in the duodenum) and an improvement in symptoms at Week 24 • Change from baseline in proportion of vomiting-free days, and change from baseline in frequency of vomiting episodes at Week 24 • Change from baseline in proportion of diarrhea-free days, and change from baseline in frequency of diarrhea episodes at Week 24 • Change from baseline in proportion of days both diarrhea- and constipation-free • Time to clinically meaningful improvement in SAGED score • Change from baseline in PROMIS Fatigue 7a score and PAGI-SYM score at Week 24

• Part A/B: To compare the effect of benralizumab 30mg Q4W with placebo on rescue corticosteroid use	• Key[b]: Proportion of patients with no rescue corticosteroid use up to Week 24
• Part A/B: To compare the effect of benralizumab 30 mg Q4W with placebo on health-related quality of life in patients with eosinophilic gastritis and/or gastroenteritis	• Change from baseline in SF-36v2 PCS and MCS at Week 24 • Change from baseline in PAGI-QOL at Week 24
• To assess the pharmacokinetics and immunogenicity of benralizumab in patients with EG/EGE	• Serum benralizumab concentration over time • Presence of anti-benralizumab antibodies (ADA)
Safety	
• To assess the safety and tolerability of benralizumab in patients with eosinophilic gastritis and/or gastroenteritis	• Safety and tolerability will be evaluated in terms of adverse events, vital signs, physical exam, and clinical laboratory parameters.
Exploratory	
• Part A/B: To describe the effect of benralizumab on clinical features of EG/EGE and disease activity	• Change from baseline in vomiting severity at Week 24 • Change from baseline in diarrhea severity at Week 24 • Change from baseline in endoscopic measures of inflammation and remodeling, EG-REFS, by central reader at Week 24 • Change from baseline in PAGI-SYM, DSQ, PGI-S, PGI-C, CGI-S, and CGI-C at Week 24
• Part C: To describe the longer term effect of benralizumab 30 mg Q4W in patients with eosinophilic gastritis and/or gastroenteritis	• Proportion of patients with histological response at Week 52 • Absolute change from baseline in SAGED score at Week 52 • Proportion of patients who achieve treatment response at Week 52 • Change from baseline in endoscopic measures of inflammation and remodeling, EG-REFS, by central reader at Week 52 • Proportion of patients with no rescue corticosteroid use up to Week 52 • Change from baseline in health related quality of life measures at Week 52 (SF-36v2 PCS and MCS, PAGI-QoL)

	• Use of concomitant medications and dietary restrictions to manage EG/EGE • Maintenance of histological and symptom response from Week 24 to Week 52 in patients randomised to benralizumab in parts A and B
• To characterize the GI tissues and symptoms of participants enrolled and evaluate the effect of benralizumab on these measures. • To evaluate the predictive value of clinical, histologic and endoscopic characteristics, and biomarkers with clinical efficacy outcomes in patients with EG/EGE.	• Histology and immunohistochemistry of GI mucosal biopsies • Blood-based biomarkers • GI mucosal transcriptomics • EG/EGE clinical characteristics and symptoms
• To compare the effect of benralizumab 30mg Q4W with placebo and evaluate the long-term effect of benralizumab 30 mg Q4W on healthcare resource utilization due to both overall GI disease and eosinophilic gastritis and/or gastroenteritis	• Proportion of patients with relevant healthcare resource utilization and number of events by HRU type, including hospitalizations, length of hospital stay, office visits, emergency room visits, tests and procedures
• To evaluate the effect of benralizumab 30 mg Q4W on work productivity and classroom impairment	• WPAI+CIQ scores
• To characterize the patient experience of EG/EGE and its treatments	• Qualitative patient interviews (sub-study) • Eosinophilic gastrointestinal disease diary (optional) • PGI-BR scores

[a] Only restricted medications given that are considered to have a potentially meaningful impact on EG/EGE outcomes will be considered intercurrent events.
[b] Key secondary endpoints are multiplicity protected in the multiple testing procedure for PartB.

1.2 Study design

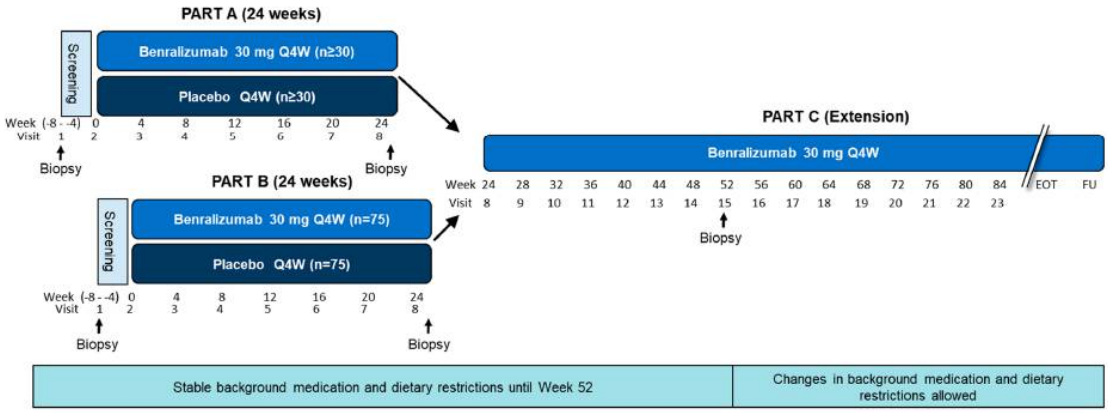
This is a 3-part study. Part A and Part B have identical designs (ie, parallel-group, randomized, double-blinded, placebo-controlled with 24-week treatment periods), with enrollment/randomization performed sequentially. Part A will be enrolled first, followed by Part B. Parts A and Parts B will include approximately 70 and 150 unique participants in total with eosinophilic gastritis (EG; with or without eosinophilic duodenitis) or duodenal-only disease, including a minimum of at least 50 and 110 patients with EG (with or without duodenal involvement) in each part, respectively. After a 4-week to 8-week run-in period, symptomatic participants with EG with or without duodenitis and patients with eosinophilic

duodenitis only, on stable background medications and diet, with histologically-confirmed disease will be randomized 1:1 to benralizumab or placebo treatments. After completing Part A or Part B, participants will continue to Part C, an extended open-label benralizumab treatment period. Participants will remain on stable background medication and diet throughout the first 52 weeks of the study (including Part A/B and the first 28 weeks of Part C). After 52 weeks of treatment with stable diet and medications, investigators may adjust background medication and diet restrictions as clinically indicated.

An Independent Data Monitoring Committee will be used for this study. Further details, composition and operation of the Independent Data Monitoring Committee are described in a separate Independent Data Monitoring Committee charter.

For an overview of the study design see Figure 1.

Figure 1 Study design flow chart



Q4W: every fourth week; EoT: end of treatment; FU: follow-up.

1.3 Sample size justification

The sample size calculations are based on the change from baseline in SAGED score at Week 24 dual primary endpoint. Changes in the other dual primary histological response rate endpoint require smaller sample sizes as larger relative differences are assumed (10% placebo response versus 80% benralizumab response).

Part A will enroll approximately 50 patients with EG. Eligible patients with duodenal only disease will also be included whilst enrolling the required 50 EG patients, with no minimum requirement on the number of duodenal only patients. A minimum of 60 and maximum of 80 patients in the overall population (i.e. with either EG or duodenal only disease) will be enrolled in part A. Thirty patients per arm in the overall population and 25 patients per arm with EG will provide >85% / >80% power for statistical significance at the 10% 2-sided level in each population respectively, if the true treatment effect for the change from baseline

in SAGED score is a standardized effect size (difference in means/ standard deviation) of 0.75. A standardized effect size is used as the variability of the SAGED Score is not yet known. An effect size of 0.75 is a similar relative effect size as that observed in the lirentelimab (formerly antolimab) Phase 2 trial. Approximately 20 adolescent patients (12 to 17 years of age) with either EG or duodenal only disease are targeted for inclusion in Part A.

Part B will enroll approximately 110 patients with EG, along with approximately 40 patients with duodenal only disease. The total part B sample size of 150 patients (75 per arm) in the overall population will consist of approximately 110 adults and target approximately 40 adolescents. Seventy-five patients per arm in the overall population in part B, and 55/arm in the population of patients with EG will provide >95% and >90% power for statistical significance at the 5% 2-sided level if the true treatment effect for the change in SAGED score is a standardized effect size of 0.65, for these two populations respectively. A smaller treatment effect is assumed in Part B to account for any increased variability or heterogeneity across regions or other subgroups. A sample size re- estimation will be performed at the end of Part A using only Part A data, to confirm the Part B size remains appropriate once the variability and responsiveness of the PRO tool is established.

Clinical benefit in the adolescent population will be concluded if broadly consistent results to the overall population are demonstrated. With 20 adolescents per arm in Part B there is a high chance of demonstrating consistent effects if they truly exist. For the histological response rate at Week 24 endpoint, there is a >95% chance of observing an adolescent treatment difference that is at least half of the overall population effect, assuming the true histological response rates are 70% on benralizumab and 10% on placebo. For the change in SAGED score endpoint at Week 24 there is an 87% chance of observing an adolescent treatment difference that is at least half of the overall population effect, assuming the true treatment effect for the SAGED score endpoint is as outlined in the sample size justification for Part B above.

2 CHANGES TO PROTOCOL PLANNED ANALYSES

There are no changes to Protocol planned analyses.

3 DATA ANALYSIS CONSIDERATIONS

3.1 Timing of Analyses

Part A will remain blinded until the primary analysis of Part A, which will be performed when all Part A patients have completed Part A, i.e., the 24-week placebo controlled period. An earlier blinded cut of Part A data may be used to perform longitudinal psychometric analysis to confirm measurement properties and responsiveness of the de novo SAGED PRO instrument. These analyses will be described separately in a psychometric analysis plan.

Part B consists of a separate patient cohort to Part A and will provide the pivotal dataset for the primary assessment of efficacy. Patients recruited in Part B will remain blinded until the primary analysis of Part B, which will be performed when all Part B patients have completed Part B, ie the 24-week placebo-controlled period. No data from Part B patients will be included in the earlier end-of-Part A analysis. Likewise, no data from Part A patients will be included in the Part B primary analysis.

The main CSR for the study will be written based on the Part B analysis and all Part B data reported up to the Part B database lock will be included. An earlier separate report of the Part A data may be written, if there is long enough time between parts A and B completing to warrant a separate report, otherwise Part A data will also be included within the main Part B CSR.

Data from the open-label benralizumab part of the study (Part C) will primarily be reported separately from Parts A and B. In addition, a key subset of summaries may also be repeated integrating data across the entire study period (i.e., Part A together with Part C and Part B together with Part C).Complete data may also be summarized together for some key subset summaries (i.e., Part A and Part B and Part C).

3.2 Analysis Populations

The following populations are defined:

Table 2 Populations for Analysis

Population/Analysis set	Description
Enrolled	All participants who sign the ICF
All patients analysis set	All patients screened for the study, to be used for reporting of disposition and screening failures
Full analysis set (FAS)	All randomized patients who received at least 1 dose of investigational product (IP), irrespective of their protocol adherence and continued participation in the study. Patients will be analyzed according to their randomized treatment irrespective of whether or not they have been given the incorrect IP or prematurely discontinued, according to the Intention to Treat principle. Patients who withdraw consent or assent to participate in the study will be

	included up to the date of their study termination.
Safety analysis set	The safety analysis set consists of all patients who have received at least one dose of investigational product. Erroneously treated patients (eg, those randomized to one treatment but actually given the other treatment) are accounted for in the treatment group of the treatment they actually received. A patient who has on one or several occasions received active investigational product is classified as active. Safety summaries and anti-benralizumab antibody (immunogenicity) analyses will be based on this analysis set and for whom any post-dose data are available.
Pharmacokinetic analysis set	All patients who received benralizumab and from whom pharmacokinetic 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 pharmacokinetic observation post-first dose. All pharmacokinetic summaries will be based on this analysis set.
Open-label benralizumab analysis set	All patients who receive at least 1 dose of benralizumab after the end of the Week 24 double-blind treatment period

3.3 General Considerations

The primary efficacy analyses will be based on the double-blind placebo-controlled first 24 weeks of the study (DB period). In this part of the study all efficacy analyses will use the full analysis set (FAS) as defined in Section 3.2, and patients will be analyzed according to their randomized treatment, following the Intent-to-Treat (ITT) principle.

A composite estimand strategy will be used for the primary analyses of endpoints at Week 24 whereby any patients who have required rescue corticosteroids, use of a restricted medication, or discontinued IP during the first 24 weeks will be considered as treatment failures at Week 24.

A review of all concomitant medications during the study will be performed prior to database lock to identify events to be considered as treatment failure in these analyses, whereby only medications considered likely to have a meaningful impact on EG/EGE outcomes would be considered intercurrent events. This could include addition of a new therapy or an increase in previously stable background medications to treat EG/EGE. For the histologic response rate endpoint, patients with these intercurrent events prior to Week 24 will be considered non-responders at Week 24; for the change from baseline in SAGED score at Week 24 endpoint, as well as other change from baseline continuous endpoints, any patients experiencing the described intercurrent events will have their worst observation (including baseline value) prior to or at the intercurrent event carried forward to Week 24.

All patients who prematurely discontinue from IP, use restricted medications, or have any changes to background therapies for EG/EGE as described above are asked to come in for all visits until they complete the overall study duration. Therefore, sensitivity analyses can be performed to assess the robustness of the efficacy results to these estimand approaches and missing data assumptions as described in individual endpoint analysis methods in Section 4.2.

The statistical analyses of the DB period are designed to compare both efficacy and safety of benralizumab to placebo, while the OLE period is designed to evaluate the long-term safety and tolerability and persistence of effect of benralizumab in this patient population. Details regarding primary and key secondary estimands with additional details including sensitivity analyses are provided in Section 4.2.

All analyses of part C of the study including week 52 endpoints will be descriptive as no placebo control is available at that timepoint and so no hypothesis testing will be performed. Week 52 analyses will primarily be presented on the FAS, but a repeat of key analyses may also be produced on the open-label benralizumab analysis set to ensure only patients who switched to receive benralizumab after 24 weeks are included in the denominator for that group and to ensure a meaningful interpretation of the placebo-to-benralizumab patients.

Demography and baseline characteristics will be summarized by treatment group for the FAS. In the event that there are major differences between the FAS and safety analysis set, these summaries may also be repeated for the safety analysis set.

Summary data will be presented in tabular format by treatment group. Categorical data will be summarized by the number and percentage of subjects in each category. Continuous variables

will be summarized by descriptive statistics including N, mean, SD, median, and range. Data listings will be sorted by treatment group and patient number.

All hypothesis testing will be reported using 2-sided tests. Any *p*-values presented for endpoints other than those included in the hierarchical testing strategy will be nominal (i.e., not multiplicity adjusted). All *p*-values will be rounded to 4 decimal places.

The data analyses will be conducted using the SAS® System version 9.4 or above (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. Pharmacokinetic analyses will be performed using NONMEM or other appropriate software.

3.3.1 General Study Level Definitions

3.3.1.1 Study period definition

- **Double Blind (DB) period**: from the date of randomization up to the first dose of Open Label benralizumab 30 mg at Week 24, for both Part A and Part B subjects. If a participant does not transition into OLE then all the data will be included in the double-blind period.
- **Open Label Extension (OLE) period**: from the first dose of Open Label benralizumab 30 mg at Week 24 to the end of study, for the subjects who participated in Part C.

For all assessments scheduled at first dose date of OLE (V8, Week 24), 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 for Adverse Events which are defined in Section 4.6.2.1.

3.3.1.2 Baseline definition

In general, the last non missing value on or prior to the date of randomization will serve as the baseline measurement for efficacy endpoints while the last non missing value prior to first dose of study treatment will serve as the baseline measurement for safety endpoints. If there is no value prior to the randomization (or the first dose of study treatment, depending on the endpoint), the baseline value will not be imputed and will be set to missing.

For the daily assessment variables including SAGED score and other daily symptoms scores which are calculated during a certain period (e.g., bi-weekly), the score calculated during the cycle prior to Baseline Visit 2 (Week 0) will be the baseline score.

For analyses re-setting baseline to the start of Part C, the baseline value is set to the last non missing value prior to the first dose of benralizumab in Part C (i.e., likely the Week 24 measurement).

3.3.2 Visit Window

The adjusted analysis-defined windows for assessments conducted every 4 weeks are summarized in the Table 3 below.

Table 3 Visit windows for assessments conducted every 4 weeks

Adjusted defined window visit	Scheduled study day	Maximum windows
Week 0	1	Study Day = 1
Week 4	29	2 ≤ Study Day ≤ 42
Week 8	57	43 ≤ 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 [1]
Week 28	197	183 ≤ Study Day ≤ 210
Week 32	225	211 ≤ Study Day ≤ 238
Week 36	253	239 ≤ Study Day ≤ 266
Week 40	281	267 ≤ Study Day ≤ 294
Week 44	309	295 ≤ Study Day ≤ 322
Week 48	337	323 ≤ Study Day ≤ 350
Week 52	365	351 ≤ Study Day ≤ 378
Week 56	393	379 ≤ Study Day ≤ 406
Week 60	421	407 ≤ Study Day ≤ 434
Week 64	449	435 ≤ Study Day ≤ 462
Week 68	477	463 ≤ Study Day ≤ 490
Week 72	505	491 ≤ Study Day ≤ 518
Week 76	533	519 ≤ Study Day ≤ 546
Week 80	561	547 ≤ Study Day ≤ 574
Week 84	589	575 ≤ Study Day ≤ 602
…	previous scheduled study day+28	previous lower bound study day+28≤ Study Day ≤ previous upper bound study day+28

[1.] The windowing will only be performed for assessments within the appropriate periods e.g., double blind versus open label.

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

Date of assessment – Date of randomization + 1

Study days before randomization will be defined as follows:

Date of assessment – Date of randomization

By this definition, the day of randomization will be study day 1 and the day before the day of randomization will be study day – 1. There is no study day 0. The planned date of Visit 3 (Week 4) will be study day 29 (= 28 + 1), for example.

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

- If there are 2 or more observations within the same visit window, then the non-missing observation closest to the scheduled visit will be used in the analysis.
- If 2 observations are equidistant from the scheduled visit, then the non-missing observation with the earlier collection date will be used in the analysis.
- If 2 observations are collected on the same day, then the non-missing observation with the earlier collection time will be included in the analysis.

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

The SAGED score and other daily symptoms are calculated using data captured over 14-day periods. The scheduled bi-weekly (14-day) study windows are listed in Table 4.

Table 4 Bi-weekly windows for daily diary assessments

Adjusted defined window visit	Scheduled study day	Maximum windows
Baseline	1	The last 14 days from Study Day -14 to Study Day -1
Week 2	15	Study Day 1 to Study Day 14
Week 4	29	Study Day 15 to Study Day 28
Week 6	43	Study Day 29 to Study Day 42
Week 8	57	Study Day 43 to Study Day 56
Week 10	71	Study Day 57 to Study Day 70
Week 12	85	Study Day 71 to Study Day 84
Week 14	99	Study Day 85 to Study Day 98
Week 16	113	Study Day 99 to Study Day 112
Week 18	127	Study Day 113 to Study Day 126
Week 20	141	Study Day 127 to Study Day 140
Week 22	155	Study Day 141 to Study Day 154
Week 24	169	Study Day 155 to Study Day 168 [1]

Week 26	183	Study Day 169 to Study Day 182
Week 28	197	Study Day 183 to Study Day 196
Week 30	211	Study Day 197 to Study Day 210
Week 32	225	Study Day 211 to Study Day 224
Week 34	239	Study Day 225 to Study Day 238
Week 36	253	Study Day 239 to Study Day 252
Week 38	267	Study Day 253 to Study Day 266
Week 40	281	Study Day 267 to Study Day 280
Week 42	295	Study Day 281 to Study Day 294
Week 44	309	Study Day 295 to Study Day 308
Week 46	323	Study Day 309 to Study Day 322
Week 48	337	Study Day 323 to Study Day 336
Week 50	351	Study Day 337 to Study Day 350
Week 52	365	Study Day 351 to Study Day 364
…	previous scheduled study day+14	From previous lower bound study day+14 to previous upper bound study day+14

[1.] The windowing will only be performed for assessments within the appropriate periods e.g., double blind versus open label.

3.3.3 Handling of Unscheduled Visits

For overall analyses not based on any particular study visit, all data will be listed and/or analysed, including any repeated or unscheduled visits, unless otherwise specified.

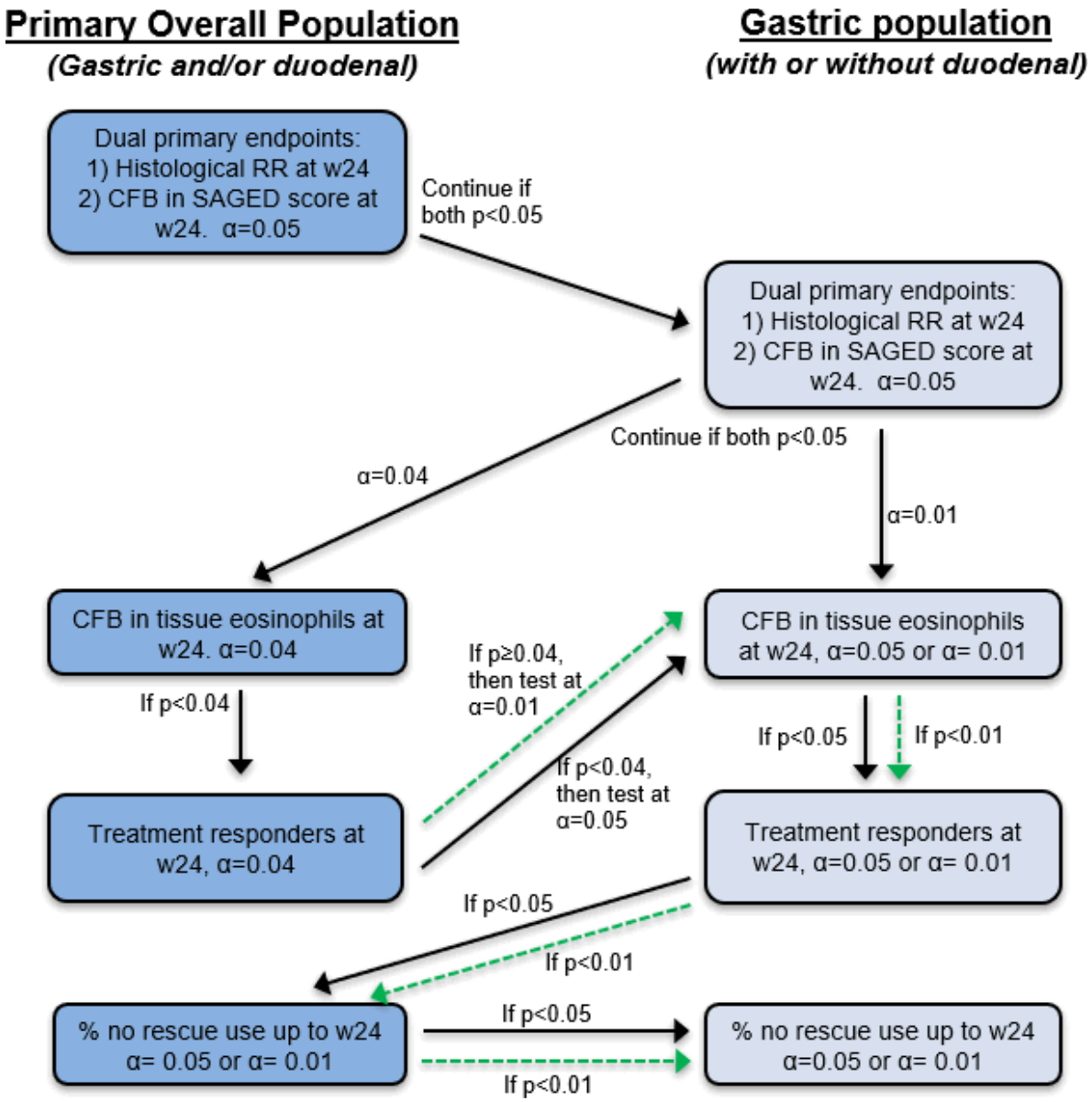
For the analyses based on particular study visit, please refer to Section 3.3.2for the visit determination and scheduled/unscheduled visit handling.

All collected data including unscheduled visits data will be listed in the listings.

3.3.4 Multiplicity/Multiple Comparisons

To account for multiplicity in the statistical testing of the dual primary endpoints (histological response rate and change from baseline in SAGED score) and 3 key secondary endpoints (change from baseline in tissue eosinophils, treatment responders, and proportion with no rescue use) in both the primary overall population and the population of patients with EG, in part B of the study, a testing strategy will be used to strongly control the overall type I error rate at the 0.05 level. The testing strategy is depicted in Figure 2.

Figure 2 Multiple Testing Strategy for the Primary and Key Secondary Endpoints



CFB change from baseline.
Note: testing within the hierarchy can only continue if the p-value is below the indicated threshold, and the treatment effect favors the benralizumab treatment arm.

The dual primary endpoints will be tested sequentially at the 2-sided 5% level (1-sided 2.5% level) in the overall and gastric populations in the order shown below:

1. Histological response rate at Week 24 in the overall population
2. Change from baseline in SAGED score at Week 24 in the overall population
3. Histological response rate at Week 24 in the gastric population

4. Change from baseline in SAGED score at Week 24 in the gastric population

If at any point a null hypothesis cannot be rejected in favour of the benralizumab treated group, further hypothesis testing will stop, and no further hypotheses will be rejected for any endpoint in either population.

If the null hypotheses for both dual primary endpoints in both populations are successfully rejected, with treatment effects in favor of benralizumab and p-values <0.05, testing will then proceed to the key secondary endpoints in the overall and gastric populations. In this step, α=0.04 will be assigned to test the key secondary endpoints in the overall population, and α=0.01 will be reserved to test the key secondary endpoints hierarchically within the gastric population. The following principles will apply:

- If the null hypotheses are rejected in the overall population for both the change in tissue eosinophils and treatment responder endpoints (both at p<0.04 with a treatment effect favoring benralizumab), then the α=0.04 will be recycled to the gastric population and key secondary endpoints in the gastric population will be tested at the full α=0.05 level.
- If the null hypotheses cannot be rejected in the overall population for both the change in tissue eosinophils and treatment responder endpoints, then the α=0.04 cannot be recycled and all subsequent testing in the hierarchy outlined in Figure 2 above will occur at the α=0.01 level.

Hierarchical testing will continue at the given alpha level until a null hypothesis cannot be rejected; at this time, further testing will stop, and no subsequent null hypotheses in the testing hierarchy will be rejected.

The above approach will be applied to the pivotal part B of the study. In part A of the study, hypothesis testing will be performed at the 10% 2-sided significance level for all endpoints (1-sided 5% level). For the primary endpoints assessed in the overall population and within the EG population, the same hierarchical testing strategy as outlined for the primary endpoints for part B above will be applied, but with null hypotheses being rejected if statistical significance is achieved at the 2-sided 10% levels.

It is noted that for the purpose of marketing approval, both primary endpoints would need to be statistically significant for Part B.

3.3.5 Handling of Protocol Deviations in Study Analysis

Patients who do not meet eligibility criteria but are still randomized will be analyzed according to the analysis sets described in Section 3.2. There is no intention to perform a per-protocol (PP) analysis in this study.

3.3.5.1 Important protocol deviations

Only important protocol deviations will be listed and tabulated in the Clinical Study Report (CSR). 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 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

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.

4 STATISTICAL ANALYSIS

In general, Part A and Part B will be summarized separately. Part C will also be summarized separately from Part A and Part B. Part C will not have baseline related summaries, e.g., demographic, medical history, Prior medication etc.

4.1 Study Population

4.1.1 Subject Disposition and Completion Status

4.1.1.1 Definitions and Derivations

Disposition information can be directly obtained from case report form (CRF).

4.1.1.2 Presentation

Patient disposition will be summarized using all patients analysis set. The total number of patients will be summarized for the following groups: those who enrolled and those who were not randomized (and reason). The number and percentage of patients within each treatment group and 'Total' will be presented by the following categories: randomized, received treatment with study drug, did not receive treatment with study drug (and reason), completed treatment with study drug in DB treatment period, discontinued treatment with study drug in DB treatment period (and reason), discontinued treatment with study drug in DB treatment period but completed study follow-up, completed DB treatment period study, and withdrawn from study in DB treatment period (and reason).

For Part C the number and percentage of patients within each treatment group and 'Total' will be presented by the following categories: enrolled in OLE treatment period, did not enrol in OLE treatment (and reason), completed OLE treatment with study drug, discontinued OLE treatment (and reason), discontinued OLE treatment but completed study follow-up, completed OLE treatment, and withdrawn from OLE treatment (and reason).

Disposition and screen failure information will be listed using the all patients analysis set.

The number of patients remaining on treatment, patients discontinued IP but still in study follow-up, and patients who withdrew from the study will be summarized by treatment group and scheduled visit, separately for patients in the full analysis set.

The number of patients by country and centre will also be summarized by treatment group and 'total' in the FAS.

4.1.2 Analysis Sets

4.1.2.1 Definitions and Derivations

Definition for different analysis sets are shown in Table 2 in Section 3.2.

4.1.2.2 Presentation

The number of patients within each treatment group and 'Total' will be presented by the following categories: included in each analysis set, excluded in each analysis set (and reason).

Reasons why subjects excluded from certain analysis sets will be listed.

4.1.3 Protocol Deviations

4.1.3.1 Definitions and Derivations

The definition of protocol deviations and derivations are described in Section 3.3.5.1.

4.1.3.2 Presentation

Protocol deviations will be summarized using FAS. The number of patients within each treatment group will be presented by the following categories: with at least one important protocol deviation, with each important protocol deviation.

Information for subjects with important protocol deviation will also be listed.

4.1.4 Demographics

4.1.4.1 Definitions and Derivations

Most demography information can be directly obtained from case report form (CRF). Age will be derived from the date of informed consent-date to birth, rounded down to the nearest integer. For patients in countries where date of birth is not recorded, the age as recorded in the electronic case report form (eCRF) will be used.

4.1.4.2 Presentation

Demography will be summarized by treatment group and 'Total' for the FAS, using frequency and percentages (for categorical variables) and descriptive statistics of *n*, mean, standard deviation, minimum, Q1, median, Q3, and maximum (for continuous variables). If there are major differences between the FAS and safety analysis set, these summaries may also be repeated for the safety analysis set.

4.1.5 Baseline Characteristics

4.1.5.1 Definitions and Derivations

Various baseline characteristics will be summarized, including patient characteristics (weight, height, BMI, baseline eosinophil count, historical eosinophil count, etc).

Baseline characteristics information can be directly obtained from CRF.

4.1.5.2 Presentation

Baseline characteristics will be summarized in a similar manner to that described for baseline demographics (see Section 4.1.4.2).

4.1.6 Disease Characteristics

4.1.6.1 Definitions and Derivations

Baseline disease characteristics will include but not limited to the location of disease, background medications use, refractory to background medication, dietary restriction, baseline steroid use, refractory to steroid, duration of EG/EGE.

4.1.6.2 Presentation

Disease characteristics will be summarized in a similar manner to that described for baseline demographics (see Section 4.1.4.2).

4.1.7 Medical History and Concomitant Disease

4.1.7.1 Definitions and Derivations

Medical history information can be directly obtained from CRF.

4.1.7.2 Presentation

Medical history will be summarized by treatment group and 'Total'. Medical history will be summarised by system organ class (SOC) and preferred term (PT) assigned to the event by MedDRA. For each PT, the number and percentage of patients reporting at least one occurrence will be presented, i.e., for a patient multiple occurrences of an medical history will only be counted once.

Specific medical and surgical histories will be summarized separately.

4.1.8 Prior and Concomitant Medications

4.1.8.1 Definitions and Derivations

A medication will be regarded as prior if it started prior to the date of randomization and 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. Medications with start date after the on-treatment period will not be considered as concomitant.

If a medication started and stopped on the date of randomization, it will be considered as concomitant.

Medications that are considered to have a potentially meaningful impact on EG/EGE outcomes will be considered as intercurrent events. These medications will be carefully reviewed and documented prior to unblinding.

4.1.8.2 Presentation

The number and percentage of patients who take prior medications, those who take allowed concomitant medications and those who take medications that are considered intercurrent events during the study, will be presented by treatment group and 'Total'. Concomitant medications will be classified according to the WHO-Drug. The summary tables will present data by generic term within ATC code.

4.1.9 Study Drug Compliance

4.1.9.1 Definitions and Derivations

Duration of study drug exposure is calculated in days as:

Study drug exposure = last dose date of IP - first dose date of IP+1.

Treatment compliance is calculated as:

Study treatment compliance = (total doses administered/total doses expected) x 100.

Patients who received no study treatment will have zero compliance. Total number of doses expected includes all visits with protocol scheduled IP administration on or before a patient's IP discontinuation or treatment complete date.

4.1.9.2 Presentation

Durations and compliance rate of the study drug exposure will be summarized descriptively by treatment group and 'Total' for the safety analysis set within each Part (Part A or Part B, Part C) as well as overall study period.

4.2 Endpoint Analyses

This section covers details related to the endpoint analyses such as primary, secondary, other endpoints including sensitivity and supportive analyses. The primary and key secondary endpoints will be analysed following the hierarchical order based on the overall population and within the subgroup of patients with gastric disease (with or without intestinal involvement), respectively. See Section 3.3.4 for details.

For part B analyses, alpha level will follow the instruction shown in Figure 2 for primary and key secondary endpoints. For other endpoints not included in the testing hierarchy, alpha will be 0.05 and all p-values will be considered nominal. For part A analyses, alpha level will be 0.1.

A composite approach will be used for all the responder analyses whereby any patients who have intercurrent events of rescue use, restricted medication use, or randomized treatment discontinuation during the analyses period will be considered as non-responders.

A composite approach will be used for all the continuous analyses, where worst observation prior to or at the intercurrent events of rescue use, restricted medication use, or randomized treatment discontinuation will be carried forward (WOCF). If a participant has had only improvements since baseline this will result in the baseline value being carried forwards.

For the descriptive summaries of each endpoint, the number and percentage of subjects with imputation based on composite approach will be displayed. The descriptive summaries will summarise the results once the imputations have been made.

Table 5 Primary, key secondary efficacy and main safety estimands

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Objective 1: Part A/B: To compare the effect of benralizumab 30 mg every 4 weeks (Q4W) with placebo on histologic signs and gastrointestinal symptoms in patients with eosinophilic gastritis and/or gastroenteritis					
Primary/MCP	Proportion of patients with a histologic response at Week 24	FAS	• Treatment discontinuation - composite (treated as non-responder) • Required rescue corticosteroids or use of a restricted medication[a] – composite (treated as non-responder)	Odds ratio and difference in proportions from stratified CMH test at Week 24.	Section 4.2.1
Primary/MCP	CFB in SAGED score at Week 24	FAS	• Treatment discontinuation – composite (WOCF) • Required rescue corticosteroids or use of a restricted medication[a] – composite (WOCF)	Mean difference between interventions (LSMD from CFB ANCOVA) at Week 24.	Section 4.2.2
Objective 2: To compare the effect of benralizumab 30 mg Q4W with placebo on clinical features of eosinophilic gastritis/ gastroenteritis and disease activity					
Key Secondary/MCP	Percentage CFB in tissue eosinophils (stomach and/or duodenum if applicable) at Week 24	FAS	• Treatment discontinuation – composite (WOCF) • Required rescue corticosteroids or use of a restricted medication[a] – composite (WOCF)	Mean difference between interventions (LSMD from percentage CFB ANCOVA) at Week 24.	Section 4.2.3

Statistical category	Endpoint	Population	Intercurrent event strategy	Population level summary (analysis)	Details in section
Key Secondary/MCP	Proportion of patients who achieve treatment response at Week 24	FAS	• Treatment discontinuation - composite (treated as non-responder) • Required rescue corticosteroids or use of a restricted medication[a] – composite (treated as non-responder)	Odds ratio and difference in proportions from stratified CMH test at Week 24.	Section 4.2.4
Objective 3: Part A/B: To compare the effect of benralizumab 30mg Q4W with placebo on rescue use					
Key Secondary/MCP	Proportion of patients with no rescue corticosteroid use up to Week 24	FAS	• Treatment discontinuation - composite (treated as rescue used)	Odds ratio and difference in proportions from stratified CMH test at Week 24.	Section 4.2.5
Objective 4: To assess the safety of benralizumab in patients with eosinophilic gastritis and/or gastroenteritis					
Safety	Adverse Event, vital signs and laboratory parameters DB+OLE	Safety analysis set	Remained adherent to intervention (on-treatment)	Descriptive	Section 4.6

MCP = Multiple comparisons procedure; EOT = End of treatment; LSMD = Least squares mean difference; CFB = Change from baseline; CMH test = Cochran-Maentel-Haenszel test; ANCOVA = Analysis of covariance; WOCF = Worst observation carried forward.
[a] Only those medications that are considered to have a potentially meaningful impact on EG/EGE outcomes will be considered as intercurrent events, therefore treated with composite approach.

4.2.1 Primary Endpoint – Histological Response Rate

4.2.1.1 Definition

There are two primary endpoints for this study, one is histology-based and the other one is symptom-based.

For the first primary endpoint, the proportion of patients achieving a histological response at Week 24, is defined as below:

≤6 eosinophils/hpf in the stomach for the patients with only gastric disease at baseline.

≤6 eosinophils/hpf in the stomach and ≤15 eosinophils/hpf in the duodenum for the patients with gastric + duodenal disease at baseline.

≤15 eosinophils/hpf in the duodenum for the patients with only duodenal disease at baseline.

4.2.1.2 Derivation

The composite approach for responder analyses mentioned in Section 4.2 will be applied.

The analysis of histological response rate at Week 24 will include data collected at the Week 24 visits including a window of +/-2 weeks within DB period.

4.2.1.3 Handling of Dropouts and Missing Data

Patients with no biopsy data at Week 24 will be considered as non-responders.

4.2.1.4 Primary Analysis of Primary Endpoint

For the primary endpoint of proportion of patients achieving a histological response at Week 24, the null hypothesis is that the proportion of responders on benralizumab 30 mg Q4W is equal to the proportion of responders on placebo. The alternative hypothesis is that the proportion of responders on benralizumab 30 mg Q4W is not equal to the proportion of responders on placebo, i.e.:

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

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

Hypothesis testing for the primary analyses of the primary endpoints in the overall population, and in the population of patients with EG, will be performed at the 2-sided 10% / 5% significance level for parts A and B of the study, respectively. If the p-value is less than 0.1 / 0.05 (for parts A / B respectively) and the treatment effect favors benralizumab, reject H_0 and accept H_1.

A multiple testing procedure will be applied to the primary and key secondary endpoints, in the overall population and in the population of patients with EG, details are provided in Section 3.3.4.

Histological response rate at Week 24 will be compared between benralizumab and placebo using a Cochran-Maentel-Haenszel (CMH) test stratified by location of disease (EG with or without duodenal involvement vs duodenal disease alone), region and baseline steroid use. The results of the analysis will be presented using an odds ratio, together with its associated confidence intervals and 2-sided p-value. Results will be transformed into a difference in proportions for ease of interpretation.

The number and percentage of histological responders will also be summarized by randomized treatment with confidence intervals around the proportions for Week 24 and Week 52.

4.2.1.5 Sensitivity Analyses of the Primary Endpoint

A sensitivity analysis to the Cochran-Maentel-Haenszel test may be performed using a logistic regression with the same covariates. Sensitivity analyses may also be performed including all post baseline biopsy data to assess the impact of any additional data collected that may have been outside of the Week 24 window.

Sensitivity analysis will also be performed by including all available data regardless of intercurrent events (i.e., treatment policy strategy).

4.2.1.6 Supplementary Analyses of the Primary Endpoint

The number and percentage of patients with histological response by visit will be summarized descriptively.

In addition, the number and percentage of patients achieving tissue eosinophil reductions to different levels by the baseline disease location will be descriptively summarized by visit to explore alternative histological response thresholds.

4.2.1.7 Subgroup Analyses

As well as analyses in the overall population and within the subgroup of patients with gastric disease (with or without intestinal involvement) as described previously, analyses of the primary endpoints will be performed to explore the consistency of the treatment effect across other pre-defined subgroups. Analyses will only be performed if sufficient patients in each level of the subgroup are available, condensing of groupings may be considered if necessary.

- Location of disease: gastric alone, gastric + duodenal, duodenal alone
- Age group: age<18 vs age≥18 years
- Gender (male vs female)
- Race –White, Asian, Other
- Background medications use: Yes vs No
- Refractory to background medication: Yes vs No
- Dietary restriction: Yes vs No
- Baseline steroid use: Yes vs No

- Refractory to steroid: Yes vs No
- Prior steroid response: Resolved, Improved, No Effect or Unknown.
- Decrease in background medications during first 24 weeks: Yes vs No
- Duration of EG/EGE: <5, 5-10, >10 years

Subgroup analyses will be performed for the proportion of patients achieving a histological response at Week 24, comparing benralizumab and placebo using logistic regression models. The dependent variable will be achieving a histological response at Week 24 (Yes, No) and the independent variables will include the same covariates as in the primary analysis along with additional terms for the subgroup main effect and the treatment×subgroup interaction. Marginal standardization methods (Bartlett 2018) will be used for the model estimates for all rate analyses, including logistic regression, unless otherwise specified. The results from logistic regression will be shown in forest plot for better visualization.

It is noted that if there are low counts in some of the treatment by subgroup response groups, the logistic regression models may not be reliably estimable, in which case data will be presented descriptively without formal analysis.

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.

4.2.2 Primary Endpoint – Change from Baseline in SAGED Score

4.2.2.1 Definition

The Symptom Assessment for Gastrointestinal Eosinophilic Diseases (SAGED) instrument was developed to measure gastrointestinal symptoms in participants diagnosed with EG/EGE. The SAGED instrument, comprising 8 items, measures the severity of abdominal pain, nausea, bloating, early satiety, lack of appetite, vomiting, diarrhea, and frequency of vomiting. Severity for each concept is assessed using an 11-point numerical rating scale (where 0 = 'none' and 10 = 'worst imaginable'). Frequency of vomiting is reported as number of episodes of vomiting within the past 24 hours.

The SAGED instrument is a daily diary to be completed by participants each evening to record their symptoms during the past 24 hours. It will be completed by the participant every evening starting on the evening from V1 (Week -8 to -4) until V21 (Week 76).

The SAGED score (range: 0-50) includes the items measuring severity of abdominal pain, nausea, bloating, early satiety and lack of appetite. Higher scores indicate greater symptom severity. The

severity of vomiting and severity of diarrhea scores (range: 0-10) are reported separately (summarized under Section 4.2.14 and Section 4.2.15). Frequency of vomiting is reported as a count (summarized under Section 4.2.6).

4.2.2.2 Derivation

The second of the dual-primary endpoints, the change from baseline in SAGED score at Week 24 (follow the visit window in Table 4 within DB period) is calculated by summing the abdominal pain, nausea, bloating, early satiety and loss of appetite severity items from SAGED on a daily basis, and then taking the 14-day mean of the daily sums.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

Responder thresholds to support the primary change in SAGED score endpoint will be explored in the longitudinal psychometric analyses that will be performed with data from patients recruited in Part A of the study. These psychometric analyses will confirm measurement properties and responsiveness of the de novo SAGED PRO instrument, and methods for these analyses will be described separately in a psychometric analysis plan. Responder thresholds will be used to determine the treatment responders (summarized under Section 4.2.4.1).

4.2.2.3 Handling of Dropouts and Missing Data

For the the change from baseline in SAGED score primary endpoint, calculation of the 14-day mean scores will require at least 8 days of evaluable data during the period; otherwise the mean scores will be set to missing.

4.2.2.4 Primary Analysis of Primary Endpoint

For the second primary endpoint of change from baseline in SAGED score at Week 24, the null hypothesis is that the change in SAGED score for patients on benralizumab 30 mg Q4W is equal to the change for patients on placebo. The alternative hypothesis is that the change in SAGED score for patients on benralizumab 30 mg Q4W is not equal to the change for patients on placebo, i.e.:

H_0: Difference in change from baseline in SAGED score at Week 24 (benralizumab 30 mg – Placebo) = 0

H_1: Difference in change from baseline in SAGED score at Week 24 (benralizumab 30 mg – Placebo) ≠ 0

Significance level and multiple testing procedure are applied the same way as mentioned in histological response rate estimand (see Section 4.1.2.4).

After composite approach imputations are made, the available SAGED score at Week 24 will then be analysed using an analysis of covariance (ANCOVA) model. The model will include change from baseline in SAGED score at Week 24 as the dependent variable, baseline SAGED score as a continuous covariate, along with location of disease (EG vs duodenal alone), region

and baseline steroid use as categorical variables. The model will be used to estimate the mean change from baseline at Week 24 for each treatment group and the difference versus placebo, with corresponding confidence limits. A p-value, corresponding to a 2-sided test, will be presented to compare the benralizumab and placebo treatment groups.

The change from baseline in SAGED score at Week 24 and Week 52 will also be summarized descriptively by randomised treatment group.

4.2.2.5 Sensitivity Analyses of the Primary Endpoint

Sensitivity analyses for the primary endpoint based on different missing data mechanism assumptions will be used to explore the robustness of any treatment effect including multiple imputation approaches. Alternative estimand approaches for the handling of intercurrent events of rescue therapy use will also be explored via sensitivity analyses, e.g., MMRM for effectiveness. Full details of sensitivity analyses can be found in Section 7.1.

Sensitivity analysis will also be performed by including all available data regardless of intercurrent events (i.e., treatment policy strategy).

Sensitivity analyses may also be considered to explore the effect of extreme outliers on individual endpoints, such as rank based methods.

4.2.2.6 Supplementary Analyses of the Primary Endpoint

The proportion of patients who achieve SAGED score response threshold at Week 24 will be compared between benralizumab and placebo using CMH along with location of disease (EG vs duodenal alone), region and baseline steroid use as categorical variables.

LS mean change from baseline in SAGED score over time will be plotted by treatment with 95% CIs. The LS means are obtained from ANCOVA model which includes change from baseline in SAGED score at each 14 days as the dependent variable, baseline SAGED score as a continuous covariate, along with location of disease (EG vs duodenal alone), region and baseline steroid use as categorical variables.

A plot will be produced showing the mean change from baseline at week 24 in each of the 5 individual symptoms which make up the SAGED score, by treatment group. 14 day mean scores of the 5 individual symptoms will be used, and the plot will highlight the changes in scores by connecting baseline and week 24 mean scores with the change and 95% CIs annotated onto the plot. Means and confidence intervals from these individual symptom scores will be based on descriptive statistics. Statistical analyses will be considered for these scores if any trends warrant further investigation.

4.2.2.7 Subgroup Analyses

The same subgroup definitions mentioned in Section 4.1.2.7 will be applied.

For subgroup analyses of the change in SAGED at Week 24 endpoint, for each of the subgroup factors in turn, a separate ANCOVA model will be fitted using the same model

terms as used for the primary analysis (defined in Section 4.2.2.4), with additional terms for the subgroup main effect and the treatment×subgroup interaction. The results from ANCOVA will be shown in forest plot for better visualization.

It is noted that if there are low counts in some of the treatment by subgroup response groups, the ANCOVA regression models may not be reliably estimable, in which case data will be presented descriptively without formal analysis.

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.

4.2.3 Key Secondary Endpoint – Percent Change from Baseline in Tissue Eosinophils at Week 24

4.2.3.1 Definition

To support the primary histological response rate endpoint, further summaries and analysis of tissue eosinophil counts from centrally-read biopsies will be performed. This will include a key secondary endpoint analysis of percent change from baseline in eosinophils/hpf.

Eight gastric biopsies are targeted for the collection from separate areas of the gastric antrum (four samples) and the gastric corpus (four samples); four duodenal biopsies, with at least two from the second part of the duodenum.

4.2.3.2 Derivation

The highest count from eight gastric biopsies will be used for patients with gastric disease at baseline. The highest count from four duodenal biopsies will be used for patients with only duodenal disease at baseline. Baseline gastric disease is defined by the baseline gastric biopsy results with ≥30 eosinophils/hpf in at least 5 hpfs, baseline duodenal disease is defined by the baseline duodenal biopsy results with ≥30 eosinophils/hpf in at least 3 hpfs.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

The tissue eosinophils at Week 24 will include data collected at the Week 24 visits including a window of +/-2 weeks within DB period.

4.2.3.3 Handling of Dropouts and Missing Data

If the tissue eosinophils information is not available at Baseline or Week 24, but the subject has not withdrawn from randomised treatment or had a rescue or restricted therapy which would be handled with WOCF, the subject will be excluded from the analysis.

4.2.3.4 Primary Analysis of Key Secondary Endpoint

The percent change from baseline in eosinophils/hpf at Week 24 will be analyzed by using ANCOVA models with change in eosinophils as the dependent variable and baseline eosinophils along with treatment, region and baseline steroid use as explanatory variables.

4.2.3.5 Sensitivity Analyses of the Key Secondary Endpoint

The sensitivity analyses mentioned in SAGED score section (Section 4.2.2.5) may be performed on secondary endpoints if the amount of missing data warrants further exploration.

4.2.4 Key Secondary Endpoint – Proportion of Patients Who Achieve Treatment Response at Week 24

4.2.4.1 Definition

This key secondary endpoint is the proportion of patients who achieve treatment response, defined as tissue remission (same histological response criteria defined in Section 4.2.1.1) and an improvement in symptoms (change from baseline in SAGED score calculated within each 14-day period as defined in Table 4, threshold is to be confirmed with psychometric analyses of Part A data) at Week 24.

4.2.4.2 Derivation

The composite approach for responder analyses mentioned in Section 4.2 will be applied.

4.2.4.3 Handling of Dropouts and Missing Data

Patients with not enough data at Week 24 to be determined as treatment responder will be considered as non-responders .

4.2.4.4 Primary Analysis of Key Secondary Endpoint

The proportion of patients who achieve treatment response at Week 24 will be compared between benralizumab and placebo using CMH test stratified by location of disease (EG with or without duodenal involvement vs duodenal disease alone), region and baseline steroid use.

4.2.4.5 Sensitivity Analyses of the Key Secondary Endpoint

The sensitivity analyses mentioned in histological response rate (Section 4.2.1.5) may be performed on secondary endpoints if the amount of missing data warrants further exploration.

4.2.5 Key Secondary Endpoint – Proportion of Patients with No Rescue Corticosteroid Use up to Week 24

4.2.5.1 Definition

This key secondary endpoint is the proportion of patients who did not use any rescue corticosteroid up to Week 24.

4.2.5.2 Derivation

Medications that are considered to have a potentially meaningful impact on EG/EGE outcomes will be considered as intercurrent events. If any intercurrent events medications was used during the first 24 weeks, then the subject will be considered as rescue corticosteroid used during the first 24 weeks.

4.2.5.3 Handling of Dropouts and Missing Data

Subjects who discontinued during the first 24 weeks will be considered as rescue corticosteroid used during the first 24 weeks.

4.2.5.4 Primary Analysis of Key Secondary Endpoint

The proportion of patients who did not use any rescue corticosteroid up to Week 24 will be compared between benralizumab and placebo using CMH test stratified by location of disease (EG with or without duodenal involvement vs duodenal disease alone), region and baseline steroid use.

The number and percentage of patients with no rescue corticosteroid use at Week 24 and Week 52 will also be summarized descriptively by randomised treatment group.

4.2.5.5 Sensitivity Analyses of the Key Secondary Endpoint

A sensitivity analysis to the CMH test may be performed using a logistic regression with the same covariates.

Sensitivity analysis may also be performed with CMH test using a treatment policy approach (i.e., regardless of restricted medication or treatment discontinuation during the study, just consider actual rescue corticosteroid use).

4.2.6 Secondary Endpoint – Change from Baseline in Proportion of Vomiting-free Days at Week 24

4.2.6.1 Definition

Proportion of vomiting-free days is defined as the number of days in which the patient dose not report vomiting in the SAGED questionnaire in each 28-day period, divided by 28.

4.2.6.2 Derivation

A minimum of 8 days out of 14 days is required for both 14-day windows (defined in Table 4) in a 28-day period. The number of vomiting-free days in that period will be scaled up over the 28 days using the proportion of the available days (e.g. 16 vomiting-free days out

of 20 days with data available in a period is 80% vomiting-free days which will be reported as 22.4 days vomiting-free for that 28-day period).

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.6.3 Handling of Dropouts and Missing Data

If there are less than 8 days (more than 6 days of missing data) out of a 14-day window, this 14-day will be set to missing. In a 28-day period, if one of the 14-day window is missing (the other one is either missing or not), the whole 28-day period will be set to missing. This matches the missing data approach for SAGED score (see Section 4.2.2.3).

4.2.6.4 Primary Analysis of Secondary Endpoint

The change from baseline in proportion of vomiting-free days at Week 24 will be analyzed using an ANCOVA model. Please follow the ANCOVA model as described for the analysis of the SAGED score endpoint (see Section 4.2.2.4 for details). LS mean change from baseline in proportion of vomiting-free days over time up to Week 24 will be plotted by treatment with 95% CIs.

In addition, frequency of vomiting episodes and changes in these frequencies within 28 days periods during the study will be summarized descriptively.

4.2.7 Secondary Endpoint – Change from Baseline in Proportion of Diarrhea-free Days at Week 24

4.2.7.1 Definition

Daily diarrhea information will be obtained from Bristol Stool Form Scale.

The Bristol Stool Form Scale is a frequently used diagnostic tool designed to evaluate human feces based on the shape and consistency of the stool (Lewis and Heaton 1997). There a 7 stool types, which are dependent upon the amount of time in which it spends in the colon:

1. Separate hard lumps, like nuts (hard to pass)
2. Sausage-shaped, but lumpy
3. Like sausage, but with cracks on its surface
4. Like a sausage or snake, smooth and soft
5. Soft blobs with clear cut edges (passed easily)
6. Fluffy pieces with ragged edges, a mushy stool

7. Watery, no solid pieces.

Types 3, and 4, and 5 are considered optimal. Types 1 and 2 indicate constipation, and types 6 and 7 indicate diarrhea.

Participants will be asked every evening to classify the stool type of each bowel movement they have had during the past 24 hours. The Bristol Stool Form Scale will be completed by the participant starting on the evening from V1 (Week -8 to -4) until V21 (Week 76).

4.2.7.2 Derivation

Proportion of diarrhea-free days is calculated as the number of days in which the patient does not report any stools of type 6 or 7 in the Bristol Stool Form Scale in each 28 day period, divided by 28.
Same derivation rule of vomiting-free days (see Section 4.2.6.2) will be applied for diarrhea-free days.

4.2.7.3 Handling of Dropouts and Missing Data

Same missing rule of vomiting-free days (see Section 4.2.6.3) will be applied for diarrhea-free days.

4.2.7.4 Primary Analysis of Secondary Endpoint

The change from baseline in proportion of diarrhea-free days at Week 24 will be analyzed using an ANCOVA model. Please follow the ANCOVA model as described for the analysis of the SAGED score endpoint (see Section 4.2.2.4 for details). LS mean change from baseline in proportion of diarrhea-free days over time up to Week 24 will be plotted by treatment with 95% CIs.

In addition, frequency of diarrhea episodes and changes in these frequencies within 28 days periods during the study will be summarized descriptively.

4.2.8 Secondary Endpoint – Change from Baseline in Proportion of Both Diarrhea and Constipation-free Days at Week 24

4.2.8.1 Definition

Daily diarrhea and constipation information will be obtained from Bristol Stool Form Scale (Section 4.2.7.1).

4.2.8.2 Derivation

Diarrhea-free days are defined in Section 4.2.7.2. Constipation-free days are defined as the number of days in which the patient does not report any stools of type 1 or 2 in the Bristol Stool Form Scale in each 28-day period, divided by 28.

The same derivation rule as used for vomiting-free days (see Section 4.2.6.2) will be applied for both diarrhea and constipation-free days.

4.2.8.3 Handling of Dropouts and Missing Data

Same missing rule of vomiting-free days (see Section 4.2.6.3) will be applied for both diarrhea and constipation-free days.

4.2.8.4 Primary Analysis of Secondary Endpoint

The change from baseline in proportion of both diarrhea and constipation-free days at Week 24 will be analyzed by using ANCOVA models. Please follow the ANCOVA model as described for the analysis of the SAGED score endpoint (see Section 4.2.2.4 for details). LS mean change from baseline in proportion of both diarrhea and constipation-free days over time up to Week 24 will be plotted by treatment with 95% CIs.

4.2.9 Secondary Endpoint – Time to Improvement in SAGED Score

4.2.9.1 Definition

This endpoint is the time to clinically meaningful improvement in the SAGED score. The threshold for meaningful improvement in SAGED score will be defined in psychometric analyses based on data from Part A of the study.

4.2.9.2 Derivation

SAGED score is calculated in 14-day windows as defined in Table 4. Time to improvement in SAGED score will be derived as below:

last date of the 14-day window when SAGED score first reached the improvement threshold – date of randomisation + 1

If there is no improvement during the first 24 weeks, patients will be censored at Week 24 (Day 169). Patients who have intercurrent events of rescue use, restricted medication use, or randomized treatment discontinuation during the first 24 weeks, will be censored at Week 24 (Day 169).

4.2.9.3 Handling of Dropouts and Missing Data

Same rules for SAGED score calculation mentioned in Section 4.2.2.3 will be applied.

4.2.9.4 Primary Analysis of Secondary Endpoint

The time to improvement in SAGED score will be compared between treatments using a Cox proportional hazard model with baseline SAGED score, location of disease (EG vs duodenal alone), region and baseline steroid use as covariates. Time to improvement in SAGED will be displayed graphically using the Kaplan-Meier estimate by treatment.

In addition, descriptive summary of the time to improvement in SAGED score will be shown for the patients who have an improvement.

4.2.10 Secondary Endpoint – Change from Baseline in PROMIS Fatigue 7a score at Week 24

4.2.10.1 Definition

The Patient-Reported Outcome Measurement Information System (PROMIS) Short Form – Fatigue 7a consists of 7 questions assessing the severity of the patient's fatigue-related symptoms and the impact of these symptoms on health-related quality of life. The questions have a 7-day recall period and the response options are on a 5-point Likert scale ranging from 1 (never) to 5 (always).

The PROMIS Fatigue 7a score will be collected at V1 (Week -8 to -4), V2 (Week 0) and every 4 weeks after V2 until V21 (Week 76).

4.2.10.2 Derivation

The standardized T-score (range 0-100) is obtained by rescaling the raw total score using the score conversion table. The mean T-score for the reference population of US adults is 50; the standard deviation of the reference population is 10. Higher T-scores represent greater fatigue.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.10.3 Handling of Dropouts and Missing Data

If the standardized T score is not available at Baseline or Week 24, the subject will be excluded from the analysis.

4.2.10.4 Primary Analysis of Secondary Endpoint

The change from baseline in PROMIS Fatigue 7a score at Week 24 will be analyzed by using ANCOVA models. Please follow the ANCOVA model as described for the analysis of the SAGED score endpoint (see Section 4.2.2.4 for details). LS mean change from baseline in PROMIS Fatigue 7a score over time up to Week 24 will be plotted by treatment with 95% CIs.

4.2.11 Secondary Endpoint – Change from Baseline in PAGI-SYM score at Week 24

4.2.11.1 Definition

The Patient Assessment of Gastrointestinal Disorders Symptom Severity Index (PAGI-SYM) measures symptom severity in patients with upper gastrointestinal disorders. The standard recall version used in this study asks the patient to report the severity of symptoms over the past two weeks.

The instrument is composed of 20 items, which form 6 sub-scales: heartburn/regurgitation (7 items), nausea/vomiting (3 items), post-prandial fullness/early satiety (4 items), bloating (2 items), upper abdominal pain (2 items), and lower abdominal pain (2 items). Subscale

scores are calculated by averaging across items comprising the subscale; scores vary from 0 (none or absent) to 5 (very severe).

The PAGI-SYM score will be collected at V1 (Week -8 to -4), V2 (Week 0) and every 4 weeks after V2 until V21 (Week 76).

4.2.11.2 Derivation

The total PAGI-SYM score is calculated by taking the mean of the subscale scores. Higher scores indicate greater symptom severity.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.11.3 Handling of Dropouts and Missing Data

If item-level data are missing, subscale scores are calculated using the mean of non-missing items. If more than 50% of items are missing, the subscale score is set to missing. If any subscale score is missing, the PAGI-SYM total score is set to missing.

4.2.11.4 Primary Analysis of Secondary Endpoint

The change from baseline in PAGI-SYM score at Week 24 will be analyzed by using ANCOVA models. Please follow the ANCOVA model as described for the analysis of the SAGED score endpoint (see Section 4.2.2.4 for details). LS mean change from baseline in PAGI-SYM score over time up to Week 24 will be plotted by treatment with 95% CIs.

4.2.12 Secondary Endpoint – Change from Baseline in SF-36v2 PCS and MCS

4.2.12.1 Definition

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

The 8 domains are: Physical Functioning, Role Limitations due to Physical Health, Bodily Pain, General Health Perceptions, Vitality, Social Functioning, Role Limitations due to Emotional Problems, and Mental Health. The component summary measures, Physical Component Summary (PCS) and Mental Component Summary (MCS), are computed from domain scores to give a broader metric of physical and mental health-related quality of life. The transformed score range for each of the 8 domains and for PCS and MCS is 0-100; higher scores indicate better health state.

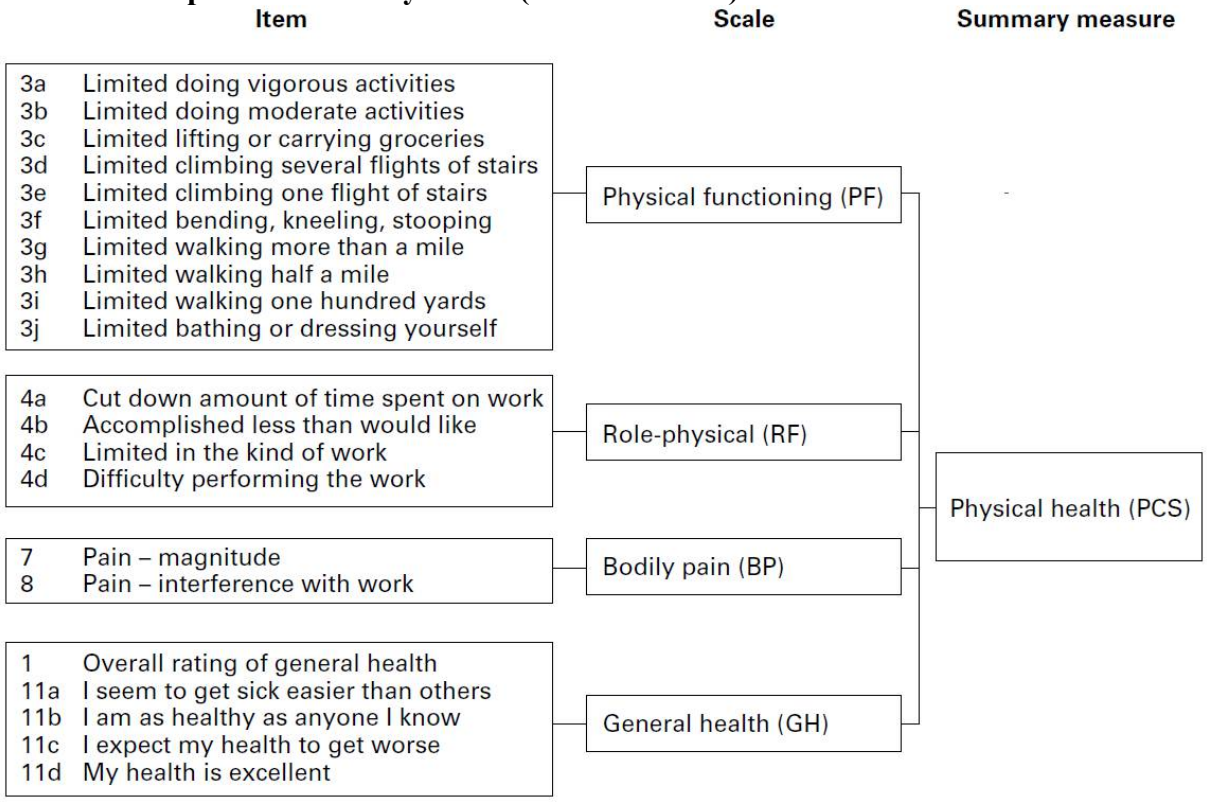
The SF-36v2 will be collected at V1 (Week -8 to -4), V2 (Week 0) and every 4 weeks after V2 until V21 (Week 76).

4.2.12.2 Derivation

Figure 3 shows the 35 questions used to compute the 8-domain profile of functional health and well-being scores. Question 2 is the remaining item referred to as the 'Health Transition' item not used to calculate domain scores and does not appear in Figure 3.

Two types of thresholds have been developed for interpretation of SF-36v2 scores. The first type is suitable for comparing group mean scores and is generally referred to as the MCID (Minimum Clinical Important Difference). The second type is suitable for interpreting change at the individual level and is referred to as the responder threshold or responder definition (Maruish ME). Threshold values for the SF-36 v2 scale and summary measures are provided in Table 6.

Figure 3 The 35 questions for the 8-domain scores and physical & mental health component summary scores (PCS and MCS)



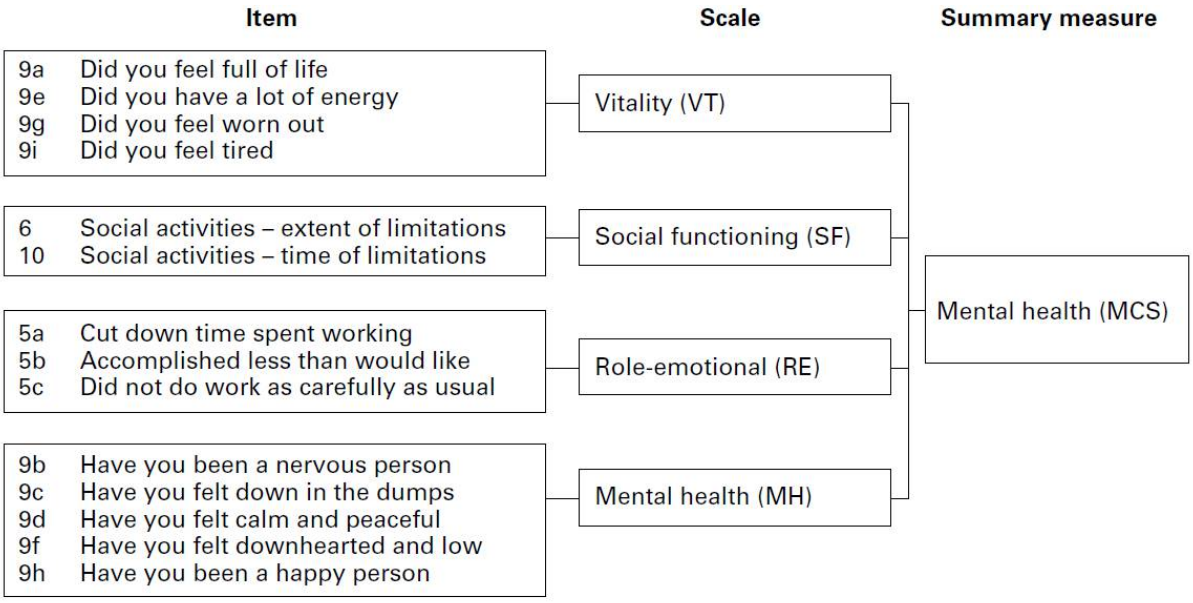


Table 6 Threshold values for the SF-36 v2 scale and summary measures

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

BP Bodily Pain; GH General Health Perceptions; MCS mental health component summary; MH Mental Health; PCS physical component summary; PF Physical Functioning; RE Emotional Problems; RP Role Limitations due to Physical Health; SF Social Functioning; VT Vitality.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.12.3 Handling of Dropouts and Missing Data

If the SF-36v2 PCS and MCS is not available at Baseline or Week 24, the subject will be excluded from the analysis.

4.2.12.4 Primary Analysis of Secondary Endpoint

The change from baseline in SF-36v2 PCS and MCS at Week 24 will be analyzed separately by using ANCOVA models. Please follow the ANCOVA model as described for the analysis of the SAGED score endpoint (see Section 4.2.2.4 for details). LS mean change from baseline in SF-36v2 PCS and MCS over time up to Week 24 will be plotted by treatment with 95% CIs.

The change from baseline in SF-36v2 PCS and MCS at Week 24 and Week 52 will also be summarized descriptively by randomised treatment group.

4.2.13 Secondary Endpoint – Change from Baseline in PAGI-QoL

4.2.13.1 Definition

The Patient Assessment of Upper Gastrointestinal Disorders-Quality of Life (PAGI-QoL), the health-related quality of life counterpart to PAGI-SYM, measures disease-specific health-related quality of life impacts in patients with upper gastrointestinal disorders. It is comprised of 30 items and five sub-scales: daily activities (10 items), clothing (2 items), diet and food habits (7 items), relationships (3 items), and psychological well-being and distress (8 items). Patients are asked to rate the impacts of their gastrointestinal symptoms over the past two weeks on a scale ranging from 0 (none of the time) to 5 (all of the time).

The PAGI-QoL score will be collected at V1 (Week -8 to -4), V2 (Week 0) and every 4 weeks after V2 until V21 (Week 76).

4.2.13.2 Derivation

The total PAGI-QoL score is calculated by averaging across subscale scores. Lower scores reflect worse health-related quality of life. Subscale scores are calculated by averaging across items comprising the subscale after reversing the scores (i.e., 5=0, 4=1, 3=2, 2=1, 1=5). If item-level data are missing, subscale scores are calculated using the mean of non-missing items. If more than 50% of items are missing, the subscale score is set to missing. If any subscale score is missing, the PAGI-QoL total score is set to missing.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.13.3 Handling of Dropouts and Missing Data

If item-level data are missing, subscale scores are calculated using the mean of non-missing items. If more than 50% of items are missing, the subscale score is set to missing. If any subscale score is missing, the PAGI-QoL total score is set to missing.

4.2.13.4 Primary Analysis of Secondary Endpoint

The change from baseline in PAGI-QoL score at Week 24 will be analyzed by using ANCOVA models. Please follow the ANCOVA model as described for the analysis of the SAGED score endpoint (see Section 4.2.2.4 for details). LS mean change from baseline in PAGI-QoL score over time up to Week 24 will be plotted by treatment with 95% CIs.

The change from baseline in PAGI-QoL score at Week 24 and Week 52 will also be summarized descriptively by randomised treatment group.

4.2.14 Exploratory Endpoint - Change from Baseline in Vomiting Severity at Week 24

4.2.14.1 Definition

The daily severity of vomiting (range: 0-10) are reported from the SAGED instrument. Vomiting severity at baseline and at Week 24 are calculated as 14-day mean scores. Each 14-day mean score will be calculated as the sum of the available daily severity of vomiting scores divided by the number of non-missing days in that period.

4.2.14.2 Derivation

A minimum of 8 days out of 14 days is required for the 14-day mean score and will follow the 14-day windows defined in Table 4.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.14.3 Handling of Dropouts and Missing Data

If there are less than 8 days (more than 6 days of missing data) out of a 14-day window, this 14-day will be set to missing.

4.2.14.4 Primary Analysis of Exploratory Endpoint

The change from baseline in vomiting severity at Week 24 will be analyzed by using ANCOVA models. Please follow the ANCOVA model described in SAGED score estimand (see Section 4.2.2.4) for details. LS mean change from baseline in vomiting severity over time up to Week 24 will be plotted by treatment with 95% CIs.

4.2.15 Exploratory Endpoint - Change from Baseline in Diarrhea Severity at Week 24

4.2.15.1 Definition

The daily severity of diarrhea (range: 0-10) are reported from the SAGED instrument. Diarrhea severity at baseline and at Week 24 are calculated the same way as mentioned for the vomiting severity endpoint (see Section 4.2.14.1).

4.2.15.2 Derivation

The same derivation rules as mentioned for the vomiting severity endpoint (see Section 4.2.14.2) will be applied.

4.2.15.3 Handling of Dropouts and Missing Data

Same missing rule mentioned in vomiting severity estimand (see Section 4.2.14.3) will be applied.

4.2.15.4 Primary Analysis of Exploratory Endpoint

The change from baseline in diarrhea severity at Week 24 will be analyzed by using ANCOVA models. Please follow the ANCOVA model described in SAGED score

estimand (see Section 4.2.2.4) for details. LS mean change from baseline in diarrhea severity over time up to Week 24 will be plotted by treatment with 95% CIs.

4.2.16 Exploratory Endpoint - Change from Baseline in EG-REFS

4.2.16.1 Definition

During endoscopy procedures at Visit 1 (Week 0), Visit 8 (Week 24), and Visit 15 (Week 52), the endoscopic manifestations of EG/EGE, including erosions/ulcerations, granularity, raised lesions, erythema, friability, fold thickness and pyloric stenosis, will be scored by the endoscopist using the Eosinophilic Gastritis Endoscopic Reference System (EG-REFS) from three areas of the stomach (antrum, body, and fundus).

The EG-REFS and duodenal findings will be centrally-read from video recordings and Investigator-read during the upper endoscopies. Centralized imaging data assessments and scoring from expert physician review will be performed for all endoscopies. Standardized methods for training Investigators on the application of these scoring systems to ensure the collection of quality data using this measure will be implemented. EG-REFS score for each location is defined based on below Table 7 as well as an overall global assessment of endoscopic gastric disease activity from 0 (Normal) to 10 (extremely severe disease) (Hirano, Collins et al. 2019, Shoda, Wen et al. 2020).

Table 7 EGREFS Grading (evaluate in 3 locations of the stomach: fundus, corpus, and antrum)

Endoscopic Abnormality ***(location of the stomach where measured)***	**Grade**
Erosion/Ulceration *(fundus, corpus, and antrum)*	0 None 1 Less than 5 erosions 2 Five or more erosions 3 Shallow/superficial ulceration(s) 4 Deep/excavated ulceration < 25% of the surface area of specified location 5 Deep/excavated ulceration 25-50% of the surface area of specified location 6 Deep/excavated ulceration > 50 % of the surface area of specified location
Granularity *(fundus, corpus, and antrum)*	0 Normal 1 Fine 2 Coarse
Raised lesion (Nodularity) *(fundus, corpus, and antrum)*	0 None 1 Mild (raised focal nodules with width greater than height) 2 Severe (raised nodules with greater height than width)
Erythema *(fundus, corpus, and antrum)*	0 None 1 Mild (pink) 2 Severe (red/hemorrhagic)
Friability/Bleeding	0 None 1 Mild (contact bleeding)

(fundus, corpus, and antrum)	2 Severe (spontaneous bleeding)
Folds *(fundus, corpus, and antrum)*	0 None 1 Thickened folds
Pyloric stenosis *(antrum only)*	0 None 1 Present (inability to pass diagnostic 8-10 mm upper endoscope)

Similarly, for evaluation of duodenal area, assessment of the visualized duodenum in total with a graded criteria and an overall assessment from 0 (Normal) to 10 (extremely severe disease).

Table 8 EoD-REFS Grading

Endoscopic Abnormality	**Grade**
Granularity	0 Normal * 1 Fine 2 Coarse
Erythema	0 None 1 Mild (pink) 2 Severe (red/hemorrhagic)
Friability/Bleeding	0 None 1 Mild (contact bleeding) 2 Severe (Spontaneous bleeding)
Ulceration	0 Normal, no erosion or ulcer 1 Erosions or pinpoint ulcerations 2 Numerous shallow ulcers with or without mucous 3 Deep or excavated ulcerations 4 Diffuse ulceration with >30% of surface area of the inspected mucosa
Denuded Patches	0 Normal/no lesions 1 Few lesions (<5) 2 Numerous lesions (>=5)
Stricture	0 Normal 1 Stricture present and can be traversed with standard adult endoscope 2 Stricture present and can NOT be traversed with standard adult endoscope

EoD: Eosinophilic Duodenal.
* normal is smooth and glistening.

The EoE-EREFS is a scoring system for assessing the presence and severity of the major endoscopic signs of Eosinophilic esophagitis (EoE), including esophageal edema, rings,

exudates, furrows, and stricture. Table 9 shows the EoE-EREFS grading. An overall assessment from 0 (Normal) to 10 (extremely severe disease) will also be performed.

Table 9 EoE-EREFS Grading

Endoscopic Abnormality	Grade
Edema	0 – Absent 1 – Loss of clarity or absence of vascular markings
Rings	0 – None 1 – Mild = subtle circumferential ridges 2 – Moderate = distinct rings that do not impair passage of a standard diagnostic adult endoscope (outer diameter 8-9.5 mm) 3 – Severe = distinct rings that do not permit passage of a diagnostic endoscope
Exudate	0 – None 1 – Mild = lesions involving <10% of the esophageal surface area 2 – Severe = lesions involving >10% of the esophageal surface area
Furrows	0 – Absent 1 – Mild = Vertical lines present without visible depth 2 – Severe = Vertical lines with mucosal depth (indentation)
Stricture	0 – Absent 1 – Present

4.2.16.2 Derivation

Reviewers may select not evaluable (NE) if the endoscopic abnormality cannot be graded due to image quality issues. An NE assessment of any category will result in an NE for the overall grading.

EG-REFS contains grading from three areas of the stomach (antrum, body, and fundus). The total EG-REFS score will be the sum of all three areas grading in the range of 0-46 (max total 15 for fundus and body and 16 for antrum).

The total score for EoD REFs for HUDSON is in the range of 0-14. The total score for EREFS for HUDSON is in the range of 0-9.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.16.3 Handling of Dropouts and Missing Data

Subjects missing REFS at Baseline or Week 24 after composite approach applied will be excluded from the analysis.

4.2.16.4 Primary Analysis of Exploratory Endpoint

The change from baseline in total EG-REFS score and overall assessment at Week 24 will be analyzed by using ANCOVA models for patients with gastric disease at baseline. Similarly, the change from baseline in EoD-REFS score and overall assessment at Week 24 will be analyzed by using ANCOVA models for patients with duodenal disease only at baseline, the change from baseline in EoE-REFS score and overall assessment at Week 24 will be analyzed by using ANCOVA models for patients with esophageal disease at baseline. The ANCOVA model will be using baseline value along with treatment, region and baseline steroid use as explanatory variables..

Baseline gastric disease is defined by the baseline gastric biopsy results with ≥30 eosinophils/hpf in at least 5 hpfs, baseline duodenal disease is defined by the baseline duodenal biopsy results with ≥30 eosinophils/hpf in at least 3 hpfs, baseline esophageal disease is defined by the baseline esophageal biopsy results with ≥15 eosinophils/hpf in at least 1 hpf. Baseline esophageal disease is defined by the baseline esophageal biopsy results with >=15 eos/hpf on at least 1 esophageal level.

The change from baseline in EG-REFS at Week 24 may also be explored using overall population.

The change from baseline in EG-REFS at Week 24 and Week 52 will be summarized descriptively by randomised treatment group and by patients with gastric disease at baseline as well as overall population.

4.2.17 Exploratory Endpoint - Change from Baseline in DSQ at Week 24

4.2.17.1 Definition

The Dysphagia Symptom Questionnaire (DSQ) is a PRO measure validated for patients age 12 and older with dysphagia related to eosinophilic esophagitis (Dellon et al 2013b). The presence and severity of dysphagia symptoms in the past 24 hours are captured in a 4-item questionnaire.

Questions 1 and 2 utilize yes/no response capture if the patient consumed solid food that day (yes/no; unscored) and instances of food going down slowly or becoming stuck in the throat or chest (scored 0 for no and 2 for yes). Question 3 asks about the severity of dysphagia, based on actions the patient took to relieve the dysphagia at its worst point during the day. It ranges from 0 (dysphagia cleared up on its own) to 4 (patient sought medical attention for dysphagia). Question 4 asks the patient to report the worst pain experienced while swallowing food over the past 24 hours (no pain [0] to very severe pain [4]).

DSQ will be completed by the participant every evening starting on the evening from V1 (Week -8 to -4) until V21 (Week 76).

4.2.17.2 Derivation

The DSQ score is comprised of daily values captured over a 14-day period and has a range of 0-84 with higher scores indicating more frequent and/or severe dysphagia. The DSQ Score is calculated as follows:

DSQ score = (Sum of points from questions 2 + 3 in the daily DSQ) × 14 days / (Number of diaries reported with non-missing data).

Daily DSQ scores are calculated as the sum of Q2 and Q3 response values. The daily DSQ score range is 0-6. Higher values indicate worse dysphagia. Daily DSQ score will be set to missing if not recorded or if Q1="No".

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.17.3 Handling of Dropouts and Missing Data

Same missing rule for SAGED score mentioned in Section 4.2.2.3 will be applied.

4.2.17.4 Primary Analysis of Exploratory Endpoint

The change from baseline in DSQ at Week 24 will be analyzed by using ANCOVA models. Please follow the ANCOVA model as described for the analysis of the SAGED score endpoint (see Section 4.2.2.4 for details). LS mean change from baseline in DSQ over time up to Week 24 will be plotted by treatment with 95% CIs.

4.2.17.5 Supplementary Analysis of Exploratory Endpoint

The change from baseline in DSQ at Week 24 will be analyzed by using ANCOVA models for patients with EoE at baseline. Please follow the ANCOVA model described in SAGED score estimand (see Section 4.2.2.4) for details. Baseline disease type will be determined based on the baseline biopsy results, criteria will be provided in the future.

4.2.18 Exploratory Endpoint - Change from Baseline in PGI-S at Week 24

4.2.18.1 Definition

Patient Global Impression of Severity (PGI-S) is a single item designed to capture the participant's perception of overall symptom severity over the past 14 days using a 6-point categorical response scale (no symptoms to very severe symptoms).

Response option	Score
No symptoms	1
Very mild	2
Mild	3

Moderate	4
Severe	5
Very severe	6

PGI-S will be collected at V1 (Week -8 to -4), V2 (Week 0) and every 2 weeks after V2 until V21 (Week 76).

4.2.18.2 Derivation

The PGI-S collected within Baseline and Week 24 window defined in Table 4 will be used for change from baseline calculation.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.18.3 Handling of Dropouts and Missing Data

Subjects with no PGI-S collected within Baseline and Week 24 window defined in Table 4 will be excluded from the analysis.

4.2.18.4 Primary Analysis of Exploratory Endpoint

The change from baseline in PGI-S at Week 24 will be summarized descriptively by treatment group.

4.2.19 Exploratory Endpoint - PGI-C at Week 24

4.2.19.1 Definition

Patient Global Impression of Change (PGI-C) is a single item assessment to capture the participant's perception of change in health status. The participant is asked to report the degree to which their health status has changed since entering the treatment period using a 7-point scale:

Response option	Score
Much better	1
Moderately better	2
A little better	3
About the same / No Change	4
A little worse	5
Moderately Worse	6
Much Worse	7

PGI-C will be collected at every 2 weeks after V2 (Week -8 to -4) until V21 (Week 76).

4.2.19.2 Derivation

The PGI-C collected within Week 24 window defined in Table 4 will be used.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.19.3 Handling of Dropouts and Missing Data

Subjects with no PGI-C collected within Week 24 window defined in Table 4 will be excluded from the analysis.

4.2.19.4 Primary Analysis of Exploratory Endpoint

PGI-C value at Week 24 will be summarized descriptively by treatment group.

4.2.20 Exploratory Endpoint - Change from Baseline in CGI-S at Week 24

4.2.20.1 Definition

Clinical Global Impression of Severity (CGI-S) is a single item designed to capture the clinician's perception of overall symptom severity at the time of the visit using a 6-point categorical response scale (no symptoms to very severe symptoms), same as PGI-S (See Section 4.2.18.1).

CGI-S will be collected at V2 (Week 0) and V8 (Week 24).

4.2.20.2 Derivation

The CGI-S collected at Week 0 and Week 24 will directly be used for change from baseline calculation.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.20.3 Handling of Dropouts and Missing Data

Subjects with no CGI-S collected in either Week 0 or Week 24 will be excluded from the analysis.

4.2.20.4 Primary Analysis of Exploratory Endpoint

The change from baseline in CGI-S at Week 24 will be summarized descriptively by treatment group.

4.2.21 Exploratory Endpoint - CGI-C at Week 24

4.2.21.1 Definition

Clinical Global Impression of Change (CGI-C) is a single item assessment to capture the clinician's perception of change in health status. The clinician is asked to report the degree to which the participant' health status has changed since entering the treatment period using a 7-point scale, same as PGI-C (see Section 4.2.19.1).

CGI-C will be collected at every 4 weeks beginning V3 (Week 4) until V21 (Week 76).

4.2.21.2 Derivation

The CGI-C collected within Week 24 window defined in Table 3 will be used.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.21.3 Handling of Dropouts and Missing Data

Subjects with no CGI-C collected within Week 24 window defined in Table 3 will be excluded from the analysis.

4.2.21.4 Primary Analysis of Exploratory Endpoint

CGI-C value at Week 24 will be summarized descriptively by treatment group.

4.2.22 Exploratory Endpoint – Proportion of Patients Who Achieve Treatment Response at Week 52

4.2.22.1 Definition

Definition of treatment responder is defined in Section 4.2.4.1.

4.2.22.2 Derivation

Derivation rule for treatment responder is mentioned in Section 4.2.4.2.

The composite approach for responder analyses mentioned in Section 4.2 will be applied.

4.2.22.3 Handling of Dropouts and Missing Data

The missing rule for treatment responder is mentioned in Section 4.2.4.3.

4.2.22.4 Primary Analysis of Exploratory Endpoint

Number and proportion of patients who achieve treatment response at Week 52 will be descriptively summarized.

In addition, the number of patients in histological response, patients with tissue eosinophil counts returning to Baseline level or worse (who may be considered as having relapsed) and patients in between these thresholds at Week 24 and Week 52 will be summarized by shift table. Similarly, the number of patients in symptomatic response, patients with symptomatic deterioration and patients in between these thresholds at Week 24 and Week 52 will be summarized by shift table.

The relationship between any changes in symptom scores and change in tissue eosinophils at week 52 will be explored. A scatterplot of changes in SAGED scores and change in tissue eosinophils at week 52 will be evaluated with locally weighted scatter plot smoothing curves (LOESS) for all subjects.

4.2.23 Exploratory Endpoint – Use of Concomitant Medications and Dietary Restrictions to Manage EG/EGE

4.2.23.1 Definition

Background medications for EG/EGE includes systemic and topical ingested or swallowed corticosteroids, PPIs and steroid treatments used for asthma or allergies that are inhaled or administered intranasally.

Dietary restrictions are any changes in diet related to the management of EG/EGE.

4.2.23.2 Derivation

Information will be directly obtained from collected data.

4.2.23.3 Handling of Dropouts and Missing Data

Not Applicable.

4.2.23.4 Primary Analysis of Exploratory Endpoint

The number and percentage of patients who reduced the background medications and/or relax dietary restrictions to manage EG/EGE at Week 24 and Week 52 will be summarized descriptively by the randomised treatment received in Part A/B.

4.2.24 Exploratory Endpoint – Proportion of Patients with Relevant Healthcare Resource Utilization and Number of Events by HRU Type

4.2.24.1 Definition

EG/EGE-related and other GI-disease complications-related healthcare resource utilization (HRU) data will be collected in the eCRF by the Investigator and study-site personnel for all participants throughout the study. Protocol-mandated procedures, tests, and encounters are excluded.

The Investigator retrospectively collects any EG/EGE-related and other GI disease complications-related HRU. At baseline, HRU will be collected with a 6-month recall period. At all subsequent visits, any EG/EGE-related and other GI disease complications-related HRU information will be collected with a recall period of 'since last HRU assessment during the scheduled visit'.

The data collected may be used to conduct exploratory economic analyses and will 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 bywards [e.g., 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

HRU will be collected at V2 (Week 0) and every 4 weeks after V2 until V21 (Week 76), then every 12 weeks after V21.

4.2.24.2 Derivation

Information will be directly obtained from collected data.

4.2.24.3 Handling of Dropouts and Missing Data

Not Applicable.

4.2.24.4 Primary Analysis of Exploratory Endpoint

Proportion of patients with relevant HRU and number of events by HRU type (including but not limited to hospitalizations, length of hospital stay, office visits, emergency room visits, tests and procedures) will be summarized by randomised treatment and visit.

4.2.25 Exploratory Endpoint – WPAI+CIQ Score

4.2.25.1 Definition

The Work Productivity and Activity Impairment Questionnaire plus Classroom Impairment Questions Version 2 (WPAI+CIQ) consists of questions about how health and health-related issues impact the ability to work, attend classes, and perform regular daily activities. The questionnaire relates to the previous 7 days. The WPAI+CIQ will be used to measure self-reported productivity loss.

This study uses the version of the instrument designed to measure the impact of a specific health problem (SHP). All instances of "problem" in the source text are replaced with "eosinophilic gastritis and/or eosinophilic gastroenteritis" to specify that the participant should respond with EG/EGE symptoms in mind.

There are a maximum of 10 questions and a minimum of 3 questions that will be completed by subjects.

1 = currently employed (yes/no)

2 = hours missed work due to health problems

3 = hours missed work due to other reasons

4 = hours actually worked

5 = degree health affected productivity while working (0-10 scale, with 0 meaning no effect)

6 = attends class in an academic setting (yes/no)

7 = hours missed class due to health problems

8 = hours actually attended class

9 = degree health affected productivity while attending class (0-10 scale, with 0 meaning no effect)

10 = degree health affected regular activities (other than work or class) (0-10 scale, with 0 meaning no effect)

If the answer to question 1 is 'No, not currently employed', then the subject should skip to question 6. If the answer to question 6 is 'No, not currently attending class', then the subject should skip to question 10.

The WPAI+CIQ provides 4 types of scores: absenteeism (work or class time missed), presenteeism (impairment at work or class/reduced on-the-job effectiveness), work productivity loss (overall work or class impairment/absenteeism plus presenteeism), and activity impairment. WPAI+CIQ outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity.

The WPAI+CIQ score will be collected at V1 (Week -8 to -4), V2 (Week 0) and every 4 weeks after V2 until V21 (Week 76).

4.2.25.2 Derivation

For the work-related questions, the following calculations should be used to create the outcomes of interest:

- Number of work hours missed = Q2
- Absenteeism = Q2/(Q2+Q4)
- Presenteeism = Q5/10
- Work Productivity Loss = Q2/(Q2+Q4) + [(1-Q2/(Q2+Q4))*(Q5/10)]

For the class-related questions, the following calculations should be used to create the outcomes of interest:

- Number of class hours missed = Q7

- Absenteeism = Q7/(Q7+Q8)
- Presenteeism = Q9/10
- Class Productivity Loss = Q7/(Q7+Q8) + [(1-Q7/(Q7+Q8))*(Q9/10)]

Additionally, Activity Impairment = Q10/10.

The composite approach for continuous analyses mentioned in Section 4.2 will be applied.

4.2.25.3 Handling of Dropouts and Missing Data

Not Applicable.

4.2.25.4 Primary Analysis of Exploratory Endpoint

The WPAI+CIQ score of each type and the change from baseline will be summarized descriptively by visit and randomised treatment group.

4.2.26 Exploratory Endpoint – PGI-BR Scores

4.2.26.1 Definition

The Patient Global Impression of Benefit-Risk (PGI-BR) is a 5-item questionnaire assessing the participant's perception of the overall benefits and risks of treatment. The 5 items assess: overall trial experience, efficacy, side effects, convenience and overall assessment of the benefits and harms of treatment. Items are rated on 5- or 6-point verbal rating or Likert-type scales.

The PGI-BR assessment will be completed by participants Visits 8 (Week 24) and 15 (Week 52) and after discontinuation if applicable as described in Table 4.

4.2.26.2 Derivation

The PGI-BR items are scored as follows:

Trial experience - "Overall, how would you rate your experience participating in the clinical trial?" (very negative, somewhat negative, neither negative nor positive, somewhat positive, very positive)". Scored from 1 (very negative) to 5 (very positive).

Efficacy – "How much has the study medication helped your condition? (not at all, a little bit, somewhat, quite a bit, very much)". Scored from 1 (not at all) to 5 (very much).

Side Effects – "How bad have the side effects of the study medication been overall? (did not experience any side effects, not at all bad, slightly bad, moderately bad, very bad, extremely bad)". Scored from 5 (did not experience) to 1 (extremely bad).

Convenience – "How convenient or inconvenient has the study medication been? (very inconvenient/somewhat inconvenient/neither convenient nor inconvenient/somewhat convenient/very convenient)". Scored from 1 (very inconvenient) to 5 (very convenient).

Benefit/risk assessment – "How would you compare the positive and negative things about the study medication? (negatives far greater than positives/negatives somewhat greater than positives/about equal/positives somewhat greater than negatives/positives far greater than negatives)". Scored from 1 (negatives far greater than positives) to 5 (positives far greater than negatives).

4.2.26.3 Handling of Dropouts and Missing Data

Not Applicable.

4.2.26.4 Primary Analysis of Exploratory Endpoint

The PGI-BR score of each category and the change from baseline will be summarized descriptively by visit and randomised treatment group.

4.3 Pharmacodynamic Endpoint(s)

Pharmacodynamic exploratory analyses will be described in a separate exploratory analysis plan finalized before database lock.

4.4 Pharmacokinetics

Blood samples (processed to serum) for pharmacokinetic assessments will be collected from all subjects at baseline prior to first benralizumab administration at V1 (Week -8 to -4), V4 (Weeks 8), V6(Week 16), V8 (Week 24), V15 (Week 52), V21 (Week 76) and every 12 weeks after. Blood samples will be collected before benralizumab administrations. Results below the lower limit of quantification (BLQ) will be set to LLOQ/2 for analysis and will be listed as <LLOQ.

Serum benralizumab concentration over time will be descriptively summarized for DB period (Part A and Part B) as well as OLE period (Part C). Serum benralizumab concentration will also be summarized by ADA status, see ADA and PK in Section 4.5 for details.

Other pharmacokinetic analyses will be described in a separate exploratory analysis plan finalized before database lock.

4.5 Immunogenicity

Anti-drug antibodies to benralizumab will be summarized using descriptive statistics at each visit by study part based on the safety analysis set. The impact of anti-drug antibodies on the pharmacokinetics of benralizumab and eosinophil levels may be

assessed. The potential association of anti-drug antibodies with safety and efficacy may be evaluated.

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, patients with a missing baseline ADA assessment will be assumed to be ADA negative at baseline as a conservative approach to ensure that all subjects 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 subject, the following ADA responses will be evaluated over the entire on-study period through EOT:

- Subjects 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 subjects in a population is known as ADA prevalence.
- Subjects who are ADA negative at all assessments, including baseline and post-baseline (also generally referred to as ADA negative).
- Subjects who are ADA positive at baseline only.
- Subjects 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 (i.e., ≥ 4-fold increase) at ≥1 post-baseline timepoint. This is called treatment-boosted ADA positive.
- Subjects 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.

- Subjects 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.

- Subjects 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 subject 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 patients 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 subject, 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 using derived visit windows (refer to Section 3.3.2 for detailed definition of visit windows). In the event a patient 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 patients 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 subjects who are ADA positive at each visit will be provided.

Key patient information will be listed for patients 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.

ADA response and demographics/patient characteristics

Demographic and baseline patient characteristics will be presented for the following ADA response categories of patients: treatment-emergent ADA positive, ADA negative, and ADA persistently positive.

ADA and eosinophil levels

Eosinophil levels will be summarised by visit for the following ADA response categories of patients: treatment-emergent ADA positive, ADA negative, ADA persistently positive and ADA positive with titer > median of maximum. A line plot of eosinophil levels by visit and ADA status will also be presented.

ADA and efficacy

The effects of ADA on the primary and key secondary endpoints calculated through EOT will be evaluated through summary statistics by treatment group and ADA status (ADA positive, treatment-emergent ADA positive, ADA negative, ADA persistently positive, and ADA positive with titre > median of maximum titre). Due to the expected small number of ADA positive subjects in the placebo group, no formal statistical comparisons of benralizumab versus placebo by ADA status (positive/negative) are planned.

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, ADA persistently positive and ADA positive with titre > median of maximum titre). The on-treatment and on-study periods are as defined in Section 4.2.6.1. AEs and SAEs will also be evaluated by causality as assessed by the investigator. The potential impact of ADA on hypersensitivity will also be assessed.

ADA and PK

Benralizumab serum concentrations will be summarised by visit and ADA status (treatment-emergent ADA positive, ADA negative, ADA persistently positive and ADA positive with titre > median of maximum titre) for patients in the PK analysis set.

4.6 Safety Analyses

All safety variables will be summarized using the safety analysis set and data presented according to actual treatment received.

The first analysis of safety data will include only data from the double-blind, placebo-controlled first 24 weeks of the study (DB periods from Part A and B, respectively). Patients will be analyzed according to the treatment they actually received (benralizumab or placebo).

Safety data from the open-label benralizumab part of the study (Part C) will primarily be reported separately from Parts A and B. In addition, a key subset of summaries may also be repeated integrating data across the entire study period (i.e. Part A together with Part C

and Part B together with Part C, and also all parts of the study together once part C has completed).

Safety data will be presented using descriptive statistics unless otherwise specified.

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

No safety data will be imputed. The handling of partial/missing dates for AEs and prior/concomitant medications is detailed in Appendix 7.2. Duration of AEs and prior/concomitant medications will not be calculated using imputed dates and will instead be set to missing.

4.6.1 Exposure

See Section 4.1.9 for details.

4.6.2 Adverse Events

4.6.2.1 Definitions and Derivations

Adverse events experienced by the patients will be collected throughout the entire study and will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA) per the Data Management Plan.

Adverse event 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 onset day before day of first dose of study treatment.
- AEs in the on-treatment period are defined as those with onset day between day of first dose of study treatment and scheduled end of treatment (EOT) visit for patients who complete study treatment or investigational product discontinuation visit (IPD) for patients who prematurely discontinue study treatment, inclusive. In the event that the EOT or IPD visit is beyond the protocol-defined visit window, AEs with onset after the last dose of study treatment date +28 days +3 days (visit window) will be excluded from the on-treatment period and instead assigned to the post-treatment period.
- AEs in the on-study period are defined as those with onset between day of first dose of study treatment and the day of the scheduled follow-up visit, inclusive.
- AEs in the post-treatment period are defined as those with onset after the on-treatment period defined above.

- On-study AEs in the DB period will be defined as those with onset date between day of first dose of study DB treatment and the day prior to the first dose of OLE period (up to and including day of V9 for patients who do not roll over to OLE).

- On-study AEs in the OLE period are defined as those with onset date on or after the day of the first dose of OLE treatment and up to EOT.

For instances where a patient attends the safety follow-up visit only, but does not attend an earlier IPD visit or EOT visit, adverse events occurring on or after the day of first dose of study treatment and on or before the last dose of study medication + 28 days will be assigned to the on-treatment period, while AEs with onset date after this time will be assigned to the post-treatment period.

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

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

4.6.2.2 Presentation

AEs will be summarized separately for the on-treatment, on-study, and post-treatment periods, as defined in Section 4.6.2.1. Additionally, only serious adverse events (SAEs) in the pre-treatment period (with start date prior to the first dose of IP) will be listed. All AEs will be listed for each subject. Summaries by SOC/PT will be presented by treatment group and will be exposure-adjusted to account for the variability in follow-up periods beyond 24 weeks.

The rate of AEs per person-years at risk will be calculated as (number of patients reporting the AE)/(total IP exposure with patients at risk of AE) for on-treatment and on-study periods. The post-treatment AEs will be listed in listings. The total period at risk for each patient will be the duration of the on-treatment, post-treatment and on-study periods as defined in Section 4.6.2.1. Rates will be expressed in terms of events per 100 patient-years.

An overall summary table will be produced showing the number, percentage, and exposure-adjusted rate of patients with at least 1 AE in any of the following categories: AEs, serious adverse events (SAEs), AEs with outcome of death, and AEs leading to discontinuation of investigational product (DAEs).

AEs, AEs with outcome of death, SAEs and DAEs will be summarised by System Organ Class (SOC) and Preferred Term (PT) assigned to the event by MedDRA. For each PT, the

number, percentage and exposure-adjusted rate of patients reporting at least one occurrence will be presented (ie, multiple occurrences of an AE for a patient will only be counted once). SOC and PT will be sorted by international order.

A summary of the most common AEs will be presented by PT. Additionally, a summary of non-serious AEs occurring in >5% of patients in any treatment group will be presented by PT. AEs and SAEs causing discontinuation of the study treatment and SAEs causing discontinuation from the study will also be summarised by PT.

AEs and SAEs will be summarised by preferred term and investigator's causality assessment (related vs. not related) and maximum intensity. If a patient reports multiple occurrences of the same AE within the same study period, the maximum intensity will be taken as the highest recorded maximum intensity (the order being mild, moderate, and severe).

Other clinically relevant adverse events will include but may not be limited to injection site reactions, hypersensitivity events, serious infections, and malignancies. Adverse events of injection site reactions (high level terms of administration and injection site reactions), hypersensitivity (standardized MedDRA query of hypersensitivity in narrow), serious infections (serious events within the SOC Infections and Infestations), and malignancy (standardized MedDRA query of malignant tumours) will be summarised by preferred term. The summary of injection site reactions will be summarised by injection site location and number of IP administrations. The summary of AEs of hypersensitivity will be presented overall and repeated for events causally related to IP as assessed by the investigator.

Plot of frequencies and risk differences (forest plots) between treatment arms will be presented for the most common adverse events and other specific events of interest. Estimates and confidence intervals based on the Miettinen Nurminen (M-N) method will also be presented for the most common adverse events and any other specific events of interest included in the structured assessment of benefit risk.

4.6.3 Clinical Laboratory, Blood Sample

4.6.3.1 Definitions and Derivations

Blood samples for determination of clinical chemistry and haematology parameters will be taken at the times detailed in the CSP, and will be assessed in a central laboratory.

In summaries, listings and figures, lab results and normal ranges will be presented in the International System (SI) unit. Eosinophils data will be presented in both SI and conventional units (eos/HPF) in summaries.

Changes in haematology and clinical chemistry variables between baseline and each postbaseline assessment will be calculated. Please refer to Section 3.3.1.2 for baseline definition. The change from baseline is defined as the post-baseline visit value minus the baseline visit value. There will be no imputation for missing values. For values recorded with a leading greater than or less than ('>', '<') symbol, the reported numeric value will be used for analysis and the value with the symbol will be included in the listings, unless otherwise specified. For example, a value of <0.01 will be analyzed as 0.01 and listed as <0.01.

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

For the purposes of haematology and clinical chemistry, baseline and maximum or minimum value post-baseline will be calculated over the entire post-baseline period, including the post-treatment period.

For the liver function tests: Aspartate Aminotransferase (AST), Alanine Aminotransferase (ALT), Alkaline phosphatase, 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.

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

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

4.6.3.2 Presentations

All continuous laboratory parameters will be summarized by absolute value at each visit by treatment group, together with the corresponding changes from baseline. The summary statistics presented will be the minimum, 1st quartile, median, 3rd quartile, maximum, mean and SD. Mean changes from baseline over time will also be plotted by treatment group.

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

Shift plots showing each individual patient's laboratory value at baseline and at maximum/minimum 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. A diagonal line indicating no change, and horizontal and vertical reference lines indicating the limits of the AstraZeneca defined reference ranges will also be displayed on the shift plots.

Data for patients who have treatment-emergent changes outside the predefined criteria will be presented. This data presentation will include all visits for this subset of patients.

The frequency of changes with respect to normal ranges between baseline and each post-treatment time point will be tabulated. Frequencies of clinically noteworthy values (using AstraZeneca defined reference ranges) occurring during the clinical study will also be given.

Any data outside the AstraZeneca normal and extended reference ranges will be explicitly noted on the listings that are produced.

4.6.4 Clinical Laboratory, Urinalysis

4.6.4.1 Definitions and Derivations

Urine samples for determination of urinalysis parameters will be taken at the times detailed in the CSP, and will be assessed in a central laboratory.

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

For the purposes of urinalysis shift tables, baseline and maximum or minimum value post-baseline will be calculated over the entire post-baseline period, including the post-treatment period.

4.6.4.2 Presentations

For urinalysis data, a shift table will be generated to present changes from baseline to EOT. The number of patients with treatment-emergent changes will also be summarized. Here, treatment-emergent changes are defined as 1) None/Trace at baseline to +, ++, +++, ++++ at any visit after baseline or 2) Increase of at least ++.

4.6.5 Other Laboratory Evaluations

Not Applicable

4.6.6 Vital Signs and weight

4.6.6.1 Definitions and Derivations

Pre-dose vital signs and weight (pulse, systolic blood pressure, diastolic blood pressure, respiration rate, and body temperature) will be obtained in accordance with the schedule provided in the protocol. Weight will be recorded in kilograms (kg).

Changes in vital signs and weight variables between baseline and each subsequent scheduled assessment will be calculated. Please refer to Section 3.3.1.2 for baseline definition. The change from baseline is defined as the post-baseline visit value minus the baseline visit value. There will be no imputation for missing values.

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

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

4.6.6.2 Presentations

Vital sign parameters will be presented for each treatment group. Summary statistics for continuous variables cover n, mean, SD, Minimum, Q1, median, Q3, and Maximum. Frequency tables cover number and percentage of patients in the respective category.

For each scheduled post-baseline visit, descriptive statistics for all vital sign parameters (systolic and diastolic blood pressure, pulse, respiration rate, oral temperature) and body weight will be presented for observed values and change from baseline.

Changes in vital signs and weight will be examined according to visit window defined in Table 3. Frequencies of clinically noteworthy values occurring during the clinical study will be presented using AstraZeneca defined reference ranges, and clinically important change criteria.

All recorded vital signs data will be listed.

4.6.7 Electrocardiogram

4.6.7.1 Definitions and Derivations

The outcome of the overall evaluation is to be recorded as normal/abnormal in the CRF, with any abnormalities being recorded as not clinically significant or clinically significant.

4.6.7.2 Presentations

The Investigator's assessment of the 12-lead ECG (normal or abnormal) will be listed for all patients, along with detailing whether any abnormalities were clinically significant or not.

The number and percentage of patients with clinically significant abnormal ECGs will be summarized by treatment group. Only ECG at baseline will be included.

4.6.8 Other Safety Assessments

Not Applicable.

5 INTERIM ANALYSIS

Not Applicable.

6 REFERENCES

Bartlett 2018
Bartlett J W, Covariate adjustment and estimation of mean response in randomised trials. Pharm Stat. 2018 Sep; 17(5):648-666.

Lewis and Heaton 1997
Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. Scan J Gastroenterol 1997;32:920-4.

User's manual for the SF-36v2 Health Survey. 3rd ed. Lincoln, RI: QualityMetric Incorporated.

Hirano, I., et al. 2019
Prospective Evaluation of a Novel, Endoscopic Activity Assessment System for Eosinophilic Gastritis. Gastroenterology 156(6): S-71-S-72.

Shoda, T., et al. 2020
Molecular, endoscopic, histologic, and circulating biomarker-based diagnosis of eosinophilic gastritis: Multi-site study. J Allergy Clin Immunol 145(1): 255-269.

Dellon et al 2013b
Dellon ES, Irani AM, Hill MR, Hirano I. Development and field testing of a novel patient-reported outcome measure of dysphagia in patients with eosinophilic esophagitis. Aliment Pharmacol Ther 2013;38:634-42.

7 APPENDIX

7.1 Accounting for missing data

7.1.1 Accounting for missing data for change in SAGED at Week 24

The amount of missing data is minimized in this study as all patients switch to receive benralizumab after the first 24 weeks and are encouraged to complete visits until they withdraw from the study even if they discontinue treatment. In addition, in the primary analyses most sources of missing data are accounted for with the composite estimand strategy that imputes outcomes for patients who discontinue randomised treatment.

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

The methodology is outlined below for the change in SAGED score primary endpoint, but similar techniques may also be used for other endpoints if appropriate.

7.1.1.1 Missing data description

Tabular summaries for the percentage of patients by the reason for discontinuation of randomized treatment as well as for withdrawal from the study will be presented by treatment to describe why patients discontinue from randomized treatment or withdraw from the study. The time to discontinuation of randomized treatment and withdrawal from the study will be presented using Kaplan-Meier plots (overall and split by treatment related/not treatment related reason for discontinuation, as defined in Tables 6 and 7). Dependent on these outputs some of the additional exploratory analyses described below may be produced as deemed necessary to further understand the pattern of missing data.

7.1.1.2 Primary analysis under the Composite Strategy Estimand

The primary analysis for the primary endpoint of SAGED at Week 24 allows for differences in outcomes over the study treatment period up to 24 weeks. In this analysis, all the data up to Week 24 visit will be included with imputation of WOCF for patients experiencing intercurrent events. The primary analysis uses the ANCOVA method, treatment group will be included as an explanatory variable along with the baseline SAGED score, region, baseline steroids use, and location of disease as explanatory variables, and assumes that missing data is missing at random (MAR) and is a direct likelihood approach (DL).

7.1.1.3 Sensitivity analysis under the effectiveness estimand using the Missing at Random (MAR) assumption.

An effectiveness estimand sensitivity analysis will be explored where all the data before IPD/EOT will be used. Data after the intercurrent event until Week 24 will be treated as missing and a mixed effect model for repeated measures (MMRM) analysis will be used for the remaining data. The dependent variable in the MMRM model will be the change from baseline in the continuous outcome at Week 24 visit. Treatment group, baseline values, region, baseline steroid use, and location of disease, visit, and treatment group × visit will be the covariates. The variance-covariance matrix will be assumed to be unstructured (UN). If the procedure does not converge, then the Toeplitz, first-order autoregressive (AR(1)), compound symmetric (CS), and variance components (VC) variance-covariance matrices will be tried in that order. The estimate of the treatment effect will be based on a contrast from this MMRM model.

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 the 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.

7.1.1.4 Imputations

Sensitivity analyses using imputation techniques will include all patients with baseline and at least one evaluable post-baseline assessment prior to the occurrence of the events. The following 4 steps will be used to build the imputation datasets and perform analyses:

1. 100 datasets with a monotone missing pattern will be obtained, induced by Markov Chain Monte Carlo (MCMC) method on patients prior to the occurrence of the events, or Week 24, whichever comes first.

2. For each of the imputed datasets obtained in step 1, the remaining missing data for patients who were not rescued by Week 24 will be imputed using the regression method for the monotone pattern with adjustment for covariates including treatment groups, region, and baseline value of the response variable.

3. Each of the 100 imputed datasets will be analysed using the main statistical model. These 100 datasets will be saved.

4. Apply Rubin's rule (Rubin et al 1986, Rubin 1987) to combine analysis results (point estimates and standard errors) from 100 imputations. Descriptive statistics

including number of patients, mean, standard error, and least squares (LS) means will be provided for each timepoint. In addition, difference in LS means and the corresponding 95% confidence intervals (CI) will be provided along with the p-values for Week 24 and all earlier time points in turn.

7.1.1.5 Sensitivity analyses using both MAR and MNAR assumptions

As described in Section 4.2.1.5, if the amount of missing data is high enough to warrant further exploration, sensitivity analyses for the primary and key secondary endpoints based on different missing data mechanism assumptions will be used to explore the robustness of any treatment effect including multiple imputation approaches.

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 methods introduced in [1, 2, 3] and further developed at AstraZeneca [4, 5] which allow for different underlying assumptions to be used. This model will assume that some pre-specified subset of patients who have missing data at Week 24 have correlations with future unobserved visits similar to patients in the placebo arm. The assumptions that will be used to impute the missing data are as follows:

(a) Missing at Random (MAR): Assumes that the trajectory for patients with missing data in each arm is similar to those observed in their own treatment arm.

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

Under the primary estimand, where treatment discontinuations are handled with a composite strategy, method (a) would be used for any missing data other than treatment/study discontinuation. In addition, to explore potentially informative drop out reasons, instead of the WOCF imputation implemented in the primary analysis, methods (a) and/or (b) may be used.

A summary of reasons for patients withdrawing from the benralizumab treatment arm and the corresponding treatment arm used to calculate the imputation under MAR and DRMI are given in Table 10.

Table 10 Summary of reasons for patients withdrawing from the benralizumab treatment arm and the corresponding treatment arm 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
Subject lost to follow up	Benralizumab	Benralizumab
Subject decision	Benralizumab	Based on review prior to study unblinding
COVID-19	Benralizumab	Benralizumab
Other	Benralizumab	Based on review prior to study unblinding

Note: Patients in the placebo arm are imputed using the mean of the non-missing values in placebo arm.

Some reasons for withdrawal are clearer to determine as treatment related (AEs, Death, Development of study-specific discontinuation criteria) or non-treatment related (patients lost to follow up, eligibility criteria not fulfilled). Other reasons are less clear such as subject decision and 'Other'; a review of each patient who withdraws from the study will therefore be carried out prior to unblinding the study. Based on this review the default assumptions for DRMI as described in (b) and Table 10 may be changed. A list of these patients and the assumptions made under DRMI will be documented prior to unblinding of the study.

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

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

The methodology described above to explore any potential impact of missing data for the SAGED change at Week 24 endpoint may be repeated for other endpoints if appropriate.

References

1. 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.

2. 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.

3. 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.

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

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

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

7. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. Biometrika 1999; 86:948-955.

7.2 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 in order to classify treatment-emergent AEs and to classify prior/concomitant medications:

Adverse Events

- The missing day of onset of an AE will be set to:
  - First day of the month that the event occurred, if the onset YYYY-MM is after the YYYY-MM of first study treatment
  - The day of the first study treatment, if the onset 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 treatment.
- The missing day of resolution of an AE will be set to:
  - The last day of the month of the occurrence. If the patient died in the same month, then set the imputed date as the death date.
- If the onset date of an AE is missing both the day and month, the onset date will be set to:

  - January 1 of the year of onset, if the onset year is after the year of the first study treatment
  - The date of the first treatment, if the onset year is the same as the year of the first study treatment
  - The date of informed consent, if the onset year is before the year of the first treatment
- If the resolution date of an AE or end date of an IP is missing both the day and month, the date will be set to:
  - December 31 of the year of occurrence. If the patient died in the same year, then set the imputed date as the death date.

Prior/concomitant medication

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

otherwise, the end date will be set to the start date of the therapy.

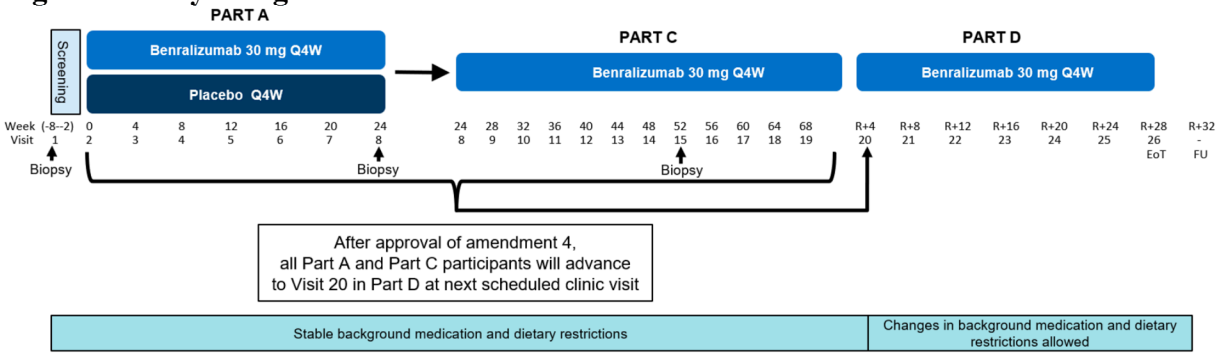
8 ADDENDUM

Protocol Version 5 (Amendment 4) was issued on March 28, 2023, after the decision to close recruitment to the study early. The amendment rationale was to fulfil a commitment to participants, from when they enrolled in the trial, of open-label access to benrazilumab following double-blind treatment.

Patients that completed Part A prior to approval of amendment 4 could choose to continue to Part C, an extended open-label benralizumab treatment period. Patients in Part A and Part C, after protocol amendment approval in their respective country, will be given the option at their next clinic visit of withdrawing from the study or moving on to Part D and receiving 6 months of open-label benralizumab Q4W dosing. Patients will not complete their current treatment period. Patients who have discontinued IP during Part A are not eligible to enter Part C or D.

Enrollment was terminated after enrollment of 34 participants, of which 12 were randomized. The figure below illustrates the current study design described above.

Figure 4 Study Design Flow Chart



Q4W: every fourth week; R: reference visit number of weeks (ie, visit prior to approval of Amendment 4); EoT: end of treatment; FU: follow-up.

Given the original sample size of the study will no longer be recruited and the study will be terminated, the analysis plan for the study was updated. The amended analysis plan will now consist of descriptive summaries of safety data and select patient listings of baseline characteristics, safety data, and key efficacy data for the double-blind placebo-controlled period (part A) and the open-label period (Parts C and D). Details on the descriptive summaries can be found below in Table 11.

Principles outlined in the main SAP document related to these descriptive summaries below will apply i.e. definition of periods for safety summaries. Other analyses described in the main SAP will no longer be implemented given the small final sample size limiting the ability to assess the original study objectives.

Table 11 List of summaries for the final analyses

Summaries	Period	Population	Format
Subject disposition	DB + OLE	All patients analysis set	Table
Duration of exposure	DB	Safety analysis set	Table
Duration of exposure	OLE	Open-label benralizumab analysis set	Table
Adverse events in any category	DB	Safety analysis set	Table
Adverse events in any category	OLE	Open-label benralizumab analysis set	Table
Serious adverse events – key subject information	DB	Safety analysis set	Table
Adverse events causing discontinuation of study treatment – key subject information	DB + OLE	Safety analysis set	Table
Haematology and clinical chemistry, baseline to minimum/maximum value post-baseline, on-treatment	DB	Safety analysis set	Table
Haematology and clinical chemistry, post-baseline values outside central laboratory reference ranges – key subject information	DB	Safety analysis set	Table
Subjects discontinued from the study	DB + OLE	All patients analysis set	Listing
Subjects completing the study	DB + OLE	All randomized subjects	Listing
Demographic and baseline characteristics		All randomized subjects	Listing

Additional demography		All randomized subjects	Listing
EG/EGE Disease History	DB	Full analysis set	Listing
Respiratory Disease History	DB	Full analysis set	Listing
Concomitant medication during study treatment	DB + OLE	Full analysis set	Listing
Peak eosinophil count (eos/hpf)	DB + OLE	Full analysis set	Listing
PGI-S and PGI-C	DB + OLE	Full analysis set	Listing
CGI-S and CGI-C	DB + OLE	Full analysis set	Listing
Symptom Assessment for Gastrointestinal Eosinophilic Diseases (SAGED) score	DB + OLE	Full analysis set	Listing
EG-REFS Score	DB + OLE	Full analysis set	Listing
EoD-REFS Score	DB + OLE	Full analysis set	Listing
EoE-REFS Score	DB + OLE	Full analysis set	Listing
Adverse events	DB + OLE	Safety analysis set	Listing
Laboratory data – Haematology	DB + OLE	Safety analysis set	Listing
Laboratory data – Chemistry	DB + OLE	Safety analysis set	Listing
Laboratory data – Urinalysis	DB + OLE	Safety analysis set	Listing
Vital signs	DB + OLE	Safety analysis set	Listing
SAGED score change from baseline in mean value by time point, WOCF, line plot	DB	Full analysis subset	Figure

SAGED score components mean value at Week 24, WOCF, line plot	DB	Full analysis subset	Figure

SIGNATURE PAGE

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

Document Name: d3258c00001-sap-ed-1		
Document Title:	Statistical Analysis Plan Edition 1	
Document ID:	Doc ID-005280488	
Version Label:	1.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
25-Oct-2023 14:06 UTC	PPD [redacted]	Author Approval
19-Oct-2023 11:31 UTC		Content Approval
19-Oct-2023 18:02 UTC		Content Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.